Understanding the Standardized Uptake Value, Its Methods, and Implications for Usage

The ancients relied on just their visual interpretations of bright objects in the heavens. But advancing technology led to quantifications that properly normalized stellar distances to obtain absolute magnitudes. Somewhat analogously, in PET the standardized uptake value (SUV) came to be used as a tool to supplement visual interpretation. Uptake normalization as a fraction of the injected dose/unit weight has been in use as far back as 1941 (*1*). It was designated as the differential absorption ratio (DAR) and in the 1980s was being used in PET (*2*). Aliases such as the differential (or dose) uptake ratio (DUR) and standardized uptake ratio (SUR) occasionally appear in the literature.

The SUV is a special member of a class of dimensionless Q (= average activity per unit volume) ratios in use: tissue Q ÷ a normalizing Q. The latter can be contralateral, a background, an organ (e.g., liver, brain, and so forth), and, in particular, the whole body as SUV = tissue Q ÷ whole-body Q including tracer excretions = tissue Q ÷ injected dose per unit body volume, weight, or area. For the (time invariant) denominator—rather than a region of interest (ROI) around the whole body or using units of volume—there are traditional and convenient uses of weight or body surface area, allowing one to obtain a dimensioned (in mg/mL or m^2/mL) result. The SUV (in mg/mL units), when averaged over the entire body, would equal the body density.

Commonly referred to as semiquantitative analysis, SUVs owe their popularity to a simplicity of method compared with others. In particular, it is of interest to compare the SUV with the more fully quantitative influx constant $K_i = k_1k_3(k_2 + k_3)$, which requires more effort. Fortunately for diagnostic purposes, the end-of-scan SUV can be diagnostically essentially as discriminating as K_i. This is because when K_i exists (i.e., $k_3 \neq 0$), the 2 are quite proportional, with a population-average proportionality constant depending only on the time of SUV evaluation and the type of tracer (*3*,*4*)—and, in particular, being virtually independent of the tissue type.

A reliable reproducible measure of uptake is sought. An underlying reason for this stems from wanting to make comparisons. Intrapatient comparisons occur during therapy monitoring. Interpatient relative uptakes in the diagnostic process can be related to degrees of pathology. Moreover, as knowledge bases form for a particular disease, contributions arise from several institutions. To facilitate comparisons in all of these settings, one hopes to address any unwanted dispersion in a marker such as the SUV. These would be about its true physiologic value for a specific disease condition when there are differing methodology preferences—the main issue in this perspective.

In seeking reproducible uptake measures, the diagnostician has a large armamentarium from which to choose according to Hoekstra et al. (*5*). The authors discuss 9 classes of analytic methods suitable for PET image analysis. These, including the SUV class, have subclasses. For the SUV, these subclasses have historically arisen out of a motivation to seek refinements in determinations that can reduce variability from method variations.

METHOD-INFLUENCING FACTORS

The SUV is subject to errors that can arise in various quantitative methods (*6*,*7*). A criticism of being casual about SUV methodology has been made (*8*), citing several factors to consider. However, as part of addressing PET procedures in a broader sense, several consensus groups (*9*–*13*)—in discussing some of these factors—have made protocol suggestions. SUV-related topics, along with other factors, are organized in Table 1, which distinguishes confounding factors (i.e., potentially addressable by adjusting SUV results) and defined factors (i.e., biologic states conventionally considered as distinct entities). A possible application of Table 1 would be an aid while writing a Methods section when SUVs are being reported—that is, indicate fully, not just partially, how SUV is being measured.

An aspect of facilitating comparisons is using variability-reducing corrections for the factors in Table 1. Statistical error in an average or reference result can be reduced to the extent that significant changes in confounding variables (e.g., in tumor size, SUV's time after injection, and so forth) are either used in corrections or kept minimal. It is desirous to have some reference conditions for the purpose of increasing the number of patients that can be grouped together in a study to gain statistical power. Also, diagnostic advantages accrue when variabilities from extraneous factors are reduced. That some SUV methods excel over others is seen in an example in which receiver-operating-characteristic (ROC) areas in ^{18}F-FDG PET breast cancer diagnosis can vary from 0.81 to 0.91 among the methods (*20*).

In this issue of *The Journal of Nuclear Medicine*, on pages 1519–1527, Boellaard et al. (*21*) use simulation to systematically investigate several of the influences in Table 1. Rather than attempting to isolate each influencing effect while all others remain constant, their approach was a practical one: vary-

Received Mar. 3, 2004; revision accepted Jul. 16, 2004.

For correspondence or reprints contact: Joseph A. Thie, PhD, University of Tennessee, 12334 Bluff Shore Dr., Knoxville, TN 37922.
E-mail: jathie@utk.edu

TABLE 1
Compilation by Category of Confounding Factors Influencing SUV Determination of Defined Tissue Type and State for Defined Population of Patients

Factor*	Comment
	Tissue activity factors
ROI shape within which to average	Either pixels or voxels can be used. Specific criteria for shape of outer boundary have precedents that include visual judgment, noise-affected maximum pixel, fixed size, contour defined by some fraction of maximum pixel, and others. Considerations also involve the character of heterogeneity encountered.
Partial-volume and spillover effects	Factors obtained in small phantom data allow observed ROI activity to be corrected to that truly present (*14*). There is dependency on reconstructed resolution, size and geometry, and the ratio of activities in ROI's region and surrounding region. Interrelated is motion blurring (e.g., from diaphragm) that undesirably averages pixel intensities.
Attenuation correction	Always required, methodology (e.g., γ-energy used, counts obtained, CT contrast agent usage, algorithm approximations) affects both absolute accuracy and noise.
Reconstruction method and parameters for scanner type	Techniques used in reconstruction—including spatial filter values, total number of pixels, and other parameters (e.g., number of iterations in some algorithms)—influence noise and resolution.
Counts' noise bias effect	Total number of noise equivalent counts from acquisition and analysis system, for a decayed dose, affects pixel randomness. ROIs based on maximum pixel value have pixel averages affected.
	Tissue state† factors amenable for corrections
Time of SUV evaluation	Injection-to-midacquisition time for SUV determination is a characterizing parameter (*15*). Traditionally, this time interval is chosen so that (dSUV/dt)/SUV is typically not excessive during acquisition. Corrections for time are possible (*16,17*).
Competing transport effects	In facilitated transport of amino acids or glucose, there is competition between variable serum concentrations of these and their tracer analogs. Where justified by data, correction for this effect can sometimes be appropriate.
	Normalization factor
Body size	In the (injected dose)/size denominator of SUV calculation, precedents for size are body weight, lean body mass, and body surface area. For ^{18}F-FDG, evidence shows the latter 2 generally reduce variability by more consistently describing a body volume into which tracer distributes (*18*).

*Factors can be interrelated. Thus, some aspect of one (e.g., induce more noise or influence resolution) may be listed elsewhere in table as a factor directly affecting SUV.

†Many biologic factors determine a particular tissue's uptake (*19*), such as kind and extent of disease, vascularity, organ usage, urine management policy, population characteristic, and so forth. Among such conditions at scan time, only 2 are singled out here as candidates for SUV corrections.

ing several readily controlled and clinically meaningful parameters (sometimes with individual effects therefore combining) in their simulation model. Their extensive data show how these affect R = the observed-to-actually-present counts ratio for an ROI = the observed ÷ the actual numerator of an SUV calculation. The scope of this research is not intended to address the last 3 factors in Table 1—that is, biologic factors and those influencing the denominator of an SUV calculation. The parameters varied are the type of ROI, the presence or absence of a spatial filter for one particular reconstruction algorithm, the noise equivalent counts collected, the number of pixels in the reconstruction matrix, uniform spheric lesion size, and the lesion-to-background ratio. The results show R ranging quite remarkably from as low as 0.4 to as high as 2.9 due to the synergism of having several parameters at once at their extremes. A much smaller, but more meaningful, measure of variability would be something like the SD of R expected due to random occurrences of various parameters having values typically encountered clinically.

Their simulation approach was to extract a part of the reconstruction code for their scanner and add features required for the investigation. Its results validate quite well against scan data from varieties of chest phantom conditions. Hence, the simulation could be used with confidence as influencing parameters are varied. It would appear that a subsequent use for this particular algorithm could possibly be using calculated R values to correct future SUVs for their off-normal parameters compared with some reference set of parameters. Strict validity of such corrections, however, would apply only to the chest region of this particular scanner and reconstruction for which the study was intended. With this approach, notwithstanding parameter differences, possible interinstitutional comparisons of corrected SUVs might be envisioned in which the PET hardware and software are a commonality. Otherwise, as Boellaard et al.

(*21*) caution, SUV comparisons among institutions cannot be made casually—in contrast to acceptable comparisons they show possible for same-patient intrainstitutional studies.

This work is significant in several respects. It quantifies, for their particular chest phantom, the SUV variability encountered due to a variety of factors in combination. In particular, it explores a little-publicized upward biasing effect of higher pixel noises in ROIs based on the maximum single pixel value. The variabilities associated with all parameters studied are useful to observe, both by those measuring SUVs and by others using these. The work also demonstrates how simulation can be useful in addressing these influences. Possibly, some day, interinstitutional SUV comparisons may be more confidently made with research of this type used in combination with some standardization of methods.

SUV USAGE

Visual interpretation, as the bulwark of radiology, is typically combined with other information in diagnoses. Of the latter, the SUV is closely allied to the image reading process. Among the reader's mental processes are qualitative comparisons of activities: within the image or with prior experience. Hence, unless comparisons are limited to images acquired by a frozen set of methods within an institution, the reader must be aware of the effects listed in Table 1. Having corrected SUVs available, along with knowledge of other factors influencing them, can be an aid.

If, after competing with or supplementing other analytic methods, the SUV has been chosen, Table 1 implies various choices to be made involving measurement parameters. The specific use of an SUV determination can have a bearing on the methods used. For example, correcting the ^{18}F-FDG SUV for serum glucose (i.e., traditionally SUV $\times$ [glucose concentration $\div$ a standard 100 mg/dL], though data for applicability to each unique tissue type should support this) can be appropriate when monitoring the same patient during therapy. On the other hand, many reports in the literature show no statistically proven advantage in applying this in studies composed of varieties of tissues. As another example, determining whether an intrinsically appropriate partial-volume correction is in fact beneficial can depend on prior experience—that is, whether an expected diagnostic advantage has been shown to be statistically significant in similar circumstances.

A popular usage of SUVs is their capability in helping to distinguish between benign and malignant lesions. For example, a study might find an SUV of 2.5 as appropriate for separating certain benign and malignant lesions. Caution, however, must be exercised using such a cutoff outside of the institution and the application for which it was determined. This is because there are institutional differences in the degree of diagnostic conservatism (i.e., choice of operating point on the ROC), the patient population, the specific pathology studied, and the specific methodology involved in determining the SUV. Interinstitutional

TABLE 2
Variability of Average SUVs Among Institutions for Particular Categories of ^{18}F-FDG PET Studies

Category	*n*	Average SUV	Average $\log_{10}$SUV $\pm$ SD
Non-Hodgkin's lymphoma	21	8.0	0.81 $\pm$ 0.29
	22	9.2	0.89 $\pm$ 0.29
	22	12.5	1.02 $\pm$ 0.27
Breast cancer	41	3.5	0.49 $\pm$ 0.20
	24	4.5	0.57 $\pm$ 0.27
	36	5.1	0.63 $\pm$ 0.23
	26	12.8	1.02 $\pm$ 0.29
Pancreatic cancer	42	3.2	0.45 $\pm$ 0.24
	34	4.4	0.60 $\pm$ 0.18
	23	6.5	0.77 $\pm$ 0.18
Head and neck squamous cell cancer	48	3.2	0.49 $\pm$ 0.11
	22	6.3	0.74 $\pm$ 0.24
	37	9.4	0.94 $\pm$ 0.19
Normal liver	82	1.7	0.22 $\pm$ 0.13
	24	2.5	0.40 $\pm$ 0.07
	37	2.7	0.43 $\pm$ 0.09

In meta-analysis (*22*) within each category, using individual patient $\log_{10}$SUV values in Kruskal–Wallis ANOVA, the *P* value for at least 2 institutions differing in means is always found to be <0.0001. Reasons can be a combination of factors in Table 1 along with variations in populations and pathologies chosen for study. The higher SD values of the logarithms (which are approximately [1/ln 10] $\times$ coefficients of variation of SUVs here) for cancers compared with those of normal liver presumably show variability stemming from extents of disease.

variability stemming from the latter 3 categories is evident from Table 2, extracted from a meta-analysis of ^{18}F-FDG PET studies, each having ≥20 patients with SUVs (*22*). But, if population character and pathology factors existing within the disease categories in Table 2 could be eliminated, a much better picture might emerge. This is suggested somewhat from a subset of 20- to 40-y olds within a meta-analysis (*23*) of institutions studying the coefficients of variation (CVs) of normal whole brains' metabolic rates: The average of the CVs of individual studies within 26 institutions = 0.15; the CV contribution due solely to interinstitutional differences = 0.14; and the total (i.e., combined patient and institutional variabilities) CV of metabolic rates among 26 institutions = 0.20. These metabolic rate CVs contrast with the much larger SUV interinstitutional variabilities apparent within categories of Table 2.

Finally, it might be tempting to suggest that, to avoid uncertainties or inaccuracies in an SUV approach, one should turn to fully quantitative methods such as K_i determinations. These determinations offer the benefit of avoiding issues of evaluation time and body size normalization, though typically they exhibit slightly higher intrainstitutional interpatient variability (*24*) than that found in SUVs. But the other factors in Table 1 still remain to be addressed. With the proportionality constant between SUV and K_i being only physiologically (and not methodology) based (*3,4*), there is a suggestion that if Table 2 were for K_i determinations, the conclusion could be the same: substantial variability among institutions within each particular category of studies.

CONCLUSION

Messages to carry away from the work of Boellaard et al. (*21*), supplemented by Table 1 presented here, are the desirability of standardization of protocols and analyses and that both measurer and user of SUVs have an awareness of all influencing factors. Many, though not all, of the latter also apply to influx and rate constant determinations and, to some extent, qualitative visual interpretations. Fortunately, within an institution there can be its preferred de facto standardized approach to SUVs. However, there might be changes over time or lack of acceptance by all, and special caution must be obviously be exercised in interinstitutional SUV usage, which presently is difficult. Helpful for the time being would be a better-documented specification of methods in publications than may be customary, considering Table 1 for guidance. To make the most of this diagnostic tool, as well as benefit from fully quantitative analytic methods, challenges for the future might include:

- Further efforts by organizations to reach consensus on standardized approaches in scanner data acquisition and analysis, building on past accomplishments (*9–13*).
- Highly automated user-friendly software that corrects and reports, perhaps by a simulation algorithm, SUVs along with values of their influencing parameters.
- Along with patient data, possible reporting of scanned standard phantoms—recommended by a PET Data Analysis Working Group (*9*) and also suggested by Boellard et al. (*21*)—and possibly including SUVs, as known activity ratios of local to whole phantom, for the actual geometries and activities being encountered.
- More research of the type reported in this issue.

Joseph A. Thie, PhD
University of Tennessee
Knoxville, Tennessee

REFERENCES

1. Kenney JM, Marinelli LD, Woodard HQ. Tracer studies with radioactive phosphorus in malignant neoplastic disease. *Radiology*. 1941;37:683–690.
2. Kubota K, Matsuzawa T, Ito M, et al. Lung tumor imaging by positron emission tomography using C-11 L-methionine. *J Nucl Med.* 1985;26:37–42.
3. Thie JA. Clarification of a fractional uptake concept. *J Nucl Med.* 1995;36:711–712.
4. Sadato N, Tsuchida T, Nakaumra S, et al. Noninvasive estimation of the net influx constant using the standardized uptake value for quantification of FDG uptake of tumors. *Eur J Nucl Med.* 1998;25:559–564.
5. Hoekstra CJ, Paglianiti I, Hoekstra OS, et al. Monitoring response to therapy in cancer using [^{18}F]-2-fluoro-2-deoxy-D-glucose and positron emission tomography: an overview of different analytical methods. *Eur J Nucl Med.* 2000;27:731–743.
6. Carson RE. Precision and accuracy considerations of physiological quantitation in PET. *J Cereb Blood Flow Metab.* 1991;11:A45–A50.
7. Alavi A, Smith R, Duncan D. What are the sources of error in measuring and calculating cerebral metabolic rates with fluorine-18-fluorodeoxyglucose and PET? *J Nucl Med.* 1994;35:1466–1468.
8. Keyes JW. SUV: standard uptake or silly useless value? *J Nucl Med.* 1995;36:1836–1839.
9. Rapoport SI. Discussion of PET workshop reports, including recommendations of PET Data Analysis Working Group. *J Cereb Blood Flow Metab.* 1991;11:A140–A146.
10. Schelbert HR, Hoh CK, Royal HD, et al. Procedure guideline for tumor imaging using fluorine-18-FDG. *J Nucl Med.* 1998;39:1302–1305.
11. Young H, Baum R, Cremerius U, et al. Measurement of clinical and subclinical tumour response using [^{18}F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. *Eur J Cancer.* 1999;25:1773–1782.
12. Bartenstein P, Asenbaum S, Catafau A, et al. European Association of Nuclear Medicine procedure guidelines for brain imaging using [^{18}F]FDG. *Eur J Nucl Med Mol Imaging.* 2002;10:BP43–BP48.
13. Bourguet P. 2002 Standards, options and recommendations for the use of [^{18}F]-FDG (PET-FDG) in cancerology [in French]. *Bull Cancer*. 2003;90(spec no.):S1–S109.
14. Hoffman EF, Huang SC, Phelps ME. Quantitation in positron emission computed tomography: 1. Effect of object size. *J Comput Assist Tomogr.* 1979;3:299–308.
15. Hamberg LM, Hunter GJ, Alpert NM, et al. The dose uptake ratio as an index of glucose metabolism: useful parameter or oversimplification? *J Nucl Med.* 1994;35:1308–1312.
16. Thie JA, Hubner KF, Smith GT. Optimizing imaging time for improved performance in oncology PET studies. *Mol Imaging Biol.* 2002;4:238–244.
17. Beaulieu S, Kinahan P, Tseng J, et al. SUV varies with time after injection in ^{18}F-FDG PET of breast cancer: characterization and method to adjust for time differences. *J Nucl Med.* 2003;44:1044–1050.
18. Schomberg A, Bender H, Reichel C, et al. Standardized uptake values of fluorine-18 fluorodeoxyglucose: the value of different normalization procedures. *Eur J Nucl Med.* 1996;23:571–574.
19. Huang SC. Anatomy of SUV. *Nucl Med Biol.* 2000;27:643–646.
20. Avril N, Bense S, Ziegler SI, et al. Breast imaging with fluorine-18 FDG PET: quantitative image analysis. *J Nucl Med.* 1997;38:1186–1191.
21. 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.
22. Thie JA, Hubner KF, Smith GT. The diagnostic utility of the lognormal behavior of PET standardized uptake values in tumors. *J Nucl Med.* 2000;41:1664–1672.
23. Wang GJ, Volkow ND, Wolf AP, Brodie JD, Hitzemann RJ. Intersubject variability of brain glucose metabolic measurements in young normal males. *J Nucl Med.* 1994;35:1457–1466.
24. Thie JA, Hubner KF, Smith GT. Standardized uptake value and influx constant: relationships and variabilities with model interpretation and clinical implications. *Clin Positron Imaging.* 1999;2:99–104.