

Partial volume correction strategies for quantitative FDG PET in oncology

Nikie J. Hoetjes · Floris H. P. van Velden · Otto S. Hoekstra · Corneline J. Hoekstra · Nanda C. Krak · Adriaan A. Lammertsma · Ronald Boellaard

Received: 25 January 2010 / Accepted: 6 April 2010 / Published online: 27 April 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract

Purpose Quantitative accuracy of positron emission tomography (PET) is affected by partial volume effects resulting in increased underestimation of the standardized uptake value (SUV) with decreasing tumour volume. The purpose of the present study was to assess accuracy and precision of different partial volume correction (PVC) methods.

Methods Three methods for PVC were evaluated: (1) inclusion of the point spread function (PSF) within the reconstruction, (2) iterative deconvolution of PET images and (3) calculation of spill-in and spill-out factors based on tumour masks. Simulations were based on a mathematical phantom with tumours of different sizes and shapes. Phantom experiments were performed in 2-D mode using the National Electrical Manufacturers Association (NEMA) NU2 image quality phantom containing six differently sized spheres. Clinical studies (2-D mode) included a test-retest study consisting of 10 patients with stage IIIB and IV non-small cell lung cancer and a response monitoring study consisting of 15 female breast cancer patients. In all studies tumour or sphere volumes of interest (VOI) were generated using VOI based on adaptive relative thresholds.

Results Simulations and experiments provided similar results. All methods were able to accurately recover true SUV within 10% for spheres equal to and larger than 1 ml. Reconstruction-based recovery, however, provided up to twofold better precision than image-based methods. Clinical studies showed that PVC increased SUV by 5–80% depending on tumour size. Test-retest variability slightly worsened from 9.8 ± 6.5 without to $10.8 \pm 7.9\%$ with PVC. Finally, PVC resulted in slightly smaller SUV responses, i.e. from -30.5% without to -26.3% with PVC after the first cycle of treatment ($p < 0.01$).

Conclusion PVC improves accuracy of SUV without decreasing (clinical) test-retest variability significantly and it has a small, but significant effect on observed tumour responses. Reconstruction-based PVC outperforms image-based methods, but requires dedicated reconstruction software. Image-based methods are good alternatives because of their ease of implementation and their similar performance in clinical studies.

Keywords Partial volume correction · Positron emission tomography (PET) · FDG · Standardized uptake value (SUV) · Oncology

N. J. Hoetjes (✉) · F. H. P. van Velden · O. S. Hoekstra · A. A. Lammertsma · R. Boellaard
Department of Nuclear Medicine & PET Research,
VU University Medical Center,
P.O. Box 7057, 1007 MB Amsterdam, The Netherlands
e-mail: n.hoetjes@vumc.nl

C. J. Hoekstra
Department of Nuclear Medicine, Jeroen Bosch Hospital,
Den Bosch, The Netherlands

N. C. Krak
Department of Radiology, Erasmus Medical Center,
Rotterdam, The Netherlands

Introduction

In oncology, positron emission tomography (PET) using ^{18}F -2-fluorodeoxy-D-glucose (FDG) is increasingly being used for differentiation between benign and malignant tumours, for tumour staging, for diagnosis of post-treatment recurrence, for determination of response to therapy and for prognostic purposes. Most of these applications take advantage of the capability of PET to quantify tissue radioactivity concentration [1].

Amongst various approaches to assess FDG uptake, the standardized uptake value (SUV) is widely used as a semi-quantitative index of tumour uptake [2]. SUV represents radioactivity concentration at a certain time, normalized for body distribution (i.e. body weight, body surface area or lean body mass) and injected dose. Apart from the fact that SUV is time dependent, i.e. changing with time after FDG injection [3], it also is affected by more technical factors, such as method of data acquisition, image reconstruction algorithm and applied region of interest (ROI) strategy [4]. In addition, the so-called partial volume effect results in increasing underestimation of SUV with decreasing tumour volume [5].

The partial volume effect is due to the limited spatial resolution of PET scanners (typically 5–7 mm full-width at half-maximum, FWHM) and it can be expressed by a recovery coefficient, i.e. the ratio between measured and true activity concentrations [6, 7]. The low spatial resolution results in partial spread of signal from a tissue of interest into neighbouring tissues ('spill-out'), but also from those neighbouring tissues into the target tissue ('spill-in'). For tumours smaller than the reconstructed spatial resolution, the partial volume effect may result in a more than 50% underestimation of the true FDG concentration [6].

Various partial volume correction (PVC) methods have been developed to overcome or correct for the limited resolution of a clinical PET scanner [6, 8–10]. Some of these methods, specifically designed for human brain studies, use high-resolution anatomical images to correct for partial volume effects [9, 11, 12]. These methods generally use a coregistered structural magnetic resonance imaging (MRI) scan that is segmented into grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) regions. These segmented MRI scans can be used to correct reconstructed PET images for partial volume effects either at a parametric [13, 14] or a ROI level [15, 16]. A disadvantage of the use of these high-resolution anatomical images is the need for accurate registration with the PET images [17]. Inaccurate segmentation or misregistration can result in PVC errors. Nevertheless, when these techniques are applied to brain studies, they work reasonably well [15, 16]. For whole-body studies computed tomography (CT) can also be used for anatomical imaging. In cases of oncological applications, however, there are several limitations: (a) metabolic PET and anatomical CT/MRI tumour boundaries do not necessarily coincide, for example in cases of necrotic tumours, (b) radiotherapy can lead to tissue scarring and both CT and MRI have limitations in differentiating scar tissue from viable tumour [18, 19], and (c) accurately delineating tumour may be difficult, as (non-enhanced) CT images lack contrast between normal soft tissue structures and tumour extensions. In addition, (d) inaccurate segmentation and misalignment also result in

PVC errors as is the case when one would use MRI data for the same purpose.

Another possible PVC strategy of SUV data is to use a lookup table with PVC factors as function of tumour volume, where tumour volumes are either derived from a PET or a (separate) CT study [20, 21]. It is, however, not easy to estimate the true metabolic size on PET or CT images and tumour shapes do not necessarily correspond with phantom shapes (spheres) used to derive these factors. Another option for generating PVC images is to include the scanner point spread function (PSF) in the reconstruction process [22, 23].

The purposes of the present study were (a) to evaluate the performance of several PET-based PVC methods for oncological whole-body FDG PET studies and (b) to study the effects of PVC on SUV accuracy, test-retest variability and observed SUV responses.

Materials and methods

Reconstruction-based partial volume correction

The PVC ordered subsets expectation maximization (PVC-OSEM) reconstruction algorithm similar to that of Brix et al. [22] and Reader et al. [23] was implemented by taking the PSF of the imaging system into account.

The PVC-OSEM algorithm, used in this study, works as follows. The starting image, in this case a uniform image with all voxels set to 1.0, is blurred with an isotropic 3-D Gaussian shaped kernel, representing the 3-D PSF of the system, resulting in a smoothed starting image. By forward projecting this smoothed starting image, an estimated sinogram is generated. Next, an updated sinogram is calculated as the ratio between measured and estimated sinograms. Subsequently, an updated image is generated by backprojection of the updated sinogram. The estimated image is multiplied by the updated image, resulting in a sharper representation of the estimated image. This process is repeated a certain number of times (iterations). The algorithm has been described in detail elsewhere [24]. PVC-OSEM with matched projectors (PVC-OSEM-M) additionally smooths the updated image with a 3-D Gaussian shaped kernel that represents the PSF of the system. This second additional smoothing is not performed for PVC-OSEM using non-matched projectors (PVC-OSEM-NM).

Image-based partial volume correction

Iterative deconvolution methods Deconvolution is an image restoration process used to recover spatial resolution, restore edges and improve contrast. Two PVC methods based on iterative deconvolution were evaluated: Lucy-Richardson (LR) [25] and Van Cittert deconvolution (VC) [17].

The LR method is initialized using a copy of the original PET image. This copy is smoothed with a Gaussian shaped kernel that represents the PSF of the scanner. This smoothed copy is subtracted from the original PET image and the difference is added to the copy of the original unsmoothed image. Subsequently, the process starts again with the sharpened copy. This loop is repeated until the smoothed copy is equal to the original PET image, resulting in a sharper PET image. The VC method is algorithmically similar to the LR method, except that the ratio instead of the difference between original PET image and smoothed copy is used to update the copy of the original PET image. Note that LR and VC are very similar with the exception that LR allows for negative pixel values.

Mask-based method In this method a tumour volume of interest (VOI) is used as a mask and smoothed using a PSF that matches the blurring effects of the PET scanner, including the smoothing effects of reconstruction filters. In the present study a isotropic 3-D Gaussian kernel of 6.5 mm FWHM provided a good description of the shape of the PSF, as the image resolution was primarily determined by (standard) Gaussian post-smoothing of the iteratively reconstructed PET images. The difference in average voxel value between smoothed and original non-smoothed mask is a direct measure of the partial volume effect (spill-out) for the tumour with its specific shape and size. Spill-in of signal from surrounding activity is derived by using an ‘inverted’ mask, i.e. setting the voxel values within the mask to zero and the surrounding pixels to the measured average background value, and following the same procedure. It should be noted that mask-based methods inherently assume homogeneous tumour uptake.

Simulation studies

Simulations, representing 2-D mode acquisitions, were performed using methodology and software as described previously [4]. In short, an anatomically realistic mathematical phantom was derived from a 2-D dynamic FDG PET scan (ECAT EXACT HR+ scanner, CTI/Siemens, Knoxville, TN, USA) of a patient. The scan data, acquired from 15 to 60 min after administration of 370 MBq FDG, were summed and the resulting sinogram was reconstructed using OSEM [26] with 2 iterations and 16 subsets and post-smoothed using a 6 mm FWHM Gaussian filter. It was assumed that a nearly noise-free image was obtained by first summing over 45 min and then applying additional smoothing. Subsequently, voxel values were scaled, so that the average soft tissue activity concentration equalled 5 kBq/ml, which approximately corresponds to the average soft tissue activity concentration observed in patient studies at 50–60 min after injection of 370 MBq FDG. Tumours of various sizes and shapes were simulated by inserting volumes with increased uptake in this mathematical phantom.

Next, noise-free sinograms were generated by forward projection of the obtained image (i.e. entire 3-D volume). These sinograms contained the number of true coincidences per line of response. Random and scattered coincidences were added to obtain sinograms with prompt coincidences. Randoms were assumed to be distributed uniformly over the sinogram. Scattered coincidences were derived from forward projection of the difference between a scatter-corrected and non-scatter-corrected reconstructed image, using the same image data as used for generating the true sinogram. Poisson noise was added to the prompts, randoms and scatter sinograms. Finally, true coincidences sinograms were generated by subtracting the noisy random and scatter sinograms from the noisy prompt sinograms. Number of counts corresponded to 2-D data obtained for 5–7 min or about 1.5×10^7 noise equivalent counts (NEC).

In total, 100 noisy images were obtained to derive the variability of the ‘measured’ activity concentration both with and without PVC and to determine to what extent reproducibility of activity concentration measurements are affected by PVC.

Phantom experiment

A phantom experiment was performed using the National Electrical Manufacturers Association (NEMA) NU2 image quality phantom containing 6 spheres with volumes of 26.5, 11.5, 5.6, 2.6, 1.2 and 0.5 ml (with diameters of 37, 28, 22, 17, 13 and 10 mm, respectively). The background compartment of the phantom was filled with an FDG solution of 2 kBq/ml. All spheres were filled with a solution of 20 kBq/ml, resulting in a tumour to background ratio (TBR) of 10. The phantom was scanned for 60 min on an ECAT EXACT HR+ scanner (Siemens/CTI, Knoxville, TN, USA) in 2-D mode [27]. This scanner has an axial field of view of 15 cm and produces 63 transaxial slices with a slice thickness of 2.5 mm. A 10-min transmission scan was then acquired for attenuation and scatter correction purposes. For this study, the average activity concentrations in the spheres were compared with the true activity concentration.

Clinical studies

Scans were performed in two separate groups of patients. To study the impact of PVC on test-retest variability, a group of ten patients (seven men and three women, mean age: 50 ± 5 years) with stage IIIB and IV non-small cell lung cancer was scanned twice on 2 consecutive days before receiving any treatment [28]. To investigate the effect of PVC on PET response, a group of 15 female patients (mean age: 52 ± 8 years) with locally advanced breast cancer was scanned before treatment and after one course of chemotherapy [29]. Both studies had been approved by the

Medical Ethics Committee of the VU University Medical Center and all patients gave written informed consent.

For both studies dynamic scans were performed using an ECAT EXACT HR+ PET scanner in 2-D acquisition mode. Patients were studied in the supine position with arms down and with the dominant lesions in the centre of the field of view. During the first scan, the distance between the suprasternal notch and the upper side of the field of view (laser beam alignment) was recorded and used for patient positioning during the second scan.

All patients fasted for 6 h prior to scanning. Patients received two venous catheters, one for injection of FDG contralateral to the tumour, the other for venous blood sampling. Prior to injection of FDG, a blood sample was collected for determination of plasma glucose level. Acquisition started with a 10- to 15-min transmission scan [28] to correct for attenuation and scatter, followed by a bolus injection of 370 MBq of FDG in 5 ml saline through an injector (Medrad International, Maastricht, The Netherlands) with 0.8 ml/s, after which the line was flushed with 42 ml saline (2.0 ml/s). At the time of the FDG injection, a dynamic emission scan (in 2-D mode) was started with a total duration of 60 min and with variable frame lengths (6×5, 6×10, 3×20, 5×30, 5×60, 8×150 and 6×300 s). All dynamic scan data were corrected for decay, dead time, scatter, random coincidences and measured photon attenuation.

In the present study, SUV with and without PVC were calculated using the sum of the last three frames (i.e. 45–60 min post-injection). In total, 30 and 18 lesions were assessed in test-retest and response studies, respectively.

Reconstructions

All studies were reconstructed using both regular 2-D OSEM and PVC-OSEM (either with matched or non-matched projectors) with attenuation correction weighting. Regular OSEM reconstructions were performed with 2 iterations and 16 subsets, and PVC-OSEM with 4 and 8 iterations for non-matched and matched projectors, respectively, and 16 subsets. Reconstructed image matrix size equalled 256×256 providing a voxel size of 2.57×2.57×2.43 mm. Image reconstructions included all routine corrections required for quantification, such as scatter correction, attenuation correction, normalization, and dead time and decay corrections. Images without PVC were post-smoothed using a Gaussian filter of 5 mm FWHM resulting in an image resolution of about 6.5 mm FWHM. After reconstruction, VC-, LR- or mask-based PVC methods were applied to the regular (smoothed) 2-D OSEM reconstructed images.

Data analysis

Tumour VOI were defined using a 3-D region growing algorithm, as described previously [4, 30]. This algorithm is

based on the 3-D search algorithm in the IDL software package version 6.3 (Interactive Data Language, Research Systems Inc., Boulder, CO, USA). In short, the program first searched for the location of the maximum voxel value within a (semi-automatically or manually) predefined region. Next, using this maximum value and its location as starting point, a 3-D VOI was defined automatically using a 3-D region growing algorithm, including all voxels above a specified threshold. This threshold was set at 41 or 50% of the sum of maximum and background values. The latter (local) background value was derived automatically using a 3-D shell of 1 voxel thickness at 1.5 cm from the border of the initially estimated or (pre-)defined tumour volume. This initial estimate is based on the 70% of maximum pixel value 3-D isocontour [4, 29]. For volumes larger than 4 ml, a threshold of 41% gives a reasonable approximation of the true (tumour) volume [31]. For smaller volumes (<4 ml), however, a decrease in TBR increases the optimum threshold value due to the edge-blurring effect of the background. A threshold value of 41%, adapted for local background, worked very well for simulation and phantom data, but during analysis of clinical data it occasionally failed to provide reliable and realistic tumour VOI. Therefore, a higher threshold of 50%, adapted for local background, was used for clinical data.

In addition to the PET-based VOI (PET VOI) mentioned above, a second set of VOI was used for the simulation studies. These VOI (exact VOI) were placed in exactly the right position, and had the exact size and shape of the simulated tumours, thereby excluding effects of PET-based VOI inaccuracies and uncertainties.

For both simulation and phantom data the average activity concentration within a VOI was measured. Subsequently, recovery coefficients were derived by dividing these observed activity concentrations by the ‘true’ simulated or experimentally determined (well counter, calibrated against the dose calibrator) activity concentrations.

For clinical studies, tumour VOI were defined on the non-PVC-corrected images, smoothed with a Gaussian kernel of 6.5 mm FWHM, before they were applied to the various PVC images. SUV were normalized to body weight and injected dose, and corrected for plasma glucose level. In the case of clinical data ‘true’ SUV are not known and therefore these data were only used to assess the impact of PVC on test-retest variability or to investigate the effect of PVC on (measured) response.

Results

LR and VC methods provided nearly identical results. Therefore, in the remainder only results of the LR method will be shown.

Simulation studies

In the case of (standard) PET-based VOI, derived recovery coefficients showed that all PVC methods investigated were able to accurately correct for partial volume effects if spheres were larger than 4 ml (Fig. 1a). When exact VOI was used, recovery coefficients of PVC data were within 20%, even for the smallest VOI size (Fig. 1c).

Figure 1b, d show corresponding coefficients of variation (COV) of the recovery coefficients for the various sphere sizes. For both PET VOI and exact VOI the COV of the recovery coefficient increases with decreasing sphere volume and after applying PVC. In general, PVC-OSEM with matched projectors showed the smallest increase (<30%) in COV for both PET and exact VOI, when compared to all other PVC methods (<46%).

Phantom experiment

The recovery coefficients of the six spheres positioned in the NEMA NU2 image quality phantom are shown in Fig. 2 for all PVC methods investigated. All these methods were able to accurately recover true activity within 20% for spheres equal or larger than 1 ml (Fig. 3).

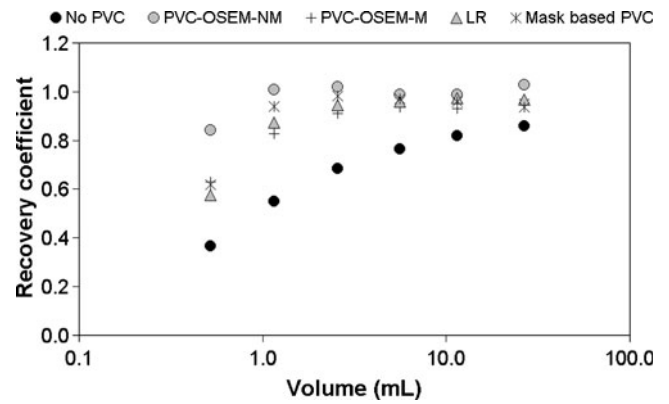


Fig. 2 Recovery coefficients of the six spheres positioned in the NEMA NU2 image quality phantom. Note that x-axis is plotted using a logarithmic scale

Clinical studies

Figure 4 shows ratios between SUV with and without PVC as function of tumour metabolic VOI, illustrating that PVC increases SUV by 5–80% depending on metabolic volume or VOI size.

In Fig. 5 percentage test-retest with PVC is shown as function of test-retest without PVC. Test-retest variability

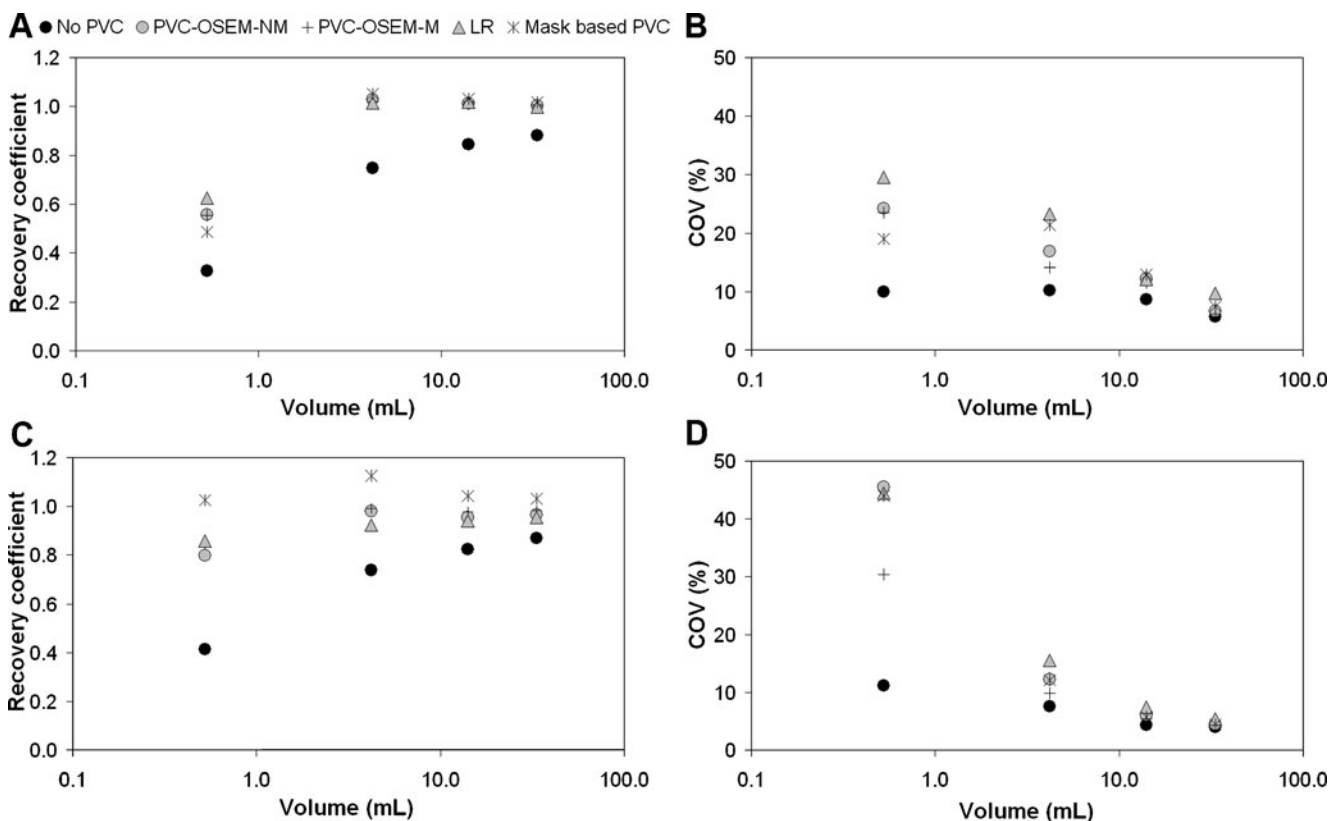


Fig. 1 Recovery coefficient (a, c) and its corresponding coefficient of variation (COV, b, d) as function of sphere size for PET-based VOI (a, b) and exact VOI (c, d), respectively. Note that x-axis is plotted using a logarithmic scale

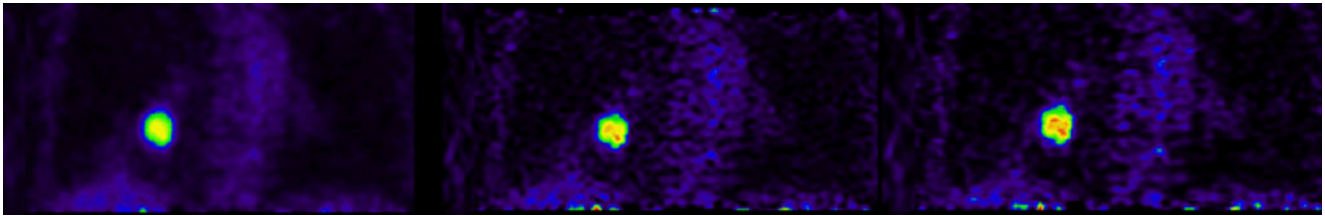


Fig. 3 Visual comparison of PVC-corrected images. From *left to right*: no PVC, imaged-based PVC (LR) and reconstruction-based PVC (PVC-OSEM-M)

slightly worsened from $9.8 \pm 6.5\%$ before to $10.8 \pm 7.9\%$ after PVC. All methods showed an increase in test-retest variability with decreasing metabolic tumour volume. The p values of applying PVC compared to not applying PVC ranged from 0.17 to 0.39 for the different methods.

Figure 6 illustrates the decrease in SUV after the first cycle of chemotherapy for the various PVC methods and provides the corresponding p values of the changes compared to baseline. PVC resulted in slightly smaller SUV responses, i.e. from -31% without to -26 or -27% with PVC. Although small, this effect was highly significant for all PVC methods ($p < 0.05$), except for the mask-based method ($p > 0.05$).

Discussion

In this study the performance of various PVC methods (either image, mask or reconstruction based) was evaluated for whole-body oncological FDG PET studies, focussing on the impact of PVC on accuracy of SUV quantification, test-retest variability and magnitude of SUV response.

Simulation studies and phantom experiment

In clinical data the ground truth is not known and consequently these data cannot be used directly to assess the accuracy of PVC methods. Therefore, simulation and phantom experiments were performed to assess the performance of PVC. Clinical data are mainly used to show the impact of using PVC on test-

retest variability and FDG response assessment. In the simulation studies, all PVC methods accurately corrected SUV for partial volume effects when tumours (spheres) were larger than 4 ml (20 mm diameter). Interestingly, when known information about location and size was used (exact VOI), the accuracy of all PVC methods improved to within 20%, even for the smallest tumour size investigated. Apparently, accurate tumour VOI definition is needed for PVC of small structures. It could be hypothesized that this information could be obtained using CT. In practice, however, this appears not to be the case, as both observer variability in tumour definition [32–35] and target motion (e.g. due to breathing) will likely result in some mismatch between the ‘real’ tumour volume and that defined on CT.

The simulation studies also showed an increase in COV of the recovery coefficient with decreasing tumour volume following PVC. Of all the PVC methods investigated, PVC-OSEM with matched projectors showed the smallest increase in COV.

Results obtained using the NEMA NU2 image quality phantom were consistent with those seen during simulations. All PVC methods were able to accurately recover true SUV within 20% for spheres equal to or larger than 1 ml.

Clinical studies

In clinical data the true activity concentration and its distribution within a tumour are not known, making it impossible to

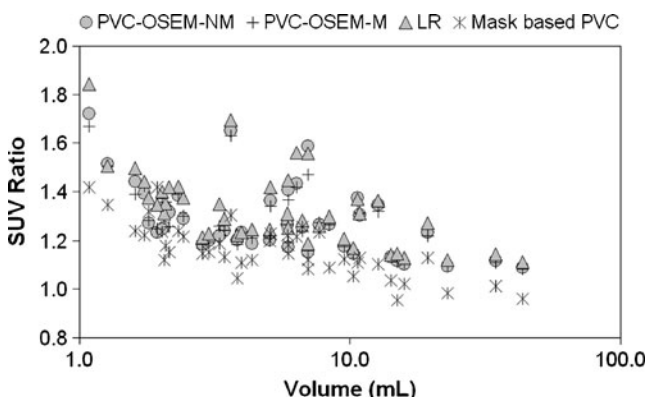


Fig. 4 Ratios between SUV with and without PVC plotted against metabolic tumour volume

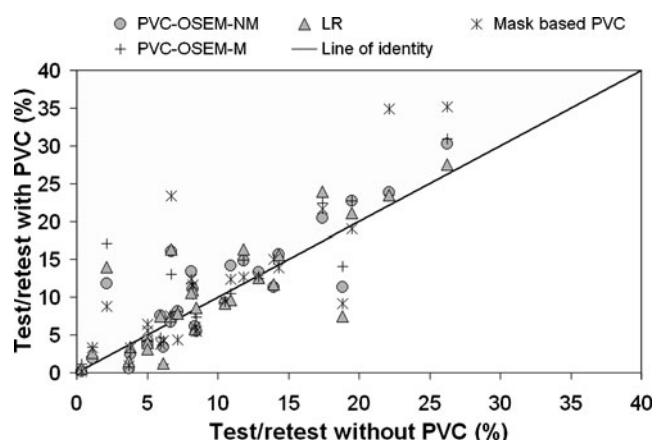


Fig. 5 Percentage test-retest with PVC plotted against test-retest without PVC

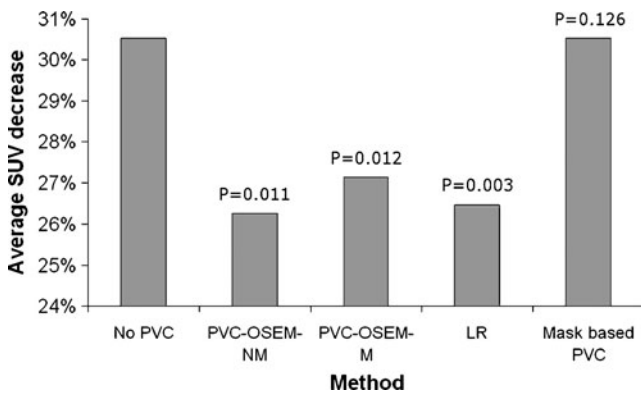


Fig. 6 Average decrease in SUV after the first cycle of chemotherapy, shown for various PVC methods. The p values of the PVC methods compared to no PVC are shown above the bars

assess the accuracy of any PVC method. Nevertheless, a number of important performance characteristics can be derived from clinical data, such as its impact on test-retest variability, on metabolic tumour volume dependence and on measured SUV response.

Test-retest variability of FDG SUV measurements acquired for the same patient on 2 consecutive days increased from 9.8% without to 10.8% with PVC, but this effect was not significant ($p > 0.17$). This may suggest that use of PVC does not worsen SUV precision and that it would therefore not affect the sensitivity of detecting changes in SUV for longitudinal studies provided that these were not caused by changes in size alone. However, the main reason for applying PVC is to rule out the effects of volumetric changes on SUV responses. As PVC did not change test-retest variability, it suggests that PVC provides more accurate assessments of treatment response without affecting its precision.

When assessing tumour response the accuracy of measured changes needs to be balanced against its reproducibility. In addition, PVC might be required to avoid incorrect SUV responses due to changes in tumour volume rather than uptake. All PVC methods investigated were able to reduce the dependency of SUV on tumour volume (Fig. 4). However, a small volumetric SUV dependency was still present for small tumours, which was also observed in the simulation studies and the phantom experiment. Eventually, application of PVC resulted in a significant effect on the observed SUV responses for most PVC methods (Fig. 6). However, use of the mask-based PVC method did not show a different SUV response as compared with the non-PVC SUV data. A possible explanation is that the mask-based method uses a VOI directly to derive correction factors (while this is not the case for the image- or reconstruction-based methods). In patient data reproducible VOI are more difficult to generate and may thus have a larger impact on the mask-based PVC method performance. As test-retest variability was not

significantly affected by using image- and reconstruction-based PVC, it may be concluded that both image- and reconstruction-based PVC can be used for more accurate, yet equally precise treatment response assessments.

Limitations/ future research

In the present implementation of all PVC methods, the PSF of the scanner was approximated by a single 3-D isotropic Gaussian function. For most scanners resolution is not spatially invariant, but decreases with increasing distance from the centre [27]. These spatial resolutions are measured using point sources in air and data are reconstructed using filtered backprojection using voxel sizes of about 1 mm (as recommended by NEMA NU 2 specifications). In oncological whole-body PET studies, however, iterative reconstruction algorithms are used with voxel sizes of 2.5–5 mm. In addition, resulting images are often post-smoothed using Gaussian filters of 5–10 mm (5 mm in the present study). Consequently, actual image resolution is primarily determined by image reconstruction settings (e.g. number of iterations and subsets used within OSEM) and post-reconstruction filtering. As these degrading effects are spatially invariant, the assumption that, in clinical studies, the PSF may be approximated by a single Gaussian function appears reasonable. This is supported by the good results obtained with the phantom experiment, where spheres were located at different positions within the field of view of the scanner. Moreover, the present results are consistent with those of Geworski et al. [6] who, when studying the feasibility of recovery correction in PET, showed that criteria for linearity and stationarity of a similar PET scanner were sufficiently met to allow for recovery correction. Nevertheless, future research should investigate whether implementation of spatial variation in PSF would improve results.

The performance of image-based PVC methods using iterative deconvolution can be further enhanced when a priori information is incorporated during the process. For example, use of Gibbs or median root priors may be helpful in reducing or avoiding image noise amplification during iterative deconvolution. The latter requires further investigations and optimization of prior settings and will be part of future research. With the present implementations of iterative deconvolution-based PVC methods, however, we have already shown that PVC can improve SUV accuracy and responses.

The present simulation studies suggest that more accurate PVC results can be obtained using exact VOI definition, or e.g. VOI derived from CT data. In clinical practice, however, CT-based VOI may suffer from observer variability [32–35]. Indeed, it has been reported that observer variation and misregistration range from 2 to 15 mm for the different image fusion methods [36, 37]. Accurate registration of PET

with CT is required, as misregistration can contribute to errors in PVC.

In both simulation and phantom studies only the effects of applying PVC on homogeneous tumours were investigated. In practice, however, tumours can be very heterogeneous. Tumour heterogeneity may lead to a decrease in contrast, which in turn may make tumour VOI definitions more problematic. From the clinical data we already observed that a higher threshold (50%) was needed to generate VOI reliably, while a lower threshold (41%) could be used for (simple) phantoms. Especially for image- and reconstruction-based PVC methods, the impact of using a somewhat higher threshold is likely minimal as both these PVC methods use only the PSF kernel for applying PVC. The use of the slightly higher threshold resulted in <5% higher SUV data for most tumours [29, 33], which is very small compared to the amplitude of PVC for most tumours, and it will therefore likely not affect the main conclusions of this paper. Of course, the best situation would be to have automated VOI methods that can take uptake heterogeneity into account and these methods will likely further increase the accuracy of SUV (with and without PVC). At present, novel algorithms are being developed to generate accurate VOI in cases of heterogeneous tracer uptake [38] and those methods will need to be explored further.

An issue not addressed in this study is the effect of breathing on PET-based VOI definition and SUV quantification. In cases of e.g. lung tumours, breathing will cause additional blurring of the PET images. This blurring will not only result in decreased SUV but also in incorrect PET-based VOI definitions [33, 39]. As the smoothing effect due to breathing is not isotropic, it cannot easily be corrected for using PVC strategies such as those in the present study. A potential solution for this problem is to perform respiratory-gated PET acquisitions. A PVC method may then be applied to each of the acquisition gates separately. The accuracy of this approach, however, needs to be established in future studies.

Finally, all image data in this study were acquired in 2-D mode. At present all modern PET or PET/CT systems are 3-D mode systems, having higher sensitivity and better count rate performances and thus providing data with better image quality and/or less noise. Collecting data with improved image quality or less noise will further enhance the applicability of the presented PVC methods. Especially image-based methods, which can be applied after reconstruction and thus off-line, cause a larger increase of image noise (Fig. 1) than reconstruction-based methods and may therefore be less suitable for scans acquired in 2-D mode. Reconstruction-based PVC should then be preferred, but it is not routinely available on all machines. Yet, already promising results were obtained for both reconstruction- as well as image-based PVC methods and it is expected that

these methods will perform at least equally well, if not better (due to smaller problems with image noise), in cases of 3-D mode acquisitions.

Conclusion

This study demonstrates that various PVC methods improve accuracy of tumour SUV estimation without decreasing test-retest variability. In addition, except for mask-based PVC, there is a significant but small effect on observed tumour responses. PVC-OSEM with matched projectors seems to be the optimal PVC method, as it provides best accuracy and precision. A disadvantage is the need for dedicated reconstruction software, which is not yet fully available for all PET/CT systems. The results in this paper demonstrate that both image- and reconstruction-based PVC can be applied to improve SUV accuracy without worsening test-retest variability. Consequently, PVC can be used for more accurate, yet equally precise treatment response assessments.

Acknowledgements This study was performed within the framework of CTMM, the Center for Translational Molecular Medicine, which provided financial support. AIRFORCE project (grant 03O-103).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Hoekstra CJ, Paglianiti I, Hoekstra OS, Smit EF, Postmus PE, Teule GJ, et al. Monitoring response to therapy in cancer using [18F]-2-fluoro-2-deoxy-D-glucose and positron emission tomography: an overview of different analytical methods. *Eur J Nucl Med* 2000;27(6):731–43.
2. Graham MM, Peterson LM, Hayward RM. Comparison of simplified quantitative analyses of FDG uptake. *Nucl Med Biol* 2000;27(7):647–55.
3. Lowe VJ, DeLong DM, Hoffman JM, Coleman RE. Optimum scanning protocol for FDG-PET evaluation of pulmonary malignancy. *J Nucl Med* 1995;36(5):883–7.
4. Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. *J Nucl Med* 2004;45(9):1519–27.
5. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. *J Nucl Med* 2007;48(6):932–45.
6. Geworski L, Knoop BO, de Cabrejas ML, Knapp WH, Munz DL. Recovery correction for quantitation in emission tomography: a feasibility study. *Eur J Nucl Med* 2000;27(2):161–9.
7. Hoffman EJ, Huang SC, Phelps ME. Quantitation in positron emission computed tomography: 1. Effect of object size. *J Comput Assist Tomogr* 1979;3(3):299–308.
8. Aston JA, Cunningham VJ, Asselin MC, Hammers A, Evans AC, Gunn RN. Positron emission tomography partial volume correction: estimation and algorithms. *J Cereb Blood Flow Metab* 2002;22(8):1019–34.
9. Boussion N, Hatt M, Lamare F, Bizais Y, Turzo A, Cheze-Le Rest C, et al. A multiresolution image based approach for correction of

- partial volume effects in emission tomography. *Phys Med Biol* 2006;51(7):1857–76.
10. Hickeson M, Yun M, Matthies A, Zhuang H, Adam LE, Lacorte L, et al. Use of a corrected standardized uptake value based on the lesion size on CT permits accurate characterization of lung nodules on FDG-PET. *Eur J Nucl Med Mol Imaging* 2002;29(12):1639–47.
 11. Baete K, Nuyts J, Van Laere K, Van Paesschen W, Ceyskens S, De Ceuninck L, et al. Evaluation of anatomy based reconstruction for partial volume correction in brain FDG-PET. *Neuroimage* 2004;23(1):305–17.
 12. Erlandsson K, Wong AT, van Heertum R, Mann JJ, Parsey RV. An improved method for voxel-based partial volume correction in PET and SPECT. *Neuroimage* 2006;31:T84.
 13. Meltzer CC, Leal JP, Mayberg HS, Wagner Jr HN, Frost JJ. Correction of PET data for partial volume effects in human cerebral cortex by MR imaging. *J Comput Assist Tomogr* 1990;14(4):561–70.
 14. Müller-Gärtner HW, Links JM, Prince JL, Bryan RN, McVeigh E, Leal JP, et al. Measurement of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects. *J Cereb Blood Flow Metab* 1992;12(4):571–83.
 15. Frouin V, Comtat C, Reilhac A, Grégoire MC. Correction of partial-volume effect for PET striatal imaging: fast implementation and study of robustness. *J Nucl Med* 2002;43(12):1715–26.
 16. Rousset OG, Ma Y, Evans AC. Correction for partial volume effects in PET: principle and validation. *J Nucl Med* 1998;39(5):904–11.
 17. Teo BK, Seo Y, Bacharach SL, Carrasquillo JA, Libutti SK, Shukla H, et al. Partial-volume correction in PET: validation of an iterative postreconstruction method with phantom and patient data. *J Nucl Med* 2007;48(5):802–10.
 18. Jones DR, Parker Jr LA, Detterbeck FC, Egan TM. Inadequacy of computed tomography in assessing patients with esophageal carcinoma after induction chemoradiotherapy. *Cancer* 1999;85(5):1026–32.
 19. Kitagawa Y, Nishizawa S, Sano K, Ogasawara T, Nakamura M, Sadato N, et al. Prospective comparison of 18F-FDG PET with conventional imaging modalities (MRI, CT, and 67Ga scintigraphy) in assessment of combined intraarterial chemotherapy and radiotherapy for head and neck carcinoma. *J Nucl Med* 2003;44(2):198–206.
 20. Pottgen C, Gauler T, Krbeek T, Kuhl H, Stergar H, Gohlert A, et al. F-18-FDG-PET/CT-progress examinations in a prospective phase-III protocol—a successful selection instrument for definitive radiochemotherapy in patients with locally advanced NSCLC? *Strahlenther Onkol* 2006;182:111–2.
 21. Vesselle H, Schmidt RA, Pugsley JM, Li M, Kohlmyer SG, Vallieres E, et al. Lung cancer proliferation correlates with [F-18] fluorodeoxyglucose uptake by positron emission tomography. *Clin Cancer Res* 2000;6(10):3837–44.
 22. Brix G, Doll J, Bellemann ME, Trojan H, Haberkorn U, Schmidlin P, et al. Use of scanner characteristics in iterative image reconstruction for high-resolution positron emission tomography studies of small animals. *Eur J Nucl Med* 1997;24(7):779–86.
 23. Reader AJ, Julyan PJ, Williams H, Hastings DL, Zweit J. EM algorithm system modeling by image-space techniques for PET reconstruction. *IEEE Trans Nucl Sci* 2003;50(5):1392–7.
 24. Mourik JE, Lubberink M, Klumpers UM, Comans EF, Lammertsma AA, Boellaard R. Partial volume corrected image derived input functions for dynamic PET brain studies: methodology and validation for [11C]flumazenil. *Neuroimage* 2008;39(3):1041–50.
 25. Tohka J, Reilhac A. Deconvolution-based partial volume correction in Raclopride-PET and Monte Carlo comparison to MR-based method. *Neuroimage* 2008;39(4):1570–84.
 26. Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans Med Imaging* 1994;13:601–9.
 27. Brix G, Zaers J, Adam LE, Bellemann ME, Ostertag H, Trojan H, et al. Performance evaluation of a whole-body PET scanner using the NEMA protocol. National Electrical Manufacturers Association. *J Nucl Med* 1997;38(10):1614–23.
 28. Hoekstra CJ, Hoekstra OS, Stroobants SG, Smit EF, Lammertsma AA. A comparison of different analytical methods for response monitoring in non-small cell lung cancer (NSCLC) using FDG PET. *J Nucl Med* 1999;40(5):140P.
 29. Krak NC, Boellaard R, Hoekstra OS, Twisk JW, Hoekstra CJ, Lammertsma AA. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. *Eur J Nucl Med Mol Imaging* 2005;32(3):294–301.
 30. Westerterp M, Pruim J, Oyen W, Hoekstra O, Paans A, Visser E, et al. Quantification of FDG PET studies using standardised uptake values in multi-centre trials: effects of image reconstruction, resolution and ROI definition parameters. *Eur J Nucl Med Mol Imaging* 2007;34:392–404.
 31. Erdi YE, Mawlawi O, Larson SM, Imbriaco M, Yeung H, Finn R, et al. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. *Cancer* 1997;80(12 Suppl):2505–9.
 32. Caldwell CB, Mah K, Ung YC, Danjoux CE, Balogh JM, Ganguli SN, et al. Observer variation in contouring gross tumor volume in patients with poorly defined non-small-cell lung tumors on CT: the impact of (18)FDG-hybrid PET fusion. *Int J Radiat Oncol Biol Phys* 2001;51(4):923–31.
 33. Nehmeh SA, Erdi YE, Ling CC, Rosenzweig KE, Squire OD, Braban LE, et al. Effect of respiratory gating on reducing lung motion artifacts in PET imaging of lung cancer. *Med Phys* 2002;29(3):366–71.
 34. Remeijer P, Rasch C, Lebesque JV, van Herk M. A general methodology for three-dimensional analysis of variation in target volume delineation. *Med Phys* 1999;26(6):931–40.
 35. Steenbakkers RJHM, Duppen JC, Fitton I, Deurloo KEI, Zijp L, Uitterhoeve ALJ, et al. Observer variation in target volume delineation of lung cancer related to radiation oncologist-computer interaction: a ‘Big Brother’ evaluation. *Radiother Oncol* 2005;77(2):182–90.
 36. Lavelly WC, Scarfone C, Cevikalp H, Li R, Byrne DW, Cmelak AJ, et al. Phantom validation of coregistration of PET and CT for image-guided radiotherapy. *Med Phys* 2004;31(5):1083–92.
 37. Nömayr A, Römer W, Hothorn T, Pfahlberg A, Hornegger J, Bautz W, et al. Anatomical accuracy of lesion localization. Retrospective interactive rigid image registration between 18F-FDG-PET and X-ray CT. *Nuklearmedizin* 2005;44(4):149–55. German.
 38. Hatt M, Cheze le Rest C, Turzo A, Roux C, Visvikis D. A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET. *IEEE Trans Med Imaging* 2009;28(6):881–93.
 39. Nehmeh SA, Erdi YE, Ling CC, Rosenzweig KE, Schoder H, Larson SM, et al. Effect of respiratory gating on quantifying PET images of lung cancer. *J Nucl Med* 2002;43(7):876–81.