The Effect of Small Tumor Volumes on Studies of Intratumoral Heterogeneity of Tracer Uptake

Frank J. Brooks[1] and Perry W. Grigsby[1–4]

[1]Department of Radiation Oncology, Washington University School of Medicine, St. Louis, Missouri; [2]Division of Nuclear Medicine, Mallinckrodt Institute of Radiology, Medical Center, St. Louis, Missouri; [3]Department of Obstetrics and Gynecology, Washington University Medical Center, St. Louis, Missouri; and [4]Alvin J. Siteman Cancer Center, Washington University Medical Center, St. Louis, Missouri

The number of studies in the literature involving quantification of the metabolic heterogeneity seen in ^{18}F-FDG PET images has increased sharply over recent years. We hypothesized that inclusion of very small regions of interest as unique data points will have deleterious effects on these studies. **Methods:** Using a combination of probability theory and clinical ^{18}F-FDG PET data, we numerically calculated the curve describing the probability a given tumor volume is large enough to adequately sample the underlying tumor biology assayed via a PET/CT scanner at a planar resolution of 4 mm and transaxial resolution of 4 mm (64 mm^3 voxel size). We then used a computer simulation to isolate the effects of tumor volume on the image local entropy. **Results:** We computed the underlying global intensity distribution for 70 cervical cancer tumors ranging from 4 to 248 cm^3), which were ensemble-averaged over the same intensity scale. From this distribution, we determined that about 700 total voxels (45 cm^3) are required to give 95% certainty that the global intensity distribution has been sufficiently sampled for common statistical comparisons of individual tumor intensity distributions to be made canonically. We demonstrated that one previously suggested measure of heterogeneity is dependent on tumor volume and that measurement of heterogeneity is about 5 times more sensitive to volume changes for volumes below the proposed minimum than for those above it. **Conclusion:** Inclusion of tumor volumes below 45 cm^3 can profoundly bias comparisons of intratumoral uptake heterogeneity metrics derived from data from the current generation of whole-body ^{18}F-FDG PET scanners.

Key Words: ^{18}F-fluorodeoxyglucose; cancer of the uterine cervix; local entropy; positron emission tomography; texture analysis

J Nucl Med 2014; 55:37–42
DOI: 10.2967/jnumed.112.116715

With advances in medical imaging techniques, there is increasing interest in the quantification of cancerous tumor microenvironments. Modern imaging enables the description of intratumor qualities in situ. One example is the use of the ^{18}F-FDG radioactive glucose analog with PET (*1*). Consider, for example, the ^{18}F-FDG PET image of a cancer of the uterine cervix shown in Figure 1. There, greater gray-scale pixel intensity (brighter) ostensibly implies greater metabolic activity. It is this type of heterogeneity that interests researchers of tumor biology (*2*).

In the specific case of ^{18}F-FDG PET images, spatial variations among differently shaded pixels are to be quantified. The goal is to objectively declare one tumor, or intratumoral region, to be more heterogenous than another tumor or intratumoral region with the hope that image heterogeneity quantifiers will provide prognostic clinical value. Toward this end, several quantifiers have been proposed (*3–8*). Regardless of the specific heterogeneity quantifier used, the distribution of gray-scale intensities constrains the values that quantifiers can attain. In short, fewer unique intensities implies less possible heterogeneity.

The distribution of measured image intensities depends on both tumor biology and imaging physics. In the case of ^{18}F-FDG PET, the well-known partial-volume effect tends to lower uptake values while increasing apparent tumor volume (*9*). In other words, the partial-volume effect is known to increase the number of unique intensities measured. This can cause distributions of measured intensities to appear more heterogeneous than would be dictated by tumor biology alone. Whatever their combined role, both physical and biological sources of image heterogeneity could yield their own prognostic information about the tumor. Therefore, interpatient comparison of objectively quantified image heterogeneity could be important clinically.

Because the value of uptake heterogeneity quantifiers depends crucially on the distribution of the gray-scale intensities for each patient, adequate sampling of those distinct distributions is paramount for comparative heterogeneity studies. Because the number of samples of each intensity distribution is the number of image pixels in the identified region of interest (ROI), the tumor volume itself indicates how well an individual intensity distribution has been sampled. We therefore hypothesized that there is some minimum tumor volume below which comparison of intratumoral uptake heterogeneity quantifiers is invalid because of under-sampling.

It was the purpose of this research to describe the computation of a lower bound on tumor volume below which the effects of latent under-sampling are profound. We then demonstrated this small-volume effect on one previously proposed metric of uptake heterogeneity, the image local entropy (*5,6,10*).

MATERIALS AND METHODS

Delineation of Tumor Regions

This was a retrospective study of 70 patients with cancers of the uterine cervix who underwent pretreatment hybrid PET/CT (Biograph

Received Feb. 15, 2013; revision accepted Aug. 1, 2013.
For correspondence or reprints contact: Frank J. Brooks, Department of Radiation Oncology, Washington University School of Medicine, 4921 Parkview Place, St. Louis, MO 63110.
E-mail: fjbrooks@wustl.edu
Published online Nov. 21, 2013.

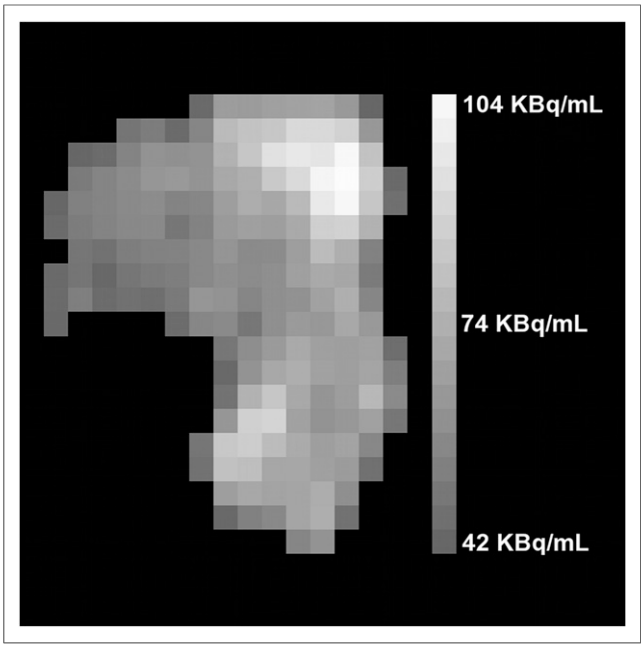

FIGURE 1. Transaxial ^{18}F-FDG PET cross-section of uterine cervical cancer with clear variations in intensity. For example, upper right corner of tumor is brightest and several darker spots are visible throughout. These intensity variations represent variations in ^{18}F-FDG uptake within tumor. Vertical edge of image corresponds to 10 cm within patient.

40 True Point Tomograph Scanner; Siemens) using the ^{18}F-FDG radiotracer assay of glucose uptake by cells. The raw ^{18}F-FDG PET data were scatter- and attenuation-corrected via the proprietary software native to the PET machine. Images were reconstructed using ordered-subset expectation maximization (8 subsets; 4 iterations). A gaussian smoothing filter of 4 mm in full width at half maximum was applied after reconstruction. No additional processing was implemented. The relevant ROI first was identified visually by an experienced oncologist. To objectively delineate tumor from background, any ROI pixel brighter than 40% of the maximum ROI pixel brightness was considered part of the tumor (*11*). The oncologist then made slight manual adjustments to the ROI to remove any obvious nontumor pixels such as those comprising bladder or bowel regions. This ROI was exported as a set of Cartesian coordinates in DICOM structure files. Use of these data for this retrospective research with waiver of informed consent was approved by the Washington University Institutional Review Board. Tumors ranged in size from 4 to 248 cm^3 and approximately followed an exponential distribution with median volume of 29 cm^3. Because our ^{18}F-FDG PET data were given to a planar resolution of 4 mm and a transaxial resolution of 4 mm, we used the conversion factor of 0.064 cm^3/voxel throughout this work.

Relative Intensity Scale

The ROIs and the original 15-bit gray-scale DICOM images were imported into custom computer software written in Python, version 2.6.2 (http://www.python.org/), using the pydicom library, version 0.9.3 (http://code.google.com/p/pydicom/). For each patient, our software automatically extracted the image pixels bounded by a given ROI, confirmed those pixels to be above the clinical threshold, and then stored the pixel intensities as observed radioactivity densities given in Bq/mL. Those values were then binned into probability histograms using the Freedman–Diaconis optimal bin width (*12*) for that group of radioactivity densities (i.e., the bin size was computed for each patient). The values were then rescaled such that the domain of each histogram was 0.4 (the clinical threshold) to 1.0. To facilitate interpatient comparison of histograms, each histogram was resampled via cubic splines at intervals of 5% intensity using the SciPy interpolation package, version 0.7.1 (http://www.numpy.org/). This interval was found to be the median Freedman-Diaconis bin size in the rescaled domain for all imaged tumors. The final result was that each tumor was associated with a common-scale histogram representing the probability that a given intensity interval appears within that tumor.

Creation of Test Images

To isolate the effects of tumor size on the example heterogeneity metric, we created shapeless "tumor" images with a known intensity histogram. A perfect square number of tumor voxels was used as the number of pixels in a square, 2-dimensional, 8-bit gray-scale image created via the Python Imaging Library, version 1.1.7 (http://www.pythonware.com/products/pil/). Each image pixel was chosen at random to have an intensity drawn from a known-intensity histogram (which is presented in the "Results" section). The simulation in no way attempted to mimic the PET scan process; it only represented distinct, variable-size samplings of the distribution of measured ^{18}F-FDG PET intensities for our set of patients. This randomization may be repeated numerous times for a given number of voxels, which we multiply by 0.064 cm^3/voxel. The result is a set of many test images that, on average, obey the identical intensity distribution while having no consistent spatial intensity patterns. Thus, each set of test images represents tumors of identical volume, identical average shape, and identical average heterogeneity.

Example Heterogeneity Statistic

We computed the local information entropy of a 2-dimensional image as described by Haralick et al. (*13*). In brief, the cooccurrence matrix describes the probability p that a pixel of a shade i occurs next to a pixel of shade j. This matrix can be computed for various directions, pixel separations, and bit depths. We computed the horizontal and vertical cooccurrence matrices for the nearest pixel neighbors of 8-bit gray-scale images. From each of these matrices, the local entropy

$$h = -\sum_{j=103}^{255}\sum_{i=103}^{255} p(i,j)\ \ln p(i,j) \qquad \text{Eq. 1}$$

was computed for each direction and then root-mean-square–averaged to obtain a single local entropy value. The limits on the summations reflect the 40% clinical threshold within the 8-bit (0–255) color scale.

RESULTS

The Ensemble Intensity Histogram

The relative intensity histograms for each patient were ensemble-averaged into a single relative intensity distribution. The resulting probability histogram is shown in Figure 2. This histogram enables the estimation of a minimum volume required for heterogeneity studies.

Derivation of the Minimum Volume

Given the probability ρ that an intensity will be in the least populated bin, the probability of having precisely L intensities in the least populated bin after choosing v voxels is

$$P(v,L) = \binom{v}{L}\rho^L(1-\rho)^{v-L} \qquad \text{Eq. 2}$$

since there are "v choose L" ways of arranging a sequence containing L intensities among the $v - L$ intensities from all other

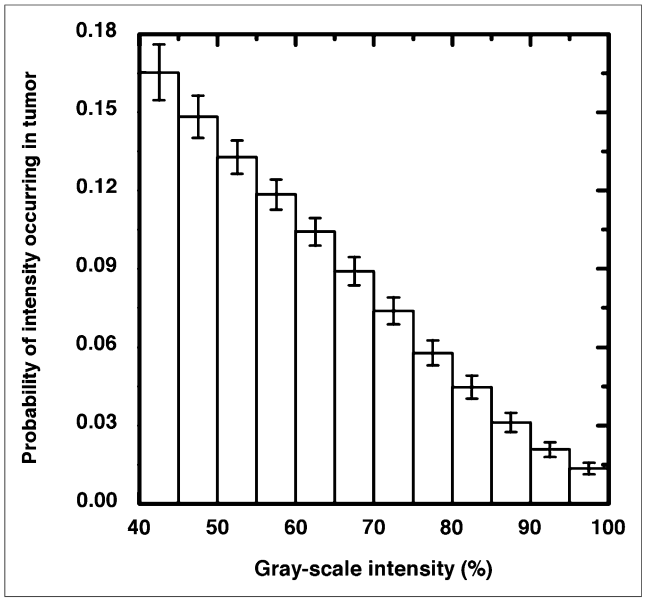

FIGURE 2. Intensity histogram resulting from the ensemble average of 70 tumors, each of which is measured on same percentage-of-maximum scale. Intensities are approximately linearly distributed until flattening of tail occurs at highest intensity values. Error bars represent SE within each intensity bin.

bins collectively (*14*). We, however, allow for at least L intensities, since higher bin-population scenarios could contribute probability to lower bin-population scenarios and since having many more samples than the bare minimum is preferable. We therefore sum the individual probabilities given in Equation 2 as

$$P(v,\lambda \geq L) = \sum_{\lambda=L}^{\infty} P(v,\lambda). \qquad \text{Eq. 3}$$

We need not derive a closed analytic form for $P(v, \lambda \geq L)$ in order to discover the requisite number of voxels to ensure a minimum least-bin population. For the present derivation, we chose the traditional minimum of 5 frequencies per tested contingency as the required minimum population. Further reasoning behind this choice and its impact on our results is presented in the "Discussion" section. In Figure 3, the probability of having precisely L intensities in the least populated bin is shown for several examples (L = 5, 10, 20, 30, and 40), where we have rescaled v to more familiar units of cm^3. The first curve (solid) shows that near a volume of 100 cm^3, the probability of having precisely 5 intensities in the least populated bin is approximately zero. This is because, for this volume, so many samples have occurred that more than 5 intensities is virtually guaranteed to be in the least populated intensity bin. We thus use 0–100 cm^3 as a practical domain on which to focus our search. As is seen in Figure 3, the precise-number probabilities are nonzero over this domain to L = 40 (dot-dot-dash). We may therefore sum Equation 3 from λ = 5 to λ = 40 and be assured that we have included all scenarios for which 5 or more intensities populate the least populated bin for volumes within the 0- to 100-cm^3 search domain. The resulting probability of having at least 5 intensities in the least populated intensity bin for ρ = 0.013 (this value is read from the ensemble histogram) is plotted in Figure 4 as the solid curve.

Impact on Clinical Data Analysis

The severity of under-sampling on clinical studies is shown by the vertical lines in Figure 4, which indicate the first and second quartiles of our tumor volume data. The adequate-sampling probability averaged over 1 cm^3 to the first quartile (12 cm^3; cross-hatched) is only 2%. Between 1 cm^3 and the median (29 cm^3; hatched), the average is 25%. The large dot at 45 cm^3 indicates the intersection with P = 0.95, that is, the volume at which one may be 95% certain that enough samples reside in the least populated intensity bin for meaningful statistical comparisons. For our clinical data, this leaves less than half (34%) the tumor volumes as viable data points for comparing ^{18}F-FDG PET heterogeneities. The other 2 curves in Figure 4 represent the probability of having at least 5 intensities in the least populated bin but using $\rho \pm 1.96(0.002)$ as the probability of populating the least populated bin (here, 0.002 is the SE read from the ensemble histogram). The bar between the nonsolid curves thus indicates the 95% confidence interval around 45 cm^3. The results of the above calculation for less stringent sampling criteria are shown in Figure 5. Requiring fewer than 5 samples in the least populated intensity bin is unlikely to sufficiently abate the under-sampling effects we describe.

Demonstrated Effect on Heterogeneity Studies

It has been proposed that local entropy may be a useful clinical measure of uptake heterogeneity (*5,6,10*). Furthermore, local entropy has been claimed to be the most reproducible metric among similar heterogeneity metrics (*10*). We therefore chose to demonstrate the small-volume effect on comparative heterogeneity studies via the local entropy metric. We first sought to isolate the effect of tumor size from the effects of intensity distribution or intensity rearrangement. This was done by creating sets of 2-dimensional, shapeless tumor images. For each tumor volume, 25 tumor images were created on which the local entropy was computed and then

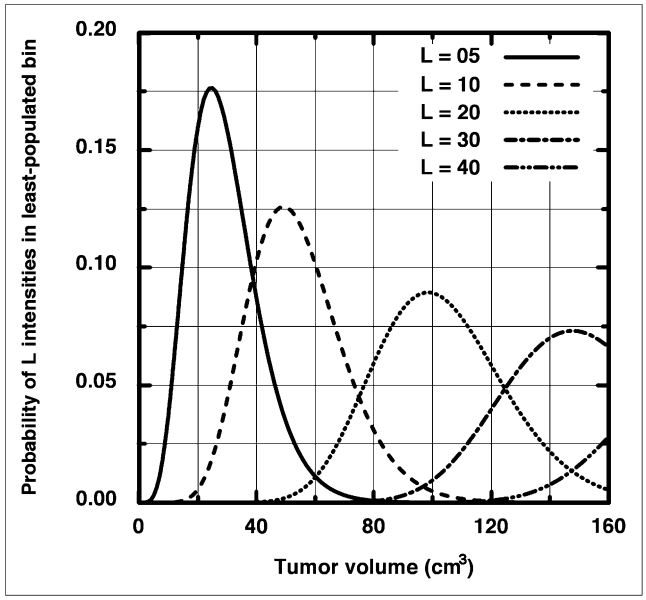

FIGURE 3. Probability that precisely L samples fall into last intensity bin given in Figure 2, plotted for several L values. Up to 100 cm^3 probability curves from higher numbers of samples ($40 > L > 5$) are not zero and therefore contribute to the probability that at least 5 samples fall into final intensity bin.

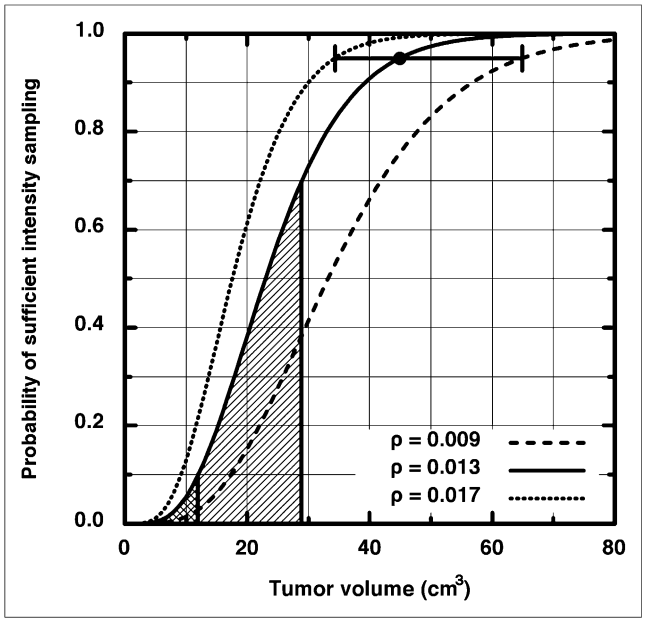

FIGURE 4. Probability (solid curve) that at least 5 samples fall into the last intensity bin given in Figure 2. Large dot indicates level of 95% certainty that adequate sampling of intensity distribution has occurred. Other curves represent probabilities computed from 1.96 times 1 SE above or below ensemble-average last-bin probability. Horizontal error bar extends from 34 cm^3 (531 voxels) to 65 cm^3 (1,016 voxels).

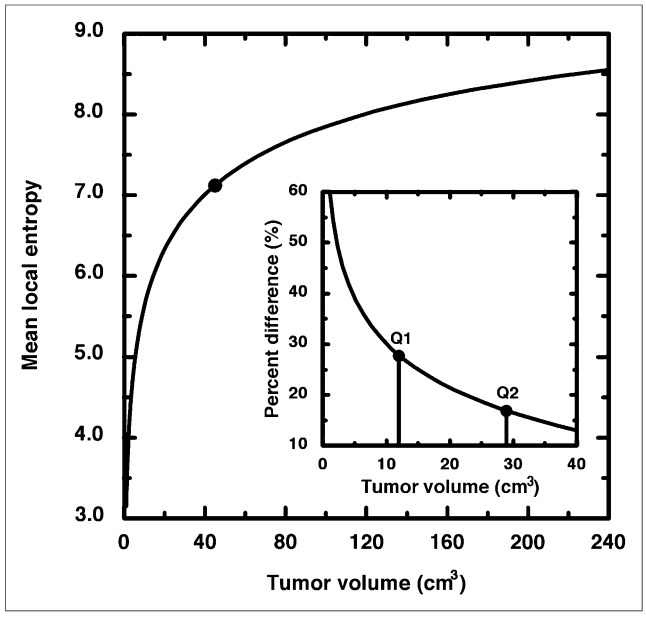

FIGURE 6. Ensemble average of root-mean-square local entropy plotted as function of image size (tumor volume). Local entropy is much less sensitive to volume for volumes greater than 45 cm^3 (large dot). Inset shows percentage difference in mean local entropy from the value averaged over only large volumes. Vertical lines indicate first and second quartiles of our tumor volume data. First quartile exhibits large deviation from large-volume average.

ensemble-averaged to a single mean heterogeneity value. The result is the plot of local entropy (h) versus tumor volume (v) given in Figure 6. Foremost is the striking increase in $h(v)$ over the first 45 cm^3 of tumor volume compared with the flatness of $h(v)$ for volumes greater than 45 cm^3. For $240 > v > 45$ cm^3, the mean value is $\langle h \rangle = 8.1$ and the individual h values differ (on average) by only 4% from that mean. We now compare this large-volume mean to the small-volume h values as is done in heterogeneity studies when tumors of widely varying volumes are analyzed together. The first quartile of tumor volumes—as indicated by the first vertical line in the inset of Figure 6—differ on average by 38% from the large-volume mean $\langle h \rangle = 8.1$. For all $v < 45$ cm^3, $h(v)$ still differs on average by 23%. Therefore, the local entropy is about 5 times more sensitive to a volume change applied to a small volume than to the same change applied to a large volume. It is thus seen that before any assessment of tumor biology has been made, the statistic ostensibly doing that assessment has been saddled with a nonnegligible value change that has nothing to do with tumor biology.

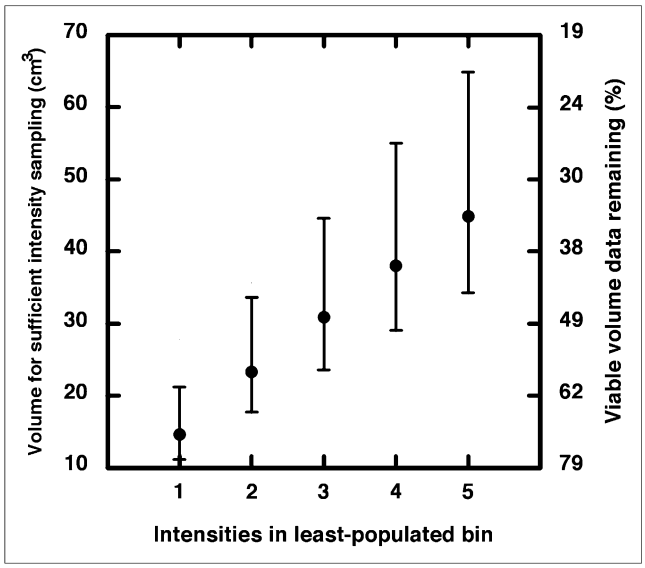

FIGURE 5. (Left-hand scale) Volume associated with sufficient sampling plotted vs. increasingly strict criterion for adequate sampling of intensity distribution. (Right-hand scale) Nonlinear scale indicating approximately how much of our volume data remains after imposition of each adequate-sampling criterion.

DISCUSSION

For any tumor assayed via ^{18}F-FDG PET, the tumor data are a distribution of gray-scale intensities. Data of this intensity distribution represent the biology of the tumor. We have demonstrated a feasible clinical scenario in which tumors following identical intensity distributions—that is, identical measured tumor biology—have heterogeneity measures that depend strongly on tumor volume. Therefore, differences in uptake heterogeneity observed between disparate tumor volumes may not indicate actual biologic differences between those tumors.

We chose to illustrate this point via the local entropy because that statistic has been proposed previously as a robust measure of uptake heterogeneity; however, we now argue that heterogeneity statistics generally are more sophisticated than the statistical moments familiar to most clinicians. Heterogeneity is a measure of the deviation from homogeneity. In image processing parlance, heterogeneity is the "texture" of the image; the differences from smoothness. In 1973, Haralick et al. described a comprehensive set of texture metrics for gray-scale images (*13*). Each of those

fundamental metrics, one of which is the local entropy, is itself computed from gray-scale co-occurrence matrices. These matrices are simply the tallies of differences between pixel neighbors. That is, co-occurrence matrices track the probability that fixed-distance pixels are shaded differently. Over the entire image, these local variations accrue into the global texture statistics some propose to use as a measure of tumor heterogeneity. Precisely because they are accrued statistics, they measure only the information that actually is contained in the image data and, as such, must to some degree depend on sample size.

This dependence on sample size is not a failing of existing texture metrics. In quantifying texture, one is interested in the spatial patterns and intensity variations observed in image data. In ^{18}F-FDG PET images, these variations ultimately are caused by some combination of the scanning process and tumor biology. If the total number of pixels is largely diminished, so too is the certainty that patterns and variations allowed by the underlying biology have had adequate opportunity to manifest. To put this another way, the set of intensities and intensity arrangements necessary to build a complete picture of the possible biology is itself incomplete for small volumes. Mathematically speaking, many texture metrics proposed to measure uptake heterogeneity are supposed to have completely different values for smaller volumes—values that may have no predictable relation to those computed for larger volumes. This is in stark contrast to the use of statistical moments such as the distribution variance as a first-order heterogeneity quantifier where, in the case of a common intensity distribution, moment values predictably regress to the mean values as sampling (of any size) is repeated across patients.

The minimum tumor volume we describe is a minimum with regard to the type of comparative heterogeneity analysis some researchers have proposed. That is, it is a minimum imposed by the desire for robust mathematical manipulation of intratumor statistics. Our point is that although very small tumor volumes may be sufficient for treatment planning or other clinical purposes, they do not necessarily contain enough intensity data to be further analyzed using the heterogeneity quantifiers earlier proposed.

We used a straightforward argument to estimate that the minimum tumor volume for adequate intensity sampling is about 700 voxels (45 cm^3 for our image data). This argument is based ultimately on a tried-and-true criterion found in classic (*14,15*) and modern (*16*) textbooks alike regarding adequate sampling of unknown distributions that are to be compared via χ^2 goodness-of-fit testing. We chose this test because, in essence, that is what statistics derived from ^{18}F-FDG PET images are—a comparison of intensity distributions. Without sufficient frequencies in every possible contingency, the χ^2 statistic does not regress to the χ^2 distribution, and table values regarding "significance" levels become moot. Although there are situations in which less strict sampling criteria are appropriate (*14–16*), for the present context, the reasoning against these lax criteria is clear from Figure 6. If, for example, a minimum intensity-bin population of only one were required, the corresponding volume ($\approx$15 cm^3; Fig. 5) would yield a local entropy value in the most steeply increasing portion of the $h(v)$ curve. In other words, if the measured intensities have not sufficiently revealed the underlying intensity distribution, the heterogeneity metric is highly sensitive to tumor volume. Thus, inclusion of tumor volumes—or intratumor regions—so small as to essentially guarantee that under-sampling has occurred must bias the results of any comparative uptake heterogeneity study. Therefore, in the context of such studies, the default presumption should be that no statistical inference whatsoever may be made from small ^{18}F-FDG PET volumes. The onus is on the researcher to demonstrate that a new heterogeneity result is not due to the effects of under-sampling.

As seen in Figure 6, $h(v)$ is monotonic in v and therefore acts as a mere surrogate for tumor volume. This means that a decrease in volume must correspond to a decrease in heterogeneity. What is important here is the relative size of decrease. From Figure 6, it is seen that the derivative of h with respect to v is much less for the larger volumes than for the smaller ones. This indicates that h is much less sensitive to changes in v for large v. Thus, discovering a 20% change in heterogeneity between 2 large volumes actually could be significant since we suspect that h is not strongly affected by volume over the domain of only large volumes. Contrast this, for example, to the comparison of heterogeneities between an 80 cm^3 tumor and a 20 cm^3 tumor where a 20% change in h is seen to be caused by volume alone.

We feel that performing the analysis we presented in a higher dimension does not offer anything new while only serving to complicate subsequent discussion. The local entropy is extensible into 3 dimensions. However, because of the increased dimensionality, one must include much more detail as to which of the directional cooccurrence matrices are computed and how statistics derived from those matrices are implemented or combined. Furthermore, because these matrices already represent spatial averages over the entire image set (*13*), rearrangement of the image data is not likely to alter the qualitative dependence on volume. In specific cases of quantitative dependence, added dimensionality increases the number of spatial intensity configurations possible while the sampling (tumor volume) remains the same. Intuitively, however, as the number of unique scenarios a quantifier can describe increases—whether through data dimensionality or quantifier sophistication—the minimum sample size for meaningfully comparing those scenarios increases as well. We therefore predict the deleterious small-volume effect we describe to be even worse for 3-dimensional data and for higher-order heterogeneity metrics but caution that the specific dependence of a given metric on dimensionality quickly goes far beyond the intended scope of this work.

Another potential influence on the quantitative value of uptake heterogeneity metrics is the partial-volume effect. In the case of ^{18}F-FDG PET, the partial-volume effect renders voxels at the tumor–microenvironment interface to be less bright than voxels filled with tumor (*9*). In fact, any intensities measuring less than the brightest biologically possible intensity could result solely from partial filling with background or necrotic regions. It is therefore feasible that the partial-volume effect creates at least some—if not all—of the heterogeneity that is visually evident in ^{18}F-FDG PET images. For this reason, one might expect that partial-volume correction could influence our calculations. It is almost certain that such a correction will alter the overall distribution of measured ^{18}F-FDG PET intensities. However, this alteration is unlikely to profoundly affect the bright end of the intensity histogram, because the brightest pixels likely are already the closest to being completely filled with tumor and therefore are the least affected by the partial-volume effect (or partial-volume correction). The nonsolid curves shown in Figure 4 represent the effect of plus or minus 1 SE around the probability of measuring the brightest intensities. The horizontal error bar in Figure 4 might therefore represent a reasonable bound on partial-volume effects (or any systematic measurement effects) because, by ensemble averaging many same-scale histograms, we

have allowed the biology of similarly sized tumors numerous chances to manifest under the same scanning process.

The calculation of minimum volume we present is better thought of as a technique to be applied to distinct image datasets, rather than a justification for a rigid rule to be applied uniformly to all image data. What is crucial to our calculation is the number of samples of the underlying intensity distribution. Although this means that calculation of the probability at which at least L observations reside in the least populated intensity bin is independent of voxel size (Eq. 3), the input probability (ρ) is not. Because different PET scanners have voxels corresponding to different physical sizes, and because both partial-volume effects and uptake can depend nonlinearly on tumor size and biology, the distribution of measured intensities is itself likely unique to the particular combination of tumor type, scanner resolution, and scanner modality. In other words, the probability that any intensity resides in the least populated bin—as well as the bin size definition itself—is sensitive to the scanning process. In general, one can compute the ensemble intensity distribution for the image dataset; find the probability (ρ) that an intensity resides in the least populated intensity bin; use Equation 2 to construct a plot similar to Figure 3, from which the practical summation limit may be read; and compute the minimum number of voxels by setting P equal to 0.95 (or whatever confidence level is desired) in Equation 3 and numerically solving for v.

CONCLUSION

Each PET-imaged tumor is a single sampling of all radioactivities that are physically and biologically permissible for that particular scanner–tumor combination. Because image heterogeneity statistics accrue manifestations of possibilities, it is the very nature of these statistics to reflect small sample sizes. Thus, inclusion of small tumor volumes necessarily biases tracer uptake heterogeneity studies toward statistically significant differences even when no difference in uptake exists. We have argued that this bias is lessened if all ROIs included in comparative heterogeneity analyses are above a minimum number of voxels. We have described a technique for computing this number that, when applied to our specific ^{18}F-FDG PET image data, yields a minimum comparison volume of 45 cm^3.

DISCLOSURE

The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734. This work was supported by the National Institutes of Health under grant 1R01-CA136931-01A2. No other potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

We thank Richard Laforest for his advice on PET image acquisition and Lauren Tran for carefully reviewing and discussing the manuscript.

REFERENCES

1. Miles KA, Williams RE. Warburg revisited: imaging tumour blood flow and metabolism. *Cancer Imaging.* 2008;8:81–86.
2. Heppner GH. Tumor heterogeneity. *Cancer Res.* 1984;44:2259–2265.
3. O'Sullivan F, Roy S, Eary J. A statistical measure of tissue heterogeneity with application to 3D PET sarcoma data. *Biostatistics.* 2003;4:433–448.
4. Kidd EA, Grigsby PW. Intratumoral metabolic heterogeneity of cervical cancer. *Clin Cancer Res.* 2008;14:5236–5241.
5. El Naqa I, Grigsby P, Apte A, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit.* 2009;42:1162–1171.
6. Tixier F, Le Rest CC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline ^{18}F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med.* 2011;52:369–378.
7. Vriens D, Disselhorst JA, Oyen WJG, de Geus-Oei L-F, Visser EP. Quantitative assessment of heterogeneity in tumor metabolism using FDG-PET. *Int J Radiat Oncol Biol Phys.* 2012;82:e725–e731.
8. Brooks FJ, Grigsby PW. Quantification of heterogeneity observed in medical images. *BMC Med Imaging.* 2013;13:7.
9. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. *J Nucl Med.* 2007;48:932–945.
10. Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in ^{18}F-FDG PET. *J Nucl Med.* 2012;53:693–700.
11. Miller TR, Grigsby PW. Measurement of tumor volume by PET to evaluate prognosis in patients with advanced cervical cancer treated by radiation therapy. *Int J Radiat Oncol Biol Phys.* 2002;53:353–359.
12. Izenman AJ. Recent developments in nonparametric density-estimation. *J Am Stat Assoc.* 1991;86:205–224.
13. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern.* 1973;3:610–621.
14. Snedecor GW, Cochran WG. *Statistical Methods.* Ames, IA: State University Press; 1967:235–238.
15. Fisher RA. *Statistical Methods for Research Workers.* Darien, CT: Hafner Publishing Co.; 1970:83–84.
16. Zar JH. *Biostatistical Analysis.* Upper Saddle River, NJ: Prentice-Hall/Pearson; 2010:470–470.