Clinical Relevance of Partial-Volume Effect: Dependence on Lesion size and Shape

Tram Nguyen
and Poul Flemming Høilund-Carlsen
Dept. of Nuclear Medicine
Odense University Hospital
Odense, Denmark

Habib Zaidi
Div. of Nuclear Medicine
and Molecular Imaging
Geneva University Hospital
Geneva, Switzerland

Werner Vach
Inst. of Medical Biometry and Statistics
University of Freiburg
Freiburg, Germany

***Abstract*—This study sought to systematically assess the influence of partial-volume effect (PVE) and segmentation on PET quantitative measures, hinging on lesion geometry. How this affects, e.g., prevalent maximum standardised uptake values, SUV_{max}, and potentially impact clinical applications, e.g., response evaluation has yet to be fully discerned. From PET simulations with open-source software of variable-sized ellipsoidal lesions inserted in an anthropomorphic phantom, images of two contrasts and resolutions were generated. SUV_{max} and volumetric indices extracted with six different segmentations were compared for variability and test-retest repeatability. The study showed similar or larger shape dependent variability and lower repeatability in SUV_{max} than SUV_{mean}. Alternative volumetric indices might provide more robust measures but require better contouring than common thresholding. Thus findings suggested significant impact of PVE and segmentation in clinically relevant lesion sizes that can bias interpretation of SUV changes.**

I. Introduction

Common practice and guidelines for oncological PET/CT [1] include measurement of standardised uptake values SUV_{max} or SUV_{peak} in disease report, e.g., for diagnostics or response monitoring [2]. Partial-volume effect (PVE) is a recognised source of bias with potentially consequential impact on PET quantification [3] but is generally ignored in the clinic. For the sizes of many clinical lesions, however, PVE is expected to have a non-negligible size and shape related influence on SUV. Our aim was therefore to investigate the repercussion of this in a first systematic assessment of relations between PVE, lesion geometry and delineation method.

II. Methods

Computer simulations of clinical scans were conducted with basic ellipsoidal lesions for clear-cut, controlled comparisons.

A. Simulations

The XCAT2 digital phantom [4] was employed and modified based on a clinical patient scan. Ellipsoids of incrementally varied axes lengths L = (L_x, L_y, L_z) for clinically relevant volumes, V, were inserted in the lung area using MATLAB 9.0 (Mathworks Inc.). Dimensions were: 1) V $\approx$ 0.1 cc: L = 8.8×8.8×3.3 mm, 2.7×10.9×3.3 mm, 2.7×13.7×3.3 mm, 2.7×16.4×3.3 mm, 2.7×19.1×3.3 mm, 2) V $\approx$ 1 cc: L = 13.7×13.7×9.8 mm, 13.7×19.1×9.8 mm, 8.2×24.6×9.8 mm, 8.3×30×9.8 mm, 3) V $\approx$ 2 cc: L = 13.7×13.7×16.4 mm, 13.7×19.1×16.4 mm, 8.2×24.6×16.4 mm, 8.2×30×16.4 mm. Two contrasts with source-to-background ratios, sbr = 4:1 and 8:1, were constructed. Scans were then created by analytic simulations with the software for tomographic image reconstruction (STIR) [5], set up to emulate the GE Discovery RX tomograph geometry with a randoms-to-trues ratio $\approx$ 0.35 (total counts $\approx$ 164 M) based on clinical data and reports [6]. Image reconstructions were with matrices = 128×128 (voxel $\approx$ 0.55×0.55×0.33 cc) and 256×256 (voxel $\approx$ 0.27×0.27×0.33 cc) and a 3D ordered subset expectation maximisation (OSEM) algorithm (iterations/subsets = 3/18) and 5 mm FWHM Gaussian smoothing. Noise and test-retest evaluation were for 50 noisy realisations of each scenario.

B. PET Segmentation

Activity levels, A $\propto$ SUV, were extracted from lesion delineations inside manually placed 3D masks encompassing each ellipsoid. This included A_{max}, A_{mean}, A_{peak} (highest mean value within a 3D sphere of 1.2 cm diameter), the sum of lesion activity, A_{tot}, and mean metabolic volumetric product, $MVP_{mean} = A_{mean} \cdot V$. Segmentations for delineation were besides true boundary contours chosen from guidelines [1], [2] plus other common methods in clinical trials: Fixed thresholds at 41% (T41), 50% (T50), 70% (T70) of maximum value, 41% of background corrected maximum value, and fuzzy C-means.

III. Results

Measurement variations are shown in Fig. 1 for true edge contours with recovery coefficients as ratios between measured and true uptake levels. A_{tot} plots were similar to those of MVP_{mean} and hence not displayed. With A_{max} deviations from true values an average 61.5%–78.7% (V $\approx$ 0.1 cc), 22.1%–58.0% (V $\approx$ 1 cc) and 8.0%–49.0% (V $\approx$ 2 cc), and A_{mean} correspondingly from 62.5%–80.8%, 44.0%–70.9% and 39.7%–71.6%, A_{max} varied slightly more with lesion shape across image resolutions and contrasts (5.1%–30.3%). Test-retest variability was 2.6%–3.0% for A_{max} and 1.6%–1.8% for volumetric indices (A_{mean}, A_{tot}, MVP_{mean}). A_{peak} were only extractable in some larger lesions (V $\approx$ 2 cc) and varied from true values generally by 25.8%–41.9%.

While automatic segmentations increased variability overall relations were as described above. Failed delineations (with

978-1-5386-2282-7/17/$31.00

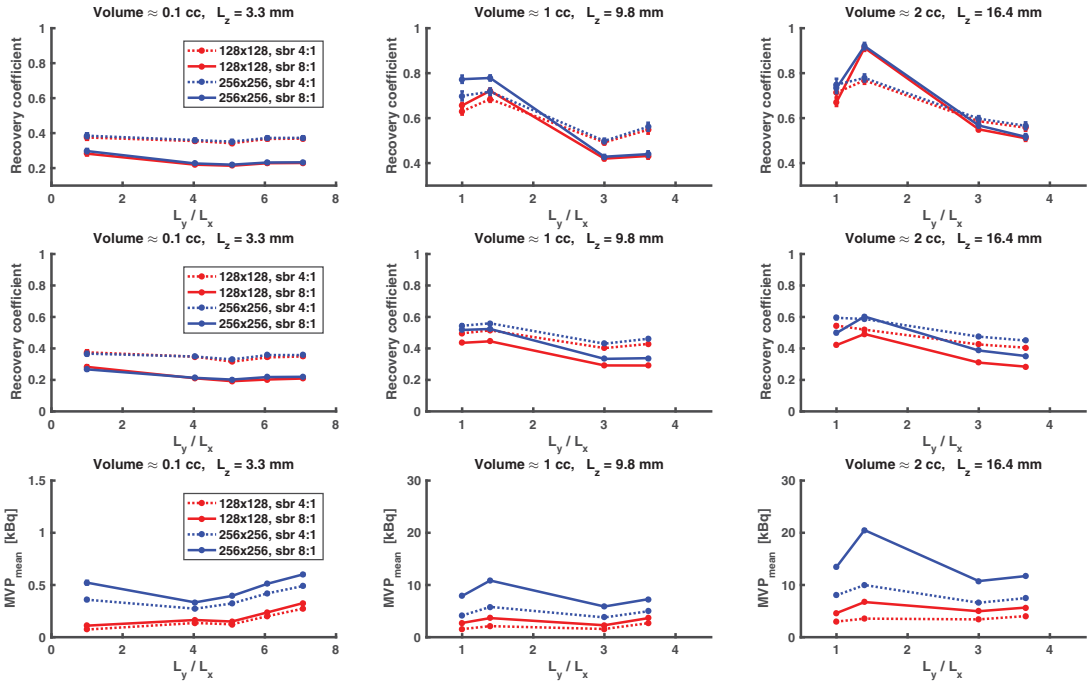

Fig. 1. Measures extracted from true lesion contours as a function of in-plane lesion axes ratios (L_y/L_x) for given through-plane axes lengths (L_z) shown for recovery coefficients = A_{max}/A_{true} (top row), A_{mean}/A_{true} (middle row), and MVP_{mean} (bottom row) for different image resolutions (matrix = 128×128 and 256×256) and contrasts (sbr = 4:1 and 8:1) demonstrating PVE impact without segmentation bias.

T41, T50, and somewhat T70) were discounted in overall accounts. Here, divergence in volume estimates of 99.8%–165.4% and test-retest delineations of 10.4%–16.0% was seen. Though A_{max} was consistent between methods, it had similar or slightly larger variability than A_{mean} with lesion shape for any given volume, both between 4.1%–9.0% at sbr = 4:1 and V $\approx$ 0.1 cc (sbr = 8:1), and 26.8%–28.5% for V $\approx$ 1–2 cc (sbr = 8:1). Test-retest variations were 2.5%–2.9% for A_{max} and 2.0%–2.3% for A_{mean}. Volumetric changes, meanwhile, impacted A_{tot} and MVP_{mean} the most. Due to this a larger dependence on contrast, lesion size and shape was seen with 26.1%–206.2% variability. Disparities on test-retest among the segmentation methods were 8.7%–14.5%.

IV. Conclusions

This precursory study confirmed a significant PVE impact on metric levels, and further supported the expected substantial relations to lesion geometry. It indicated that SUV measures, while PVE independent in small lesions as acknowledged [2], will be affected with shape changes alone in larger lesions ($\sim$ 1–2 cc). This, however, depended on contrast. Thus reduced lesion contrast, e.g., in treatment response, can greatly negate shape related variation due to concomitantly lower PVE. Results showed a lower repeatability of SUV_{max} than SUV_{mean}, which confirmed other reports [7], and here further implied that SUV_{max} is less lesion shape invariant than SUV_{mean}. Although analysis on test-retests suggested seemingly more stable volumetric indices, especially SUV_{mean} but also SUV_{tot} and MVP_{mean}, segmentation there was the dominant source of variability, calling for more robust methods than generally insufficient thresholding. More accurate simulations need to be performed for verification. Current observations, nevertheless, were consistent with known relations and provide basic insight into quantification aspects that presumably only will amplify in more complex clinical scenarios. It at least signifies non-negligible PVE impact on SUVs, which should be interpreted critically in clinical applications.

References

[1] R. Boellaard, R. Delgado-Bolton, et al.,"FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 42, pp. 328-354, Feb. 2015.

[2] R. L. Wahl, H. Jacene, Y. Kasamon, M. A. Lodge, "From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors," *J. Nucl. Med.*, vol. 50, pp. 122S-150S, May 2009.

[3] M. Soret, S. L. Bacharach, I. Buvat,"Partial-Volume Effect in PET Tumor Imaging," *J. Nucl. Med.* vol. 48, pp. 932-945, June 2007.

[4] W. P. Segars, G. Sturgeon, et al., "4D XCAT phantom for multimodality imaging research," *Med. Phys.*, vol. 37, pp. 4902–15, Sep. 2010.

[5] K. Thielemans, C. Tsoumpas, et al., "STIR: software for tomographic image reconstruction release 2," *Phys. Med. Biol.*, vol. 57, pp. 867-83, Feb. 2012.

[6] B. J. Kemp, C. Kim, et al., "NEMA NU 2-2001 Performance Measurements of an LYSO-Based PET/CT System in 2D and 3D Acquisition Modes," *J. Nucl. Med.*, vol. 47, pp. 1960-1967, Dec. 2006.

[7] A. J. de Langen, A. Vincent, et al., "Repeatability of 18F-FDG Uptake Measurements in Tumors: A Metaanalysis," *J. Nucl. Med.* vol. 53, pp. 701-708, May 2012.