LETTER TO THE EDITOR

Area under the cumulative SUV-volume histogram is not a viable metric of intratumoral metabolic heterogeneity

Frank J. Brooks

Dear Sir,

There is increasing interest in quantifying the heterogeneity of intratumoral metabolic activity. One currently employed assay of metabolic activity is ^{18}F-fluorodeoxyglucose (FDG) positron emission tomography (PET) where metabolic heterogeneity is seen as intensity variations in grayscale images. Here, the goal of the clinician is to objectively declare one tumor to be more heterogeneous than another and subsequently assess the risk imposed by increased heterogeneity upon ultimate disease outcome. One metric proposed for this purpose is the area under the "cumulative SUV-volume histogram" (CSH) [1, 2]. The independent axis of the CSH is the normalized standard uptake value (SUV) for all predetermined tumor voxels. The dependent axis is the total volume fraction *above* a given SUV threshold. These cumulative curves are computed for each tumor from standard grayscale intensity histograms where the volume fraction *at* a given percent grayscale intensity (or, equivalently, percent SUV) is recorded. The claim is that a lower area under the CSH curve (AUC) corresponds to increased heterogeneity [2]. Note that because this metric is derived solely from the original FDG PET grayscale image histograms, it is an inherently nonspatial metric; the AUC does not depend upon the specific location of grayscale intensities, but instead, only the probability that a given intensity appears *somewhere* within the tumor.

F. J. Brooks (✉)
Department of Radiation Oncology, Washington University
School of Medicine, St. Louis, MO, USA
e-mail: fbrooks@radonc.wustl.edu

Figure 1 shows two distinct intensity histograms. The first, shown in light shading, is a uniform histogram where each possible SUV level occurs with equal probability. This is the case of maximal intratumoral heterogeneity since the greatest number of unique SUV "shades" are available to color the tumor volume. The other histogram, shown in dark shading, contains only two different SUV levels, each occurring with equal probability. This is a case of greatly reduced heterogeneity since there are only two SUV shades with which to color every tumor voxel. Furthermore, those shades are adjacent in the shading scheme; that is, they are as close to being the same shade as is possible. In other words, the first histogram represents many different levels of metabolic activity while the second represents only two levels of (approximately equal) metabolic activity. Another way to see the disparity in total intensity variation is to compute the Shannon informational entropy for each histogram [3]. The Shannon entropy (S) gives an indication of the amount of unique information contained in each histogram and thus gives some indication as to the level of nonspatial variation that is possible to create using only the metabolic levels appearing in the histogram. For the broad, light-shaded histogram, $S = \ln 10 \approx 2.3$ while for the narrow, dark-shaded histogram $S = \ln 2 \approx 0.69$. There is thus a factor of three difference in informational content between the two histograms.

In Fig. 2a, the CSH for the maximally heterogeneous histogram is shown. Because each permissible intensity is represented in the original image histogram, the CSH decreases monotonically. The CSH for the more homogeneous histogram is shown in Fig. 2b. Because, in that case, the original image histogram does not contain every possible intensity shade, the CSH has a pronounced flat run followed
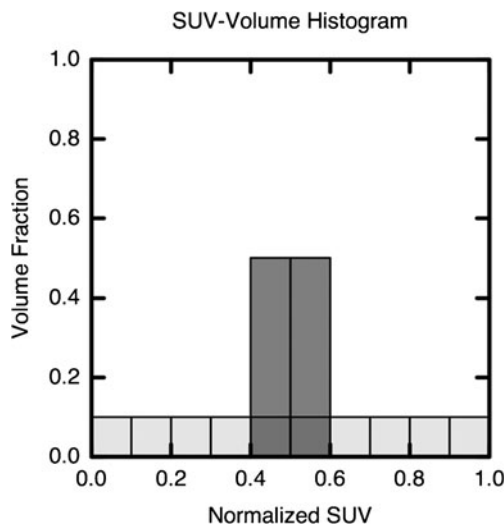
Fig. 1 Two distinct intensity histograms

by a precipitous drop. Still, that CSH is, overall, decreasing. From Fig. 2, it is readily seen that the AUC for each CSH is 5.5. Thus, the AUC for the nearly homogeneous histogram equals the AUC for the maximally heterogeneous histogram. In this context, it is difficult to see how the AUC can be used as a distinguishing quantifier of nonspatial heterogeneity.

There are two simple explanations as to why some researchers have reported statistically significant results after employing the AUC as a heterogeneity metric. First, both the SUV and volume themselves are inherently noisy measurements [4, 5]. Thus, any measure based jointly upon them has an increased uncertainty and is therefore particularly sensitive to the fundamental experimental concerns such as sample size, significant digits, and reproducibility. Second, because the AUC is ultimately dependent upon intensity histograms derived from individual tumors, comparisons of AUCs likely are influenced strongly by tumor volume. To appreciate this, consider the case where tumor biology—and thus, the theoretical intensity histogram—is known a priori to be identical for two, very differently sized tumors. The *measured* histograms can differ solely because a very small tumor necessarily under-samples the theoretical intensity histogram (i.e., the underlying intratumoral biology) while a much larger tumor more closely approximates the same theoretical histogram. In the present work, it has been demonstrated that the area under the CSH cannot distinguish a nearly homogeneous tumor from a maximally heterogeneous tumor. This suggests that any statistically significant result indicating otherwise stems from strictly statistical phenomena such as those described above.
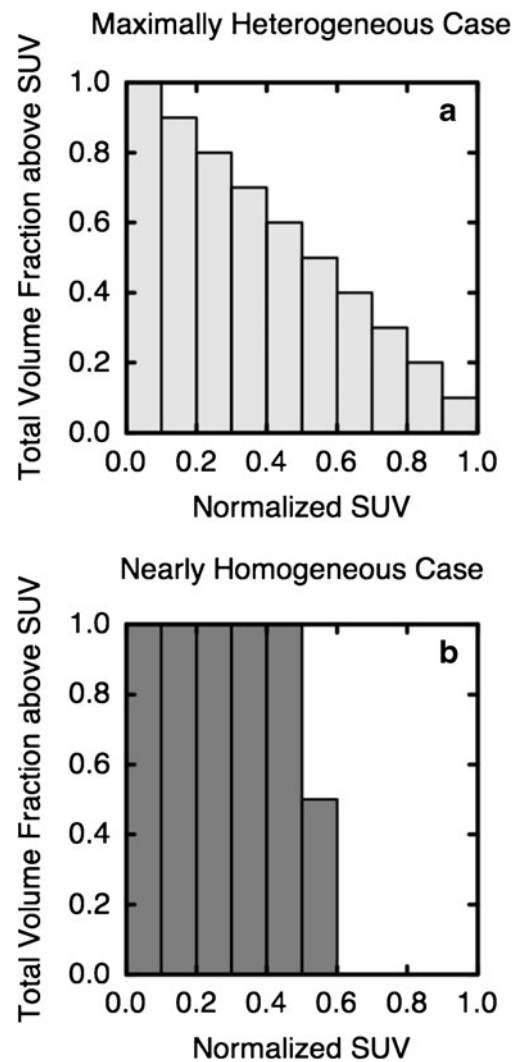

Fig. 2 a CSH for the maximally heterogeneous histogram. b CSH for the more homogeneous histogram

References

1. El Naqa I, Grigsby P, Apte A, Kidd E, Donnelly E, Khullar D, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. Pattern Recognit 2009;42(6):1162–71.
2. van Velden FHP, Cheebsumon P, Yaqub M, Smit EF, Hoekstra OS, Lammertsma AA, et al. Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. Eur J Nucl Med Mol Imaging 2011;38(9):1636–47.
3. Jaynes ET, Bretthorst GL. Probability theory: the logic of science. Cambridge: Cambridge University Press; 2003.
4. 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(9):1519–27.
5. Vanderhoek M, Perlman SB, Jeraj R. Impact of the definition of peak standardized uptake value on quantification of treatment response. J Nucl Med 2012;53(1):4–11.