Journal of Nuclear Medicine, published on May 26, 2016 as doi:10.2967/jnumed.116.173831

Accuracy and precision of partial volume correction in oncological PET/CT studies

Cysouw MCF[1], Kramer GM[1], Hoekstra OS[1], Frings V[1], de Langen AJ[2], Smit EF[2,3], Van den Eertwegh AJM[4], Oprea-Lager DE[1], Boellaard R[1,5]

Departments of [1]Radiology and Nuclear Medicine, VU University Medical Centre, Amsterdam, The Netherlands, [2]Pulmonary Diseases, VU University Medical Centre, Amsterdam, The Netherlands, [3]Thoracic Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands, [4]Medical Oncology, VU University Medical Centre, Amsterdam, The Netherlands, [5]Nuclear Medicine & Molecular Imaging, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands.

First author: Matthijs Cysouw

Department of Radiology and Nuclear Medicine

VU University Medical Centre

PO-Box 7057, 1007MB Amsterdam, Netherlands

Email:m.cysouw@vumc.nl, Telephone:+31-20-4444214

Corresponding author: Ronald Boellaard

Department of Nuclear Medicine & Molecular Imaging

University of Groningen

University Medical Centre Groningen

PO-Box 30001, 9700RB Groningen, Netherlands

E-mail:r.boellaard@umcg.nl, Telephone:+31-50-3613471

Word-count: 5183

Running title: PVE correction in oncological PET/CT

ABSTRACT

Accurate quantification of tracer uptake in small tumors using positron emission tomography (PET) is hampered by partial volume effects as well as by methods of volume of interest (VOI) delineation. This study aimed to investigate the effect of partial volume correction (PVC) combined with several VOI methods on accuracy and precision of quantitative PET. **METHODS:** Four image-based PVC methods and resolution modeling (applied as PVC) were used in combination with several commonly used VOI methods. Performance was evaluated using simulations, phantom experiments, and clinical repeatability studies. Simulations were based on a whole-body ^{18}F-fluorodeoxyglucose-PET scan in which differently sized spheres were placed in lung and mediastinum. National Electrical Manufacturers Association NU2 Quality phantom was used for the experiments. Repeatability data consisted of a ^{18}F-fluorodeoxyglucose-PET/CT study in 11 advanced non-small cell lung cancer patients and a ^{18}F-fluoromethylcholine-PET/CT study in 12 metastatic prostate cancer patients. **RESULTS:** Phantom data demonstrated PVC is strongly affected by the applied resolution kernel, with accuracy differing ~20-50% between full-width at half-maximum setting of 5.0 and 7.5mm. For all PVC methods, large differences in accuracy were seen between VOI methods. Additionally, variable sensitivity of image-based PVC methods to volumetric accuracy of VOI methods was observed. For most PVC methods, accuracy was strongly affected by >2.5mm misalignment of simulated true VOI. When using most optimal VOI method per PVC method, high accuracy could be achieved. For example, resolution modeling for mediastinal lesions and iterative deconvolution for lung lesions were 99±1.5% and 99±0.9% accurate, respectively, for spheres ≥15mm in diameter. Precision worsened slightly for resolution modeling, and to a larger extent for some image-based PVC methods. Uncertainties in delineation propagated into uncertainties in PVC performance, as confirmed in clinical data. **CONCLUSION:** Accuracy and precision of PVC methods tested depended strongly on VOI method, resolution settings, contrast, and spatial alignment of VOI. PVC has the potential to improve accuracy of tracer uptake assessment substantially, provided robust and accurate VOI methods become available. Commonly used delineation methods may not be adequate for this purpose.

Keywords: Positron-emission tomography, partial volume correction, resolution modeling, delineation, oncology

INTRODUCTION

Quantitative positron-emission tomography (PET) provides clinical oncology with a powerful tool for diagnosis, staging, restaging, and response monitoring (*1,2*). To allow for appropriate quantification of radioactive tracer uptake, PET data needs to be corrected for several physical effects, such as decay, scatter, random coincidences, and attenuation. An effect not regularly corrected for, but with major impact on PET accuracy in small tumors, is the partial volume effect (PVE) (*3*).

PVE originates from the finite spatial resolution of the PET scanner, described in the Point Spread Function (PSF), and the tissue fraction effect (*4*). In hot lesions, PVE causes a net spill-out of activity into background, leading to considerable underestimation of the measured activity concentration (AC) (*3-6*). Whereas clinical application of partial volume correction (PVC) has led to contradictory results to date (*7*), both accurate and precise PVC methods may have significant clinical impact and substantially change quantitative reads (*8*).

Many PVC methods have been developed (*4,7,9*), e.g. the recovery coefficient (RC) method (*5,6,10*), the geometric transfer matrix (GTM) (*11*), the Müller-Gärtner method (*12*), and iterative deconvolution (IDC) (*13,14*). However, each method has its limitations, and new methodology is still being developed. Some are adaptations of the RC-method (*15,16*), but others are more refined, e.g. resolution modeling (*17,18*), adaptations of IDC (*19-21*) and the GTM (*22,23*), and background-adapted PVC algorithms (*24*).

Besides PVE, PET accuracy is strongly affected by the applied volume of interest (VOI) method, noise-level and tumor-to-background ratio (TBR) (*25*). In addition, several PVC methods use pre-defined VOI boundaries to correct for PVE. Hoetjes et al. (2010) argued performance of PVC methods may benefit from exact (e.g. CT-based) VOI definition (*3*). We therefore hypothesize that PVC performance and hence PET accuracy, is strongly affected by VOI definition methodology.

Since PVC performance is a function of PVC- as well as of VOI methods and settings, their interplay might affect accuracy and precision of PVE-corrected quantitative PET metrics. In the present study, we investigate the effects of PVC methods in combined several VOI methods on a) PET accuracy and precision using phantoms and simulations, and b) repeatability of ^{18}F-fluorodeoxyglucose- (^{18}F-FDG)

and ^{18}F-fluoromethylcholine- (^{18}F-FCH) PET in advanced non-small cell lung (NSCLC) and metastatic prostate cancer (mPC) patients, respectively.

MATERIALS AND METHODS

We used phantom experiments, simulations, and clinical data to evaluate PVC performance as a function of noise, PVC settings, and VOI method (Supplemental Table 1 summarizes the analyses). Abbreviations and their definitions are provided in Table 1.

PVC methods

Four image-based PVCs were applied: iterative deconvolution Lucy-Richardson (IDC-LR) (*3,14*), background-adapted PVC (HH) (*24*) using a local (HH-LCL) and global (HH-GLBL) background region, and mask-based spill-over PVC (*3,4*) (henceforth referred to as spill-over method).

For the image-based PVC methods we optimized spatial kernel settings using phantom data, setting the Gaussian kernel at 5.0-7.5mm (0.5mm intervals).

Reconstruction-based PVC

We applied the resolution modeling (*17*) (henceforth referred to as PSF) approach as part of the reconstruction process provided by the vendor (Philips Healthcare,Cleveland,USA). Default settings were used with noise-regularization (1 PSF-iteration, 6 regularization), implemented within time-of-flight iterative reconstruction (BLOB-OS-TF).

VOI methods

The following threshold-based VOI methods (in-house developed software (*26*)) were applied to all data: 42% and 50% of the maximal voxel value (42MAX and 50MAX, respectively), 42% and 50% of the maximal voxel value adapted for local background uptake (A42MAX and A50MAX, respectively), 50% and 70% of the peak value (i.e. average value of a 12mm sphere positioned to yield the highest value) adapted for local background uptake (A50PEAK and A70PEAK, respectively), and iteratively defined

background-adapted relative threshold level (RTL), using the system PSF (*27*). In simulations we also used the true sphere volume as VOI.

Phantom experiments

We used a National Electrical Manufacturers Association NU2 Quality phantom to perform spatial resolution kernel calibration for image-based PVC methods. The phantom contained six spheres with diameters ranging 10 to 37mm. Spheres and background were filled with ^{18}F-FDG-solutions of 12.38 and 1.46kBq/mL, respectively. A 30min scan was performed on an Ingenuity TF PET/CT (Philips Healthcare,Cleveland,USA). Reconstruction was performed using BLOB-OS-TF with and without PSF-reconstruction.

Simulations (*25*)

A mathematical phantom was derived from a ^{18}F-FDG whole-body scan.. Next, 10-40mm diameter spheres (5mm intervals) were placed within mediastinum and lung. Voxel values within spheres were set to 10kBq/mL, providing local TBRs of ~6.7 and ~3.3 for lung and mediastinum, respectively.

Using forward projection, we generated noise-free sinograms. In addition, we added noise to the sinograms using Poisson statistics simulating 3 noise-levels, corresponding to data collected for 4, 3, and 2 minutes per bed position, as typical for clinical practice. Noiseless images and images corresponding to data collected for 4, 3, and 2 minutes, had liver uptake coefficients of variation of 6.2%, 13.2%, 13.6% and 18.2%, respectively (as determined by a 3cm spherical VOI placed in the right liver lobe). For each combination of sphere size and noise-level, 10 sinograms were generated (except for noiseless sinograms).

Images were reconstructed using ordered-subset expectation-maximization, with and without PSF-reconstruction, and were post-smoothed with a 5mm Gaussian filter. The number of iterations (6) and subsets (16) was set such to assure a minimal level of convergence and to avoid limited contrast recovery. In this way partial volume effects are (mainly) affected by the spatial resolution and voxel size.

Clinical data

Clinical repeatability data consisted of a ^{18}F-FDG-PET/CT study (*28*) in 11 advanced NSCLC patients and a ^{18}F-FCH-PET/CT study (*29*) in 12 mPC patients. At time of PET, patients received no treatment. Both studies were approved by the Medical Ethical Committee of the VU Medical Centre and patients gave informed consent for participation.

Patients were scanned using a Gemini TF-64 PET/CT scanner (Philips Healthcare,Cleveland,USA). Patients fasted for 6 and 4 hours before ^{18}F-FDG- and ^{18}F-FCH-PET, respectively. PET/CT scans were acquired at 60 and 40min after injection of 185MBq ^{18}F-FDG and 200MBq ^{18}F-FCH, respectively. Images were reconstructed using BLOB-OS-TF with 3 iterations and 33 subsets, with and without PSF-reconstruction. All data were corrected for decay, scatter, random coincidences, and attenuation.

PVC performance metrics

For phantom experiment and simulations, accuracy was calculated using RCs defined as:

$$\mathrm{RC} = \frac{\mathrm{ACmeasured}}{\mathrm{ACtrue}}, \qquad \text{(Eq. 1)}$$

[$AC_{measured}$= measured mean activity concentration in (Bq/mL), AC_{true}= true or simulated activity concentration in (Bq/mL)]. Bias was calculated as:

$$\mathrm{Bias} = \frac{(\mathrm{ACmeasured} - \mathrm{ACtrue})}{\mathrm{ACtrue}}. \qquad \text{(Eq. 2)}$$

For volumetric accuracy, RCs and bias were calculated in the same manner (volumes [mL] instead of ACs).

AC-ratios were defined as:

$$\mathrm{Ratio} = \frac{\mathrm{ACpvc}}{\mathrm{ACuncorrected}}, \qquad \text{(Eq. 3)}$$

[AC_{pvc}= mean AC with PVC, and $AC_{uncorrected}$= mean AC without PVC].

SUV_{mean}, normalized to bodyweight, was calculated for clinical data. Total lesion glycolysis (TLG) was calculated as [SUV_{mean}*metabolically active lesion volume (mL)]. All metrics were derived without and with PVC.

Statistical analysis

Normality of SUVs and TLGs was assessed with Shapiro-Wilks test. Intraclass correlation coefficients (ICC; two-way mixed model with an absolute agreement definition) were calculated for each combination of VOI and PVC method. For non-normal distribution, log-transformed SUV_{mean} and TLG were used to calculate ICCs. Analyses were performed using SPSS® Statistics v.22.0 (IBM®,Armonk,NY,USA).

RESULTS

Phantom experiments

Image-based PVC methods required optimization of the applied spatial kernel, per VOI method. For all VOI methods, IDC-LR and spill-over method RCs increased ~0.2-0.5 from full-width at half-maximum (FWHM) of 5.0 to 7.5mm. Accuracy of HH-GLBL was not affected by FWHM, and HH-LCL RCs only increased 0.03-0.3 between 5.0-5.5, 6.0 and 6.5-7.5 mm. Large differences in accuracy were seen for all image-based PVC methods between different VOI methods applied, mainly for 13 and 17mm spheres (typically yielding overcorrection). Even for the optimal FWHMs, PVC methods still fail for the 10mm sphere.

Volumetric accuracy was better in non-PSF-reconstructions for the 17-37mm spheres, besides d=37mm delineated with 42MAX (Figs.1A and 1B). Notably, 10 and 13mm spheres were delineated more accurately using 42MAX, 50MAX, and RTL in PSF-reconstruction. Smallest differences in volumetric accuracy were seen for background-adapted VOIs. Volumes of PET-based VOIs generated on PSF-reconstructed images were smaller compared to non-PSF (Fig.1C). No difference in volume was found for 10mm sphere, delineated with A42MAX and A50MAX, whereas delineation with A50PEAK and A70PEAK provided negligible larger volumes (+0.064mL).

Simulations

Large differences in PVC performance were seen between all VOI methods (Supplemental Figs.1 and 2). Optimal combinations of PVC and VOI methods are shown in Figure 2 and Table 2. Generally,

RCs in lung were lower compared to mediastinum. For spheres ≥15mm, PSF-reconstruction with A70PEAK in mediastinum and IDC-LR with A42MAX in lung yielded highest accuracies: 99±1.5% and 99±0.9%, respectively. HH-GLBL and HH-LCL considerably overcorrected true AC when using A42MAX, A50MAX, A50PEAK, or A70PEAK. HH-LCL and the spill-over method both performed excellently (100% accuracy overall) when using true (simulated) VOI, and were within 10% accurate when using RTL (≥15mm). Figure 3 demonstrates percentage bias in sphere volumes in lung. We found a strong relationship between underestimation of true volume and overcorrection of AC RCs for HH-GLBL and HH-LCL (RCs≤3 and ≤2.25, respectively). The spill-over method was moderately affected (RCs≤1.33) and IDC-LR and PSF-reconstruction RCs were not significantly correlated with negative bias in volume (RCs:0.6-1.05). There was a moderate inverse correlation between RCs and positive bias in volume for all methods besides PSF-reconstruction. Similar correlations were observed for mediastinal spheres, but bias in volume, and thus in AC, was larger.

Figure 4 illustrates AC-ratios as function of misalignment of true VOI. HH-GLBL demonstrated only a slight decrease in RC when misaligned ≥10mm, and was >94% accurate in lung, while overcorrecting up to 20% in mediastinum. HH-LCL was 98-100% accurate when misaligned<5mm, in lung and mediastinum. The spill-over method performed slightly worse than HH-LCL. IDC-LR and PSF-reconstruction performed poorest when using true VOI, but were equally sensitive to misalignment as HH-LCL and spill-over method. Similar trends were obtained for all sphere sizes, but sensitivity to misalignment increased with decreasing sphere size.

There was a positive association between noise-level and RCs, RCs becoming larger as VOI thresholds increased. AC-ratios of mediastinal spheres increased with noise-level for the spill-over method, HH-GLBL, and HH-LCL when using background-adapted VOIs, while in lung these ratios were equal for all noise-levels (Fig.5; similar but inverse trends were observed for volumes). In contrast, results from noise-less images were similar to the highest noise level. With true simulated VOI, RCs were similar at all noise-levels, both in mediastinum and lung.

Impact of PVC on precision for spheres in lung is illustrated in Figure 6. In general, PVC increased SDs, most pronounced for HH-GLBL and HH–LCL. Precision depended on the combination of

applied VOI and PVC methods. When using the true volume, SDs were smallest, suggesting uncertainties in PET based VOI performance propagate into uncertainties in PVC. PET-based VOIs generally resulted in larger SDs in mediastinum compared to lung.

Clinical data

Table 3 describes the clinical cohorts. Feasibility (i.e. percentage of lesions successfully delineated) of VOI methods was better in PSF than non-PSF reconstructed images (Supplemental Tables 2 and 3). HH-GLBL failed, providing negative ACs, in 2.4% and 2.8% of lesions in the ^{18}F-FDG- and ^{18}F-FCH-PET cohorts, respectively.

ICCs were calculated to quantify, and facilitate comparison between, repeatability of SUVmean and TLG (Fig.7). Repeatability of uncorrected SUVmean was best (ICC~0.97-0.98), with comparable ICCs for SUVmean of IDC-LR, spill-over, and PSF-reconstruction. ICCs of HH-LCL were slightly lower, depending on VOI method. HH-GLBL demonstrated worst repeatability of SUVmean, for all VOI methods (ICC~0.77-0.83). ICCs of SUVmean were comparable between VOI methods, except for HH-GLBL and –LCL. All PVE-corrected TLGs (pTLG) had ICCs almost equal to uncorrected TLG, except for PSF-reconstruction. Similar trends in ICCs between VOI methods are seen for volumes (Supplemental Table 4) and their respective TLGs. Overall, ICCs were lower for the ^{18}F-FCH-PET cohort compared to the ^{18}F-FDG-PET cohort.

DISCUSSION

PVE introduced substantial error in quantification of tracer uptake in lesions with diameters <25 and <30mm in mediastinum and lung, respectively. Current guidelines of response evaluation with PET do not include PVC (*2,30,31*). PERCIST (*2*) advises to only assess tumors >2cm at baseline, to avoid overestimation of metabolic response with shrinkage during therapy, whereas EORTC (*31*) merely recommends documentation of tumor size in relation to scanner resolution. However, it is unclear how lesion selection strategies in metastasized patients affect clinical performance of imaging biomarkers of response, especially in case of targeted therapy with potential heterogeneous inter- and/or intra-lesional target expression. Of note, median volumes of lesions in the FDG- and FCH-cohorts corresponded to

d=20-22mm equivolumetric spheres, being well within the range of lesions affected by PVE. PVE may also compromise diagnosis or prognosis when using SUV-based thresholds in small tumors (*7*), even when guidelines for scanner calibration, image acquisition, and reconstruction are implemented (*32*). Taken together, we estimate that appropriate PVC might become of greater clinical importance than considered so far. Our results demonstrate PVC methods have potential for accurate and precise PVE-correction. However, PVC performance heavily depends on the applied VOI method and factors influencing VOI method performance, such as lesion size, TBR, noise, and spatial alignment. We recommend future research into PVC to focus on development of robust and standardized PVC/VOI combinations and their clinical impact, using valid clinical reference parameters for the latter.

Phantom and simulation studies

Adjustment of FWHM for the image-based PVC methods had a major effect on performance of most methods. IDC-LR and the spill-over method substantially differed in accuracy between different FWHM settings, RCs increasing with FWHM. This is most likely because both methods directly use the applied FWHM for PVC, warranting accurate calibration. Performance of HH-GLBL was equal for all settings, whereas HH-LCL only differed for some FWHM settings. For the latter method, it is advised not to underestimate the FWHM, ensuring the entire spill-out of signal is contained within the spill-out region, in accordance with results from Hofheinz *et al*. (*23*).

VOIs were generated on both non-PSF and PSF-reconstructed images. Therefore, differences in volumes and volumetric accuracy between non-PSF and PSF reconstruction were assessed. In general, PSF-reconstructions resulted in smaller VOIs, most likely due to improved TBR and enhanced edges. However, volumetric accuracy was worse, apart for some VOIs generated on the smallest spheres.

In simulations, performance of PVC differed between VOI methods. RCs tended to be lower for spheres in lung versus mediastinum, which in case of the simulated uniform spheres can be explained by larger PVE in lung due to higher TBR. Without PVC, ACs obtained with A70PEAK proved most accurate. This VOI results in very small volumes, including only the core of spheres and thereby bypassing the PVE, which mainly occurs at lesion edges. PSF-reconstruction increased accuracy 2-16%, most pronounced for smallest spheres. Even though in lung improvement in accuracy was moderate, VOI methods tended to

have higher feasibility on PSF-reconstructed images (Supplemental Table 2 and 3). Whereas Teo et al. (2007) found IDC to perform optimally with a 80%MAX VOI in phantom (*12*), our simulation study suggested IDC-LR performs excellently using background-adapted VOIs with a fixed threshold. HH-GLBL and -LCL were very sensitive to underestimation of volume, probably due to inclusion of sphere AC within the spill-out region, thus substantially overestimating true AC (Fig.3). Overall, HH-LCL performs better than HH-GLBL, most likely because it can account for heterogeneity of activity within background. The spill-over method had excellent performance using RTL and A42MAX, with accuracies of 104±2.0% and 105±3.3% for spheres d≥15mm in mediastinum and lung, respectively. Notably, when using the true VOI, the spill-over method, HH-LCL and HH-GLBL performed excellently (accuracy ~100%). This is understandable since theoretically, with homogeneous uptake, accuracy should be 100% when these methods are applied using perfect tumor boundaries and true FWHM. In addition, the true VOI demonstrated highest precision. However, in clinical settings perfectly aligned CT- and PET-images are not realistic due to patient movement and breathing, and CT-based anatomical volume might comprise non-viable tumor tissue. Thus, application of CT-based VOIs when using HH-LCL or the spill-over method might result in less accurate results due to their sensitivity to misalignment of VOI (Fig.4) and inclusion of non-viable tumor tissue. HH-GLBL was unaffected by misalignment for spheres d≥20 mm, most likely due to its large background region. However, some dependency on TBR was seen using the true VOI (20% overcorrection in the mediastinum).

PVC methods directly using VOI boundaries (i.e. spill-over method, HH-GLBL, and HH-LCL) differed considerably between noise-levels for mediastinal spheres, when using background-adapted VOIs. In contrast, similar performance of all PVC methods at each noise-level was observed in lung, where high contrast resulted in very similar VOI delineation between noise-levels. Thus, at low contrast background-adapted VOIs become unreliable, propagating into unreliable performance of PVC methods sensitive to volumetric accuracy.

PVC negatively affected precision to a small extent (Fig.6), besides for HH-GLBL and –LCL where SDs increases considerably. Overall, A50PEAK seemed most precise when applying image-based PVC, most likely due to using peak values, which are less sensitive to noise than maximal voxel values.

Clinical studies

Previous research showed PVC to have no significant effect on tracer uptake repeatability, but only one PET-based VOI (A50MAX) was used (*3*). In the present study, repeatability of SUVmean based on different delineation methods was consistent after PVC, with comparable ICCs between VOI methods, except for HH-LCL and HH-GLBL. The latter methods demonstrated large differences in ICC between different VOI methods, with broader confidence intervals overall, illustrating worsened precision for these methods, in accordance with precisions observed in the simulations.

Erlandsson *et al.* (2012) proposed using pTLG in clinical settings, since uncorrected SUV might yield important volumetric information eliminated in SUV_{PVC} (*7*). Our results demonstrate that pTLG and uncorrected TLG have similar repeatability characteristics (Fig.6). Among the PVC methods ICCs were very similar, except those obtained from PSF-reconstruction. This difference is most likely caused by trends in ICCs of TLG between VOI methods being similar to trends in ICCs of VOI volumes (Supplemental Table 4), emphasizing the importance of volumetric information on precision in PVC. For HH-GLBL and –LCL applied with their optimal VOI methods, pTLG might be suitable to acquire data with optimal accuracy and precision.

Limitations

In phantoms and simulations, lesions were spherical with homogeneous uptake. In reality, tumors rarely have spherical dimensions let alone homogeneous uptake. Yet, similar trends regarding the performance of PVC were observed as in the clinical data, and the simulations allowed us to gain insight in PVC method performance with the advantage of known (simulated) truth. In addition, motion blurring due to breathing and peristaltic movement may lead to significant measurement errors. To mitigate effects of e.g. breathing, respiratory gated PET/CT studies may be performed (*33*). Respiratory gated PET/CT is, however, not yet routinely applied in all centers. Yet, the reader should be aware that, besides PVE, motion also has a negative impact on the PET quantitative accuracy.

CONCLUSION

We investigated performance of PVC as function of VOI delineation, resolution settings, TBR, and noise. We observed that accuracy of quantitative oncology PET studies may improve to a large extent using the PVC methods investigated while maintaining good precision. However, PVC performance heavily depends on the VOI method, and differs considerably between lesions in lung and mediastinum. For most image-based PVC methods ≤2.5mm error in spatial alignment of VOI and tumor is critical. Some methods directly using predefined VOIs to correct PVE are less dependent on correct alignment, but are more sensitive to volumetric accuracy. Furthermore, uncertainties in PET-based VOIs propagate into precision of PVC performance. PVC can substantially improve accuracy of quantification of tumors measuring 15-25mm in oncological PET studies. However, without highly accurate and precise VOI methods, PVC may even worsen accuracy and precision. With contemporary scanners and reconstructions, quantifying uptake of tumors <15mm in diameter is still not recommended.

CONFLICT OF INTEREST

The authors declare having no conflicts of interest.

REFERENCES

1. Hoekstra CJ, Paglianiti I, Hoekstra OS, 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;27:731-743.
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-150S.
3. Hoetjes NJ, Van Velden FH, Hoekstra OS, et al. Partial volume correction strategies for quantitative FDG PET in oncology. *Eur J Nucl Med Mol Imaging*. 2010;37:1679-1687.
4. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. *J Nucl Med*. 2007;48:932-945.
5. Hoffman EJ, Huang SC, Phelps ME. Quantitation in positron emission computed tomography: 1. effect of object size. *J Comput Assist Tomogr*. 1979;3:299-308.
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;27:161-169.
7. Erlandsson K, Buvat I, Pretorius PH, Thomas BA, Hutton BF. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. *Phys Med Biol*. 2012;57:R119-159.
8. Kuhnert G, Boellaard R, Sterzer S, et al. Impact of PET/CT image reconstruction methods and liver uptake normalization strategies on quantitative image analysis. *Eur J Nucl Med Mol Imaging*. 2016;43:249-258.
9. Rousset OG, Rahmim A, Alavi A, Zaidi H. Partial volume correction strategies in PET. *PET Clin*. 2007;2:235–249
10. Avril N, Dose J, Jänicke F, et al. Metabolic characterization of breast tumors with positron emission tomography using F-18 fluorodeoxyglucose*. J Clin Oncol*. 1996;14:1848-1857.
11. Rousset OG, Ma Y, Evans AC. Correction for partial volume effects in PET: principle and validation. *J Nucl Med*. 1998;39:904-911.

12. Müller-Gärtner HW, Links JM, Prince JL, 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;12:571-583.
13. Teo BK, Seo Y, Bacharach SL, et al. Partial-volume correction in PET: validation of an iterative postreconstruction method with phantom and patient data. *J Nucl Med*. 2007;48:802-810.
14. Tohka J, Reilhac A. Deconvolution-based partial volume correction in Raclopride-PET and Monte Carlo comparison to MR-based method. *Neuroimage*. 2008;39:1570-1584.
15. Gallivanone F, Canevari C, Gianolli L, et al. A partial volume effect correction tailored for 18F-FDG-PET oncological studies. *Biomed Res Int*. 2013;2013:780458.
16. Krempser AR, Ichinose RM, Miranda de sá AM, et al. Recovery coefficients determination for partial volume effect correction in oncological PET/CT images considering the effect of activity outside the field of view. *Ann Nucl Med*. 2013;27:924-930.
17. Rahmim A, Qi J, Sossi V. Resolution modeling in PET imaging: theory, practice, benefits, and pitfalls. *Med Phys*. 2013;40:064301.
18. Wallstén E, Axelsson J, Sundström T, Riklund K, Larsson A. Subcentimeter tumor lesion delineation for high-resolution 18F-FDG PET images: optimizing correction for partial-volume effects. *J Nucl Med Technol*. 2013;41:85-91.
19. Boussion N, Cheze le rest C, Hatt M, Visvikis D. Incorporation of wavelet-based denoising in iterative deconvolution for partial volume correction in whole-body PET imaging. *Eur J Nucl Med Mol Imaging*. 2009;36:1064-1075.
20. Bhatt R, Adjouadi M, Goryawala M, Gulec SA, Mcgoron AJ. An algorithm for PET tumor volume and activity quantification: without specifying camera's point spread function (PSF). *Med Phys*. 2012;39:4187-4202.
21. Merlin T, Visvikis D, Fernandez P, Lamare F. A novel partial volume effects correction technique integrating deconvolution associated with denoising within an iterative PET image reconstruction. *Med Phys*. 2015;42:804-819.

22. Sattarivand M, Kusano M, Poon I, Caldwell C. Symmetric geometric transfer matrix partial volume correction for PET imaging: principle, validation and robustness. *Phys Med Biol*. 2012;57:7101-7116.
23. Sattarivand M, Armstrong J, Szilagyi GM, Kusano M, Poon I, Caldwell C. Region-Based Partial Volume Correction Techniques for PET Imaging: Sinogram Implementation and Robustness. *Int J Mol Imaging*. 2013;2013:435959.
24. Hofheinz F, Langner J, Petr J, et al. A method for model-free partial volume correction in oncological PET. *EJNMMI Res*. 2012;2:16.
25. 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;45:1519-1527.
26. Frings V, Van Velden FH, Velasquez LM, et al. Repeatability of metabolically active tumor volume measurements with FDG PET/CT in advanced gastrointestinal malignancies: a multicenter study. *Radiology*. 2014;273:539-548.
27. Van Dalen JA, Hoffmann AL, Dicken V, et al. A novel iterative method for lesion delineation and volumetric quantification with FDG PET. *Nucl Med Commun*. 2007;28:485-493.
28. Kramer GM, Frings V, Hoetjes N et al. Repeatability of quantitative uptake measures of whole body [18F]FDG PET/CT in NSCLC patients.[Abstract]. *J Nucl Med*. 2015;56(suppl 3):1379.
29. Oprea-Lager DE, Kramer G, van de Ven P, et al. Repeatability of quantitative 18F-fluoromethylcholine PET/CT studies in prostate cancer. *J Nucl Med*. December 23, 2015.[Epub ahead of print].
30. Young H, Baum R, Cremerius U, et al. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer*. 1999;35:1773-1782.
31. Shankar LK, Hoffman JM, Bacharach S, et al. Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials. *J Nucl Med*. 2006;47:1059-1066.

32. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging*. 2015;42:328-354.
33. Daouk J, Fin L, Bailly P, Meyer ME. Respiratory-gated positron emission tomography and breath-hold computed tomography coupling to reduce the influence of respiratory motion: methodology and feasibility. *Acta Radiol*. 2009;50:144-155.

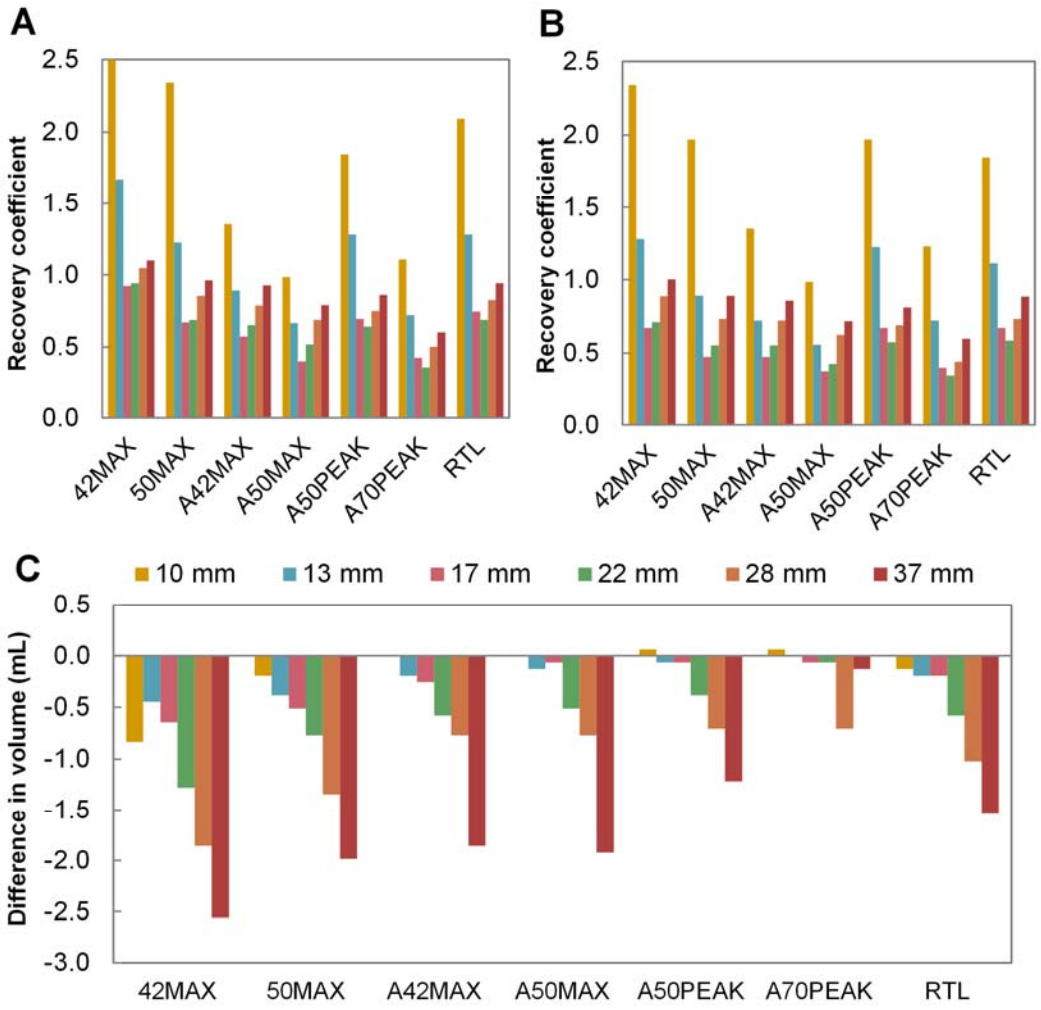

Fig. 1) Volume RCs of non-PSF- (A) and PSF-reconstructed images (B) per VOI method and sphere size, and differences in PET-based volumes between non-PSF- and PSF-reconstructed images (C). Negative volume differences indicate smaller PSF-reconstruction based volumes compared to non-PSF-reconstruction. Sizes in key indicate sphere diameters. Sphere d=10mm delineated with 42MAX had a 3.9 RC in non-PSF-reconstruction.

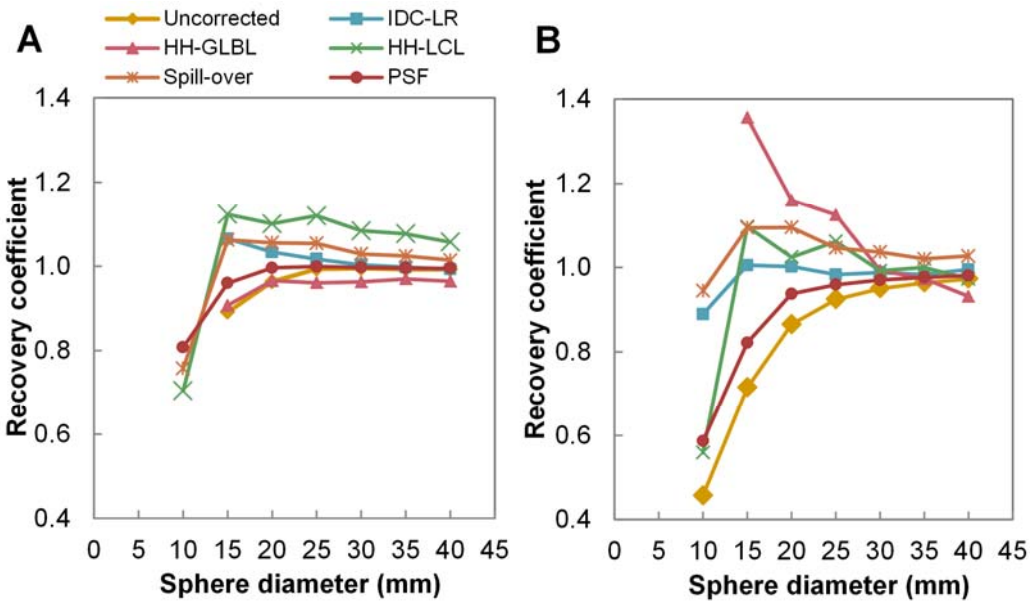

Fig 2) AC RCs as function of sphere diameter for all PVC methods, and uncorrected data, with their optimal PET-based VOI method (Table 2) for spheres in mediastinum(A) and lung(B). Missing values are due to delineation failure.

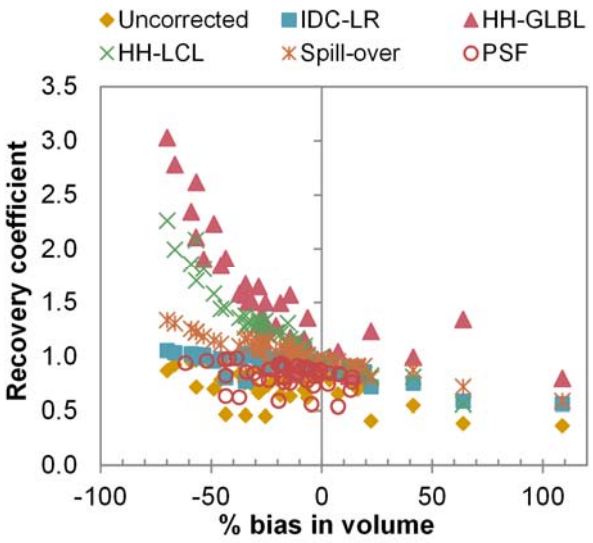

Fig. 3) AC RCs as function of volumetric bias(%). Shown are results of all VOI methods for spheres in lung (noise-less images).

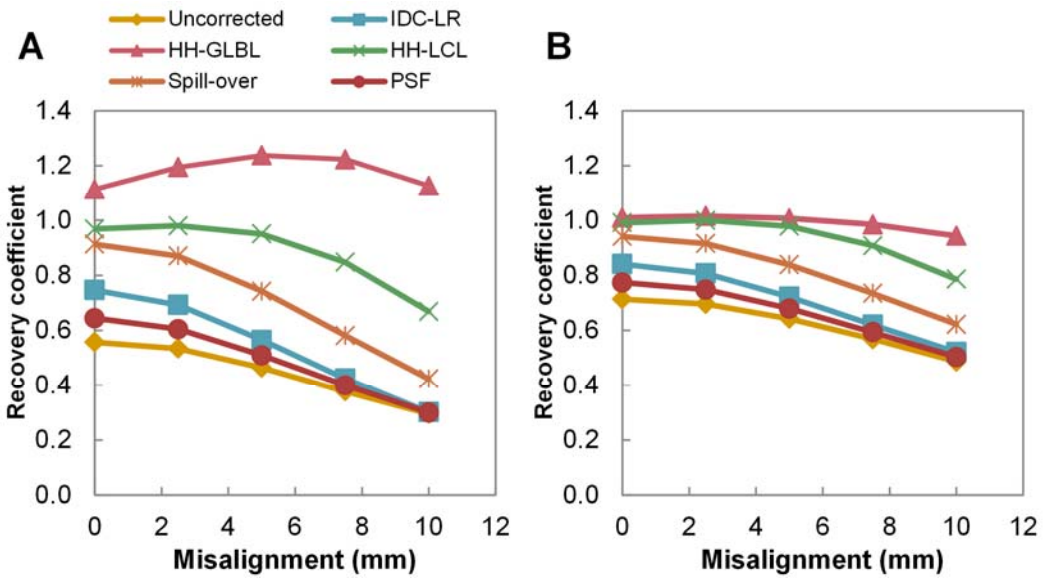

Fig. 4) AC RCs as function of misalignment of true VOI (mm). Shown are results from spheres with d=15(A) and 25mm(B) in lung (noise-less images).

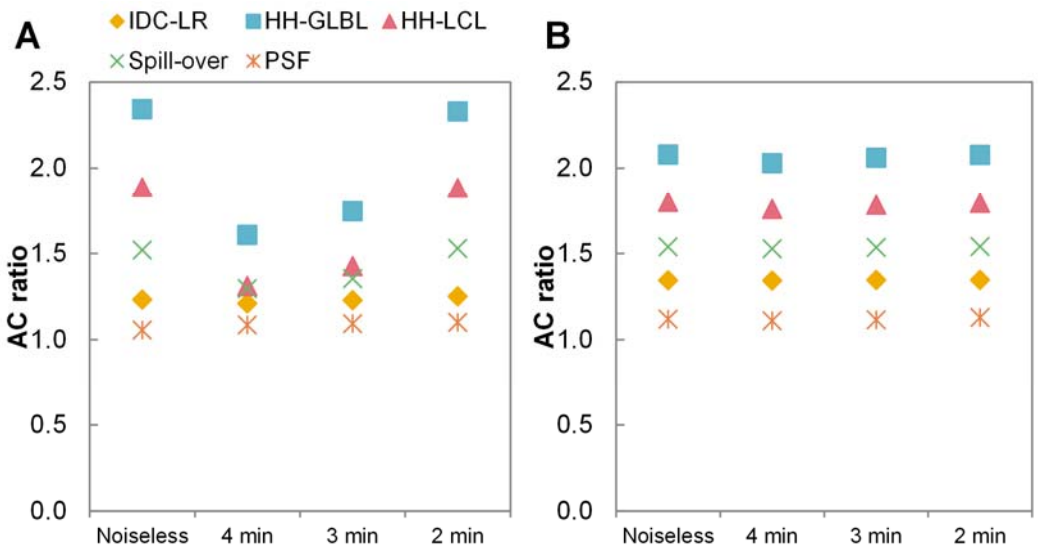

Fig. 5) AC-ratios as function of simulated acquisition time (thus noise-level). Shown are the results from A50PEAK for sphere d=20mm (corresponding to median volumes of ^{18}F-FDG- and ^{18}F-FCH-PET cohorts delineated with A50PEAK) in mediastinum(A) and lung(B), respectively.

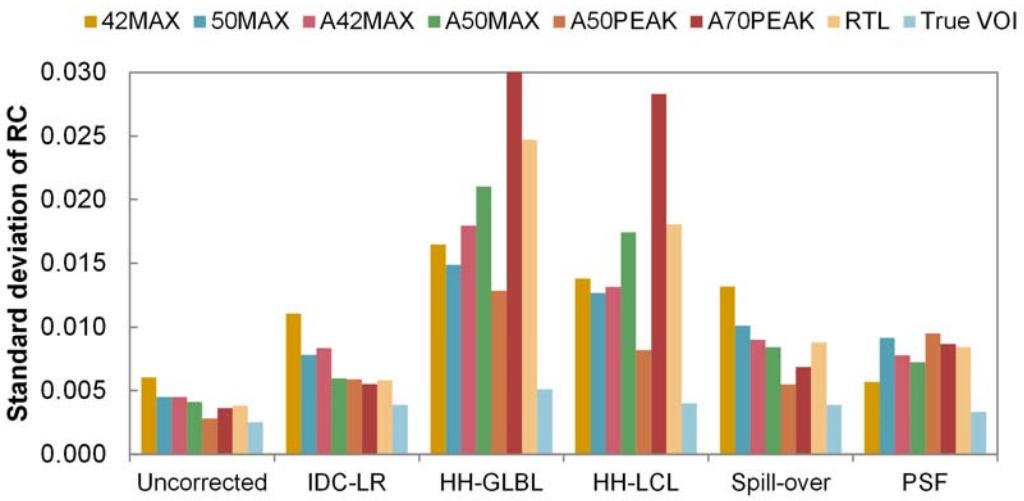

Fig. 6) SDs of RCs for all combinations of VOI+PVC method. Shown are results of spheres in lung, with d=20mm (corresponding to median volumes of ^{18}F-FDG- and ^{18}F-FCH-PET cohorts delineated with A50PEAK). Y-axis scaled for visual interpretation; SD of HH-GLBL using A70PEAK was 0.049.

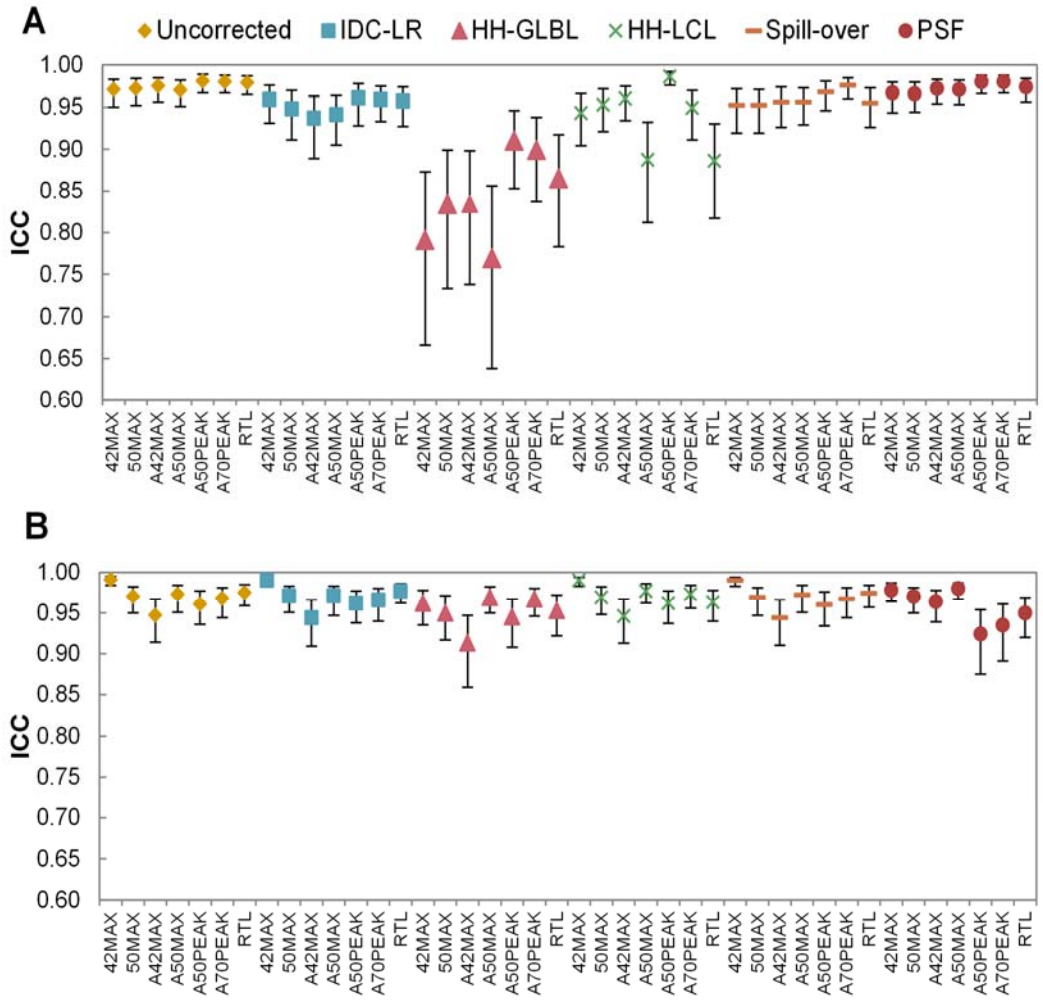

Fig. 7) ICCs of SUVmean (A) and TLG (B) of all combinations of VOI+PVC method. Shown are results of the ^{18}F-FDG-PET cohort. Error-bars represent 95%-CIs. Similar results were obtained in the ^{18}F-FCH-PET cohort (Supplemental Fig.3).

Abbreviation		Definition
^{18}F-FDG		^{18}F-fluorodeoxyglucose
^{18}F-FCH		^{18}F-fluoromethylcholine
PVE		Partial volume effect
PVC		Partial volume correction
PVC methods:		
	IDC-LR	Iterative deconvolution Lucy-Richardson
	HH-GLBL	Global background-adapted PVC
	HH-LCL	Local background-adapted PVC
	Spill-over	Mask-based spill-over PVC
	PSF-reconstruction	Point Spread Function reconstruction
VOI		Volume of interest
VOI methods:		
	42MAX	42% of maximal voxel
	50MAX	50% of maximal voxel
	A42MAX	42% of maximal voxel + background
	A50MAX	50% of maximal voxel + background
	A50PEAK	50% of peak voxel + background
	A70PEAK	70% of peak voxel + background
	RTL	Relative Threshold Level
RC		Recovery coefficient
AC		Activity concentration
SUVmean		Mean Standardized Uptake Value
TLG		Total lesion glycolysis
ICC		Intraclass correlation coefficient
FWHM		Full-width at half-maximum
BLOB-OS-TF		Iterative time-of-flight reconstruction
NSCLC		Non-small cell lung cancer
mPC		Metastatic prostate cancer

Table 1) Abbreviations and their definitions.

	Uncorrected	IDC-LR	HH-GLBL	HH-LCL	Spill-over	PSF
VOI method:	A70PEAK	A50MAX	50MAX	RTL	RTL	A70PEAK
(Mediastinum)	(97±4.1)	(102±2.7)	(96±2.4)	(109±2.6)	(104±2.0)	(99±1.5)
VOI method:	A70PEAK	A42MAX	42MAX	50MAX	A42MAX	A70PEAK
(Lung)	(90±9.8)	(99±0.9)	(109±15.8)	(103±4.7)	(105±3.3)	(94±6.0)

Table 2) Optimal PET-based VOI methods for each PVC method, in lung and mediastinum. Optimal combinations as determined in simulations, on noise-less images. Mean accuracy (percentages±SD) of spheres ≥15mm in parentheses.

		^{18}F-FDG-PET [28]	^{18}F-FCH-PET [29]
Type		NSCLC	mPC (n=4 castration-resistant)
No. of patients		11	12
No. of lesions		70	67
Age (mean±SD, years)		60±7	64±8
Gender		7 male, 4 female	12 male
Lesion localization		16 intrapulmonary, 54 extrapulmonary	44 bone metastases 23 lymph node metastases
Median volume (mL)	**Non-PSF**	3.94 (IQR 10.85)	5.76 (IQR 8.64)
	PSF	3.90 (IQR 20.10)	5.28 (IQR 7.92)

Table 3) Patient characteristics. Median volumes determined with A50PEAK, the most accurate VOI method as determined in phantom experiment, on baseline.