SUV: From Silly Useless Value to Smart Uptake Value

One of the early promises of PET was that it would be quantitative (*1*), as opposed to imaging modalities such as x-ray CT, MRI, or SPECT, in which quantification is more of a challenge than in PET. By *quantitative* is meant that a raw PET signal can be transferred into absolute activity concentrations of the radiopharmaceutical in all tissues in units of, for example, kBq/cm^3. In this sense, PET can be classified as quantitative molecular imaging, since it provides the absolute number of radioactive molecules in each desired volume, whether this be a single image voxel or a larger volume of interest. Once the specific activity of the radiopharmaceutical is known,

PET images can be used for quantitative studies of the biologic processes in which the radioactive molecules are involved.

The transformation of raw PET signals into absolute numbers of radioactive molecules requires several steps, starting with hardware corrections, often referred to as scanner set-up, which deals with, for example, differences in photomultiplier sensitivities and other hardware-related issues. Remaining inhomogeneities are then corrected by a software procedure called normalization. To obtain quantitative images, one has to apply corrections for random

coincidences, detector dead time, photon attenuation, and scatter. The first 3 corrections are relatively straightforward, but scatter correction is more complicated and generally uses information on the source distribution obtained from an initial estimate without scatter correction and an attenuation map (*1*). Finally, for absolute quantification, a further, straightforward step is necessary: cross-calibration of the raw reconstructed image counts with the real activity concentration as derived from measurements in a dose calibrator.

QUANTIFICATION OF TUMOR VOLUME AND ^{18}F-FDG UPTAKE

Given that quantitative PET images are available, these can be used for pharmacokinetic modeling, oncologic staging, treatment planning, and response monitoring or for other numeric analyses. In general, these studies require volumes of interest, drawn manually, automatically, or semiautomatically around organs, tumors, or other tissues. Volumes of interest provide the sizes and volumes of these regions, as well as the total activity, mean activity concentration, standard deviation, and other measurements. For ^{18}F-FDG PET in oncology, these analyses have become part of routine clinical practice, next to visual interpretation of the images. Standardized uptake value (SUV), defined as the maximum or mean uptake in a tumor, scaled by the administered activity and by some measure of patient size (body weight, length, or a combination), has become a standard parameter of interest.

However, because of the modest spatial resolution of PET scanners (typically 5–6 mm in full width at half maximum, or down to 2 mm for the most advanced scanner types using special image reconstruction algo-

rithms), SUV and tumor volumes critically depend on the type of scanner, the reconstruction algorithm, and details of the delineation method used, especially for tumor sizes close to or smaller than the spatial resolution of the scanner. Simply stated, the measured activity concentration in a small volume of a PET image will not exactly represent the real value but depends on the activity in neighboring voxels. This is often called partial-volume effect or spillover, which can be separated into spill-out and spill-in depending on whether the volume considered is hotter or colder than surrounding tissues (*2*).

The partial-volume effect seriously hampers the determination of tumor volumes and corresponding SUVs in PET images. Most commonly, the tumor is delineated using a certain threshold in ^{18}F-FDG activity concentrations, and the voxels belonging to the tumor are assumed to be those that are above this threshold. However, determining how to define this threshold is not at all trivial. A simple approach, available in most commercial nuclear medicine software packages, is to define the threshold as a fixed percentage of the maximum voxel value within the tumor (e.g., using 50% isocontours).

TUMOR DELINEATION METHODS

Many refinements in threshold definition and other delineation methods have been proposed and investigated.

One can still use a percentage of the maximum voxel value as the threshold, but this percentage can be adapted and can be made dependent on the background activity concentration (*3,4*). Alternatively, a threshold can be defined as a percentage of the mean SUV plus a constant (*5*). In that

method, starting with a first guess for the percentage and the constant, the threshold and the mean SUV are updated by regression until convergence has been reached. A slightly different method uses region growth, starting with a single voxel in a tumor, whereby voxels are connected that have intensities higher than a certain percentage of the mean (6). This process stops when no additional voxels can be added to the region. Other iterative methods exist in which the tumor image is modeled as a convolution of the actual volume and a gaussian function describing the spatial resolution as a 3-dimensional point-spread function (7). Also, gradient-based methods exist (8,9) in which the borders of a tumor are characterized by gradients in activity concentration. Finally, statistical methods dealing with spatial correlations between voxels are being used (10,11).

Clearly, generally used and accepted methods for tumor delineation in PET do not yet exist. All methods have separately been described in the literature, have been validated or optimized using phantom experiments, and have sometimes been applied to realistic patient images. However, a comprehensive study in which different methods have been used on the same phantom and patient data, aiming at optimizations and definite conclusions as to which method is best, has been lacking.

A NEW AND VALUABLE CONTRIBUTION TO THE SUBJECT

The report by Tylski et al. in this issue excellently fills this gap (12). The authors present a systematic approach to the topic, based on both phantom and realistically simulated patient data for a large variety of tumor sizes and tumor-to-background-activity concentration ratios. They clearly identified 2 superior methods based on analyses of their large dataset.

In their experiments, the authors not only used a torso phantom containing lung inserts, a liver insert, and 17 spheres representing the "tumors" of variable size and activity concentra-

tions, inserted in the background, the lungs, and the liver inserts, but they also used an elegant method to simulate nonspheric tumors of known size and activity concentration in a real patient PET scan. To this purpose, a real PET scan of a patient with no tumors in the lungs was taken, and 41 realistic tumors of variable sizes and SUVs were inserted in this scan.

All phantom tumors and simulated patient tumors were analyzed by 5 different methods previously described in the literature. Using the terminology of the authors, these methods were T_{max}, T_{reg}, T_{mean}, T_{bgd}, and Fit, where T_{max} uses a threshold based on the maximum voxel value within the tumor, T_{reg} is the regression method using a percentage of the mean SUV plus a constant as the threshold (5), T_{mean} also uses mean SUV but is based on region growth starting with a single voxel within the tumor (6), T_{bgd} uses a combination of the mean SUV and the background activity concentration (4), and Fit is based on convolution of the tumor image with the point-spread function corresponding to the spatial resolution of the scanner (7). The Fit method was initialized using T_{bgd} with a fixed starting parameter; therefore, background activity was intrinsically accounted for in Fit. The SUV estimates were obtained with and without partial-volume correction (13).

Because all segmentation methods contain 1 or 2 adjustable parameters, the authors optimized these parameters for each method using the phantom data. For the simulated patient scans, the parameters were optimized using additional, simulated phantom data, obtained for the same scanner and acquisition settings. As a measure for the goodness of the method, the authors used the percentage errors in volume estimates and in SUV estimates. Of these parameters, both bias and variability were investigated.

Without addressing all details of the statistical analyses, the overall conclusion of the paper is that T_{bgd} and Fit gave the most accurate tumor volume estimates, with mean errors of 2% ±

11% and −8% ± 21%, respectively. Also with regard to SUV estimates, T_{bgd} and Fit with partial-volume correction performed best, resulting in mean errors of −2% ± 10% and 3% ± 24%, respectively. Remarkably, the commonly used T_{max} yielded considerably larger bias and larger variability in tumor volume and SUV estimates, as shown in Figures 3 and 5 of the paper. The authors also found that some methods were more sensitive than others to the proper settings of parameters. T_{max}, T_{mean}, and T_{reg} had very different biases for the simulated patient data, compared with the phantom data, whereas T_{bgd} and Fit were more constant across these datasets and optimization strategies. This finding led to the conclusion that the latter methods are more robust with regard to parameter settings.

PERSPECTIVE

Tylski et al. have shown for a large variety of "tumors" with different sizes, shapes, and tumor-to-background ratios, both in phantoms and simulated in the lungs of a real patient scan, that after optimization of parameters, 2 delineation methods gave satisfactory results whereas others clearly resulted in tumor volumes and SUVs deviating from their true values. These results are a significant step forward, especially in multicenter oncology studies where uniform delineation methods are required (14). However, implementation of the results does require an effort on site, since the parameter optimizations depend on the type of scanner and image reconstruction algorithm.

Extension to different types of tumors would be interesting, as, for example, liver tumors generally have a higher background activity than lung tumors. Further, it might be worthwhile to also include gradient-based and statistical delineation methods.

Finally, to make intercomparisons of SUVs and tumor volumes more of a routine procedure, it is necessary that advanced delineation algorithms become implemented into commercial

nuclear medicine software packages and that consensus be achieved with regard to the preferred method. Standard procedures and phantoms should be defined to optimize these algorithms for different types of scanners and reconstruction algorithms, possibly with the support of PET scanner manufacturers.

CONCLUSION

Tylski et al. have made an important contribution to the transformation of the "silly useless value" as described by Keyes et al. (*15*) into more of a "smart uptake value."

Eric P. Visser
Otto C. Boerman
Wim J.G. Oyen
Radboud University Nijmegen Medical Centre
Nijmegen, The Netherlands

REFERENCES

1. Cherry SR, Sorenson JA, Phelps ME. *Physics in Nuclear Medicine.* Philadelphia, PA: Saunders/Elsevier Science; 2003.
2. Zaidi H. *Quantitative Analysis in Nuclear Medicine Imaging.* New York, NY: Springer Science + Business Media; 2006.
3. Erdi YE, Mawlawi O, Larson SM, et al. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. *Cancer.* 1997;80(12, suppl):2505–2509.
4. Nestle U, Kremp S, Schaefer-Schuler A, et al. Comparison of different methods for delineation of ^{18}F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-small cell lung cancer. *J Nucl Med.* 2005;46:1342–1348.
5. Black QC, Grills IS, Kestin LL, et al. Defining a radiotherapy target with positron emission tomography. *Int J Radiat Oncol Biol Phys.* 2004;60:1272–1282.
6. Green AJ, Francis RJ, Baig S, Begent RH. Semiautomatic volume of interest drawing for ^{18}F-FDG image analysis—method and preliminary results. *Eur J Nucl Med Mol Imaging.* 2008;35:393–406.
7. Chen CH, Muzic RF Jr, Nelson AD, Adler LP. Simultaneous recovery of size and radioactivity concentration of small spheroids with PET data. *J Nucl Med.* 1999;40:118–130.
8. Geets X, Lee JA, Bol A, Lonneux M, Gregoire V. A gradient-based method for segmenting FDG-PET images: methodology and validation. *Eur J Nucl Med Mol Imaging.* 2007;34:1427–1438.
9. Li H, Thorstad WL, Biehl KJ, et al. A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours. *Med Phys.* 2008;35:3711–3721.
10. 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;28:881–893.
11. Montgomery DW, Amira A, Zaidi H. Fully automated segmentation of oncological PET volumes using a combined multiscale and statistical model. *Med Phys.* 2007;34:722–736.
12. Tylski P, Stute S, Grotus N, et al. Comparative assessment of methods for estimating tumor volume and standardized uptake value in ^{18}F-FDG PET. *J Nucl Med.* 2010;51:268–276.
13. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. *J Nucl Med.* 2007;48:932–945.
14. Boellaard R, Oyen WJ, Hoekstra CJ, 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;35:2320–2333.
15. Keyes JW Jr. SUV: standard uptake or silly useless value? *J Nucl Med.* 1995;36:1836–1839.