A novel approach to assess the treatment response using Gaussian random field in PET

Mengdie Wang^{a)}

Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China and Center for Advanced Medical Imaging Science, Division of Nuclear Medicine and Molecular Imaging, Massachusetts General Hospital, Boston, Massachusetts 02114

Ning Guo

Center for Advanced Medical Imaging Science, Division of Nuclear Medicine and Molecular Imaging, Massachusetts General Hospital, Boston, Massachusetts 02114

Guangshu Hu

Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China

Georges El Fakhri

Center for Advanced Medical Imaging Science, Division of Nuclear Medicine and Molecular Imaging, Massachusetts General Hospital, Boston, Massachusetts 02114 and Department of Radiology, Harvard Medical School, Boston, Massachusetts 02115

Hui Zhang^{b)}

Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China

Quanzheng Li^{b)}

Center for Advanced Medical Imaging Science, Division of Nuclear Medicine and Molecular Imaging, Massachusetts General Hospital, Boston, Massachusetts 02114 and Department of Radiology, Harvard Medical School, Boston, Massachusetts 02115

(Received 17 June 2015; revised 29 December 2015; accepted for publication 3 January 2016; published 15 January 2016)

Purpose: The assessment of early therapeutic response to anticancer therapy is vital for treatment planning and patient management in clinic. With the development of personal treatment plan, the early treatment response, especially before any anatomically apparent changes after treatment, becomes urgent need in clinic. Positron emission tomography (PET) imaging serves an important role in clinical oncology for tumor detection, staging, and therapy response assessment. Many studies on therapy response involve interpretation of differences between two PET images, usually in terms of standardized uptake values (SUVs). However, the quantitative accuracy of this measurement is limited. This work proposes a statistically robust approach for therapy response assessment based on Gaussian random field (GRF) to provide a statistically more meaningful scale to evaluate therapy effects.

Methods: The authors propose a new criterion for therapeutic assessment by incorporating image noise into traditional SUV method. An analytical method based on the approximate expressions of the Fisher information matrix was applied to model the variance of individual pixels in reconstructed images. A zero mean unit variance GRF under the null hypothesis (no response to therapy) was obtained by normalizing each pixel of the post-therapy image with the mean and standard deviation of the pretherapy image. The performance of the proposed method was evaluated by Monte Carlo simulation, where XCAT phantoms (128^2 pixels) with lesions of various diameters (2–6 mm), multiple tumor-to-background contrasts (3–10), and different changes in intensity (6.25%–30%) were used. The receiver operating characteristic curves and the corresponding areas under the curve were computed for both the proposed method and the traditional methods whose figure of merit is the percentage change of SUVs. The formula for the false positive rate (FPR) estimation was developed for the proposed therapy response assessment utilizing local average method based on random field. The accuracy of the estimation was validated in terms of Euler distance and correlation coefficient.

Results: It is shown that the performance of therapy response assessment is significantly improved by the introduction of variance with a higher area under the curve (97.3%) than SUVmean (91.4%) and SUVmax (82.0%). In addition, the FPR estimation serves as a good prediction for the specificity of the proposed method, consistent with simulation outcome with ~ 1 correlation coefficient.

Conclusions: In this work, the authors developed a method to evaluate therapy response from PET images, which were modeled as Gaussian random field. The digital phantom simulations demonstrated that the proposed method achieved a large reduction in statistical variability through incorporating knowledge of the variance of the original Gaussian random field. The proposed method has the potential to enable prediction of early treatment response and shows promise for application to

clinical practice. In future work, the authors will report on the robustness of the estimation theory for application to clinical practice of therapy response evaluation, which pertains to binary discrimination tasks at a fixed location in the image such as detection of small and weak lesion. © 2016 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4939879>]

Key words: therapeutic assessment, Gaussian random field, PET imaging

1. INTRODUCTION

For decades, effectiveness of anticancer therapy is mostly evaluated using morphological changes as assessed by anatomical imaging techniques including computed tomography (CT) and magnetic resonance imaging (MRI). However, these techniques rely on changes in tumor size which may take weeks or even months to know the success or failure of the therapy. Thus, traditional criteria for efficacy of treatment, including the WHO criteria,^{1,2} RECIST,^{3,4} and EASL criteria,⁵ are based solely on systematic assessments of tumor size. Unfortunately, the morphological changes are usually not early enough to identify patients who do not respond to treatments such as chemotherapy. And it demonstrated some restrictions in differentiating post-treatment changes (fibrosis scar) from residual disease. It is important to identify these patients early in the course of treatment to save them from unnecessary side effects and costs. Moreover, alternative therapeutic strategies could be proposed to the patients in time to achieve curative effects. As a result, early understanding and assessment of therapeutic success play an important role in determining whether to continue or change treatment.^{6,7}

Positron emission tomography (PET) combined with ¹⁸F-fluorodeoxyglucose (¹⁸F-FDG) is one of the promising imaging techniques to provide measurements of tumor metabolism as an early response marker to treatment.^{8–10} In clinical practice and research studies, the interpretation of differences between before- and after-treatment PET images, usually measured using standardized uptake values (SUVs), is used to predict or evaluate the clinical outcome.^{11,12} Such analysis is widely used in nowadays because SUV is easy to obtain for everyday scans. However, such a convenient approach can only provide limited information about tumor with unsatisfied accuracy and reproducibility.^{11,13–20} Among the various limiting factors, Poisson-distributed random counting noise in the raw data is particularly important, leading to a high susceptibility to precision loss.²¹ Developing an advanced statistical methodology with more accurate information from PET imaging data becomes urgently needed.

When the image is reconstructed by the iterative, nonlinear, expectation maximization (EM) method, the uncertainty of SUV can be calculated by modeling the bias and noise characteristics of SUV.^{22,23} Inspired by this approach, a new criterion of therapeutic assessment could be derived incorporating image noise. In this study, we take into consideration the effect of variation caused by random counting noise and propose a new approach for therapy response assessment based on Gaussian random field (GRF). Due to the variation introduced by the noise, each pixel of the reconstructed image can be regarded as a random variable and the reconstructed tomographic image is often approximately

multivariate normal.^{24–26} We consider the spatially correlated random variables in a specific PET image comprising a Gaussian random field. The variance of this random field can be approximately estimated from the noisy data for penalized likelihood reconstruction algorithm.^{27,28} By introducing the image noise model, a statistically robust criterion is then derived. The false positive rate (FPR) of this proposed method is estimated based on the local average theory of random field and can be utilized to determine the appropriate threshold for therapy response assessment.

The proposed approach aims at early response assessment through pretreatment and post-treatment images in which the change of tumor size is not pronounced. We assumed that the pretreatment and post-treatment images could be registered, or the target organ (e.g., liver) could be segmented and registered, to make sure that the only difference observed in the distribution of pixel intensity is treatment-related. The performance of the proposed method is compared with standardized uptake values (SUVmean- and SUVmax) based assessment in terms of receiver operating characteristic (ROC) and area under the curve (AUC). Euler distance (ED) and correlation coefficient (CC) are utilized to measure the accuracy of FPR estimation. The proposed method is evaluated by Monte Carlo (MC) simulation using 2D XCAT phantom.

2. METHODS

We first established a new criterion of treatment response evaluation based on GRF. We then investigated the statistical property of the proposed criterion using local average theory of random field. The resulting approximation of FPR was computed to optimize the threshold in the criterion and control the error rate in the evaluation. In this study, we only considered the lesions with approximately homogeneous region of interest (ROI) in this study; the case of inhomogeneous tumor is discussed in Sec. 5.

2.A. New criterion for therapeutic assessment

First, the pretreatment and post-treatment images were reconstructed using the penalized likelihood algorithm. Similar to the approximate expression proposed by Qi,²⁸ we estimated the pixelwise variance of the pretreatment image. Then, the post-treatment image was normalized to a zero mean unit variance GRF using the estimated standard deviation under null hypothesis (no response to therapy). At last, the normalized GRF was averaged over the ROI that comprises multiple pixels to calculate a statistical decision score T . We compared this score to a threshold to decide on the presence or

absence of a response. The performance of this method was evaluated by ROC curve and the AUC, and compared with that of traditional method.

2.A.1. Image reconstruction

In this project, the PET images were reconstructed by maximum *a posteriori* (MAP) algorithm. Given the log-likelihood function $L(y|x)$ and the log of prior distribution $U(x)$, the MAP estimator can be written as

$$\hat{x} = \arg \max_{x \geq 0} L(y|x) - \beta U(x), \quad (1)$$

where y is the measured sinogram and x is the image. For the proposed method, we consider a Gaussian Markov random field prior for the image with Gibbs energy that can be written as $U(x) = (1/2) \sum_j \sum_{k \in N_j} \omega_{jk} \psi(x_j - x_k)$, where N_j is the neighborhood of pixels near pixel j and the potential function $\psi(x)$ is $x^2/2$ for quadratic penalty, and $\omega_{jk} = 1$ for horizontal and vertical neighboring pixels, $\omega_{jk} = 1/\sqrt{2}$ for diagonal neighboring pixels, and $\omega_{jk} = 0$ otherwise. The parameter β determines the relative influence of the prior and likelihood terms. Different β values, ranging from 2.0×10^{-3} to 20, were applied to the reconstruction to optimize β for assessing treatment response. For both SUV and the proposed method, the pretreatment and post-treatment images were reconstructed with fixed $\beta = 0.01$. Equation (1) was solved by a preconditioned gradient ascent algorithm, which can be written as^{29,30}

$$\hat{x}^{k+1} = \hat{x}^k + \alpha C^k(\hat{x}^k) [\nabla_x L(y|x) - \nabla_x \beta U(x)]_{x=\hat{x}^k}, \quad (2)$$

where $\alpha > 0$ is a fixed step size and $C^k(\hat{x}^k)$ is a preconditioner matrix. ∇_x denotes gradient operation with respect to x . The optimization was done with 60 iterations. Two representative images of pretreatment and post-treatment are shown in Fig. 1. As mentioned above, the therapy assessment is usually performed by determining the percentage change in SUVmean or SUVmax between two measurements as $(\text{SUV}_{\text{post}} - \text{SUV}_{\text{pre}}) / \text{SUV}_{\text{pre}}$.

After image reconstruction, we computed a normalized GRF as the difference of the reconstructed post-treatment image and the mean pretreatment image normalized by the

standard deviation of pretreatment measurement σ_{pre} ,

$$t = -(\hat{x}_{\text{post}} - \bar{\hat{x}}_{\text{pre}}) / \sigma_{\text{pre}}, \quad (3)$$

where $\hat{x}_{\text{post}}$ and $\bar{\hat{x}}_{\text{pre}}$ are the estimations of true image after therapy and the expectation of the estimation of true image before therapy, respectively. As we know, in simulation study, either noiseless data or sufficient numbers of independent data sets are available to determine the expected value of the pretreatment true image. However, to make our simulation study more clinical-oriented, we used an oversmoothed MAP reconstruction to approximate the mean as in practice situations, which is believed to be a more accurate estimator of the mean than the original reconstruction.²⁸ In the oversmoothed MAP reconstruction, the smooth parameter is equal to 1. t was defined as an intermediate criterion, whose properties will be further discussed in Secs. 2.A.2 and 2.A.3. Note that in this work, the final score T (see Sec. 2.A.3) is determined by the average of the paired tumor information $(x_{\text{post}} - x_{\text{pre}}) / \sigma_{\text{pre}}$, which indicates the “overall change” and therefore alleviates the influence of the resolution of the approximate mean estimator on final criterion.

2.A.2. Fisher information matrix (FIM)-based variance evaluation and Gaussian approximation

σ_{pre} was introduced to quantify the image noise in the new criterion. An analytical method based on approximate expressions of FIM was used to predict the variance of individual pixels of pretreatment image. Previous expression of image variance proposed by Qi and Leahy²⁸ was adopted in our specific case with minor modification. To be specific, the FIM was simplified by taking the system response matrix as an inseparable parameter,

$$F \approx D[k_j] P' P D[k_j], \quad k_j = \sqrt{\frac{\sum_i p_{ij} / y_i}{\sum_i p_{ij}}}, \quad (4)$$

where p_{ij} is the (i,j) th element of the system response matrix P , $D[k_j]$ is a diagonal matrix with diagonal elements k_j , and y_i is i th element of the measured sinogram. It is worth noting that since no attenuation was simulated in this paper, the probability matrix P is defined as the geometric projection matrix with each element (i,j) equal to the probability that a photon pair produced in voxel j reaches the front faces of the detector pair i in the absence of attenuation. In practice, more factors including attenuation effect are taken into consideration and the probability matrix can be written in the form $P = P_{\text{sens}} P_{\text{blur}} P_{\text{geo}} P_{\text{att}} P_{\text{positron}}$,¹⁶ where P_{sens} , P_{blur} , P_{geo} , and P_{att} correspond to detector normalization, sinogram blurring, geometric projection, and attenuation effect, respectively. The variance of pixel j is then calculated by

$$\text{var}_j \approx e_j^T [F(j) + \beta R(j)]^{-1} F(j) [F(j) + \beta R(j)]^{-1} e_j, \quad (5)$$

where $F(j)$ is a block Toeplitz matrix formed from the j th column of the FIM, $R(j)$ is the second derivative of the prior energy function $U(x)$, and e_j is j th unit vector. Note that

FIG. 1. True phantoms and reconstructed images. Top: true phantom images of pretreatment (left) and post-treatment (right); bottom: reconstructed images.

Eq. (5) does not account for the non-negativity constraint used in MAP reconstruction. Therefore, the pixelwise variances were then modified by the assumed truncated Gaussian model as detailed below.²⁸

Let u and u' be the mean of the origin and truncated Gaussian distribution, respectively, the variance of the truncated Gaussian distribution var' is given by

$$\text{var}' = u \sqrt{\frac{\text{var}}{2\pi}} e^{-(u^2/2\text{var})} + \frac{1}{2}(u^2 + \text{var}) \left[1 + \text{erf}\left(\frac{u}{\sqrt{2\text{var}}}\right) \right] - u'^2, \quad (6)$$

$$u' = \sqrt{\frac{\text{var}}{2\pi}} e^{-(u^2/2\text{var})} + \frac{u}{2} \left[1 + \text{erf}\left(\frac{u}{\sqrt{2\text{var}}}\right) \right], \quad (7)$$

where erf is the error function. Since u is unknown, it is not a straightforward work to compute var' . In fact, the mean of the truncated Gaussian distribution u' is the mean of the corresponding voxel in the MAP reconstructions, which we already estimated in Sec. 2.A.1 by an oversmoothed MAP reconstruction. According to Eqs. (6) and (7), $u'/\sqrt{\text{var}} = f(\zeta)$ and $\text{var}'/\text{var} = g(\zeta)$, where $\zeta = u/\sqrt{\text{var}}$,

$$f(\zeta) = \frac{1}{\sqrt{2\pi}} e^{-(\zeta^2/2)} + \frac{\zeta}{2} \left[1 + \text{erf}\left(\frac{\zeta}{\sqrt{2}}\right) \right] \quad (8)$$

and

$$g(\zeta) = \frac{\zeta}{\sqrt{2\pi}} e^{-(\zeta^2/2)} + \frac{1}{2}(\zeta^2 + 1) \left[1 + \text{erf}\left(\frac{\zeta}{\sqrt{2}}\right) \right] - f(\zeta)^2. \quad (9)$$

With the mean u' and the unconstrained variance var , we can first compute $u'/\sqrt{\text{var}}$ to find ζ and use Eq. (9) to get $g(\zeta)$. Finally, the variance of truncated Gaussian distribution can be calculated by $\text{var}' = g(\zeta)\text{var}$.

The accuracy of variance estimation was validated using MC simulations. The distribution of individual pixel of the liver phantom was calculated. The distributions of pixel A and pixel B in Fig. 1, overlaid with Gaussian distributions fitted to the histogram, are shown in Fig. 2. The sample histograms and the truncated Gaussian match well. Figure 3 shows the standard deviation images that are obtained by simulation and our approximation; they match with each other too.

FIG. 2. Histogram and fitted Gaussian probability density function (PDF) of a single pixel.

FIG. 3. Variance images using (a) Monte Carlo simulation from 1000 reconstructions and (b) the theoretical approximation.

2.A.3. New criterion and dichotomous assessment

After the variance of pretreatment image was computed, a map of statistics t was obtained by normalizing each pixel of post-treatment image with the mean and standard deviation of the pretreatment image. We then formulated the treatment response evaluation as a hypothesis test; therapy response absent or therapy response present is denoted as H_0 and H_1 , respectively. Under the null hypothesis (H_0) that the therapy has no effect, the t map is a zero mean unit variance GRF or $t \sim N(0,1)$.^{31,32} The average value of the target ROI in t map was calculated and the final criterion T is obtained, which was essentially a Gaussian random variable as well. The decision between the two hypotheses (H_0 : nonresponsive or H_1 : responsive) was determined by

$$H_0: T < u,$$

$$H_1: T > u,$$

$$T = \frac{1}{N} \sum_{j \in \text{ROI}} t_j, \quad (10)$$

where T is the final score for a targeted ROI (tumor). To evaluate the performances at different scenarios (lesion-to-background contrast ratio, lesion diameter, and percentage change in activity concentration), we calculated the ROC curve for 11 simulated lesions separately.

For the completeness of this study, we also evaluated the overall performance of the proposed criterion under multiple scenarios (or tumor ROIs), since in some cases, there are multiple tumors and one single criterion is preferred. Let T_i be the score for the tumor in the i th scenario. The sensitivity (true positive rate, TPR) and specificity (1-FPR) of the proposed criterion under multiple scenarios can be defined as follows:

$$\text{Sensitivity} = \text{TPR} = P(T_{\text{test}} > u | H_1), \quad (11)$$

$$\text{Specificity} = 1 - \text{FPR} = 1 - P(T_{\text{test}} > u | H_0), \quad (12)$$

where $T_{\text{test}} = \{T_i | i = 1, 2, \dots, M\}$. ROC curves can be generated by using different threshold value u in Monte Carlo simulations (see Sec. 3). The AUCs were then calculated to compare the performances of the traditional methods (percentage change of SUVmean and SUVmax) and the proposed method. As a potential criterion, the peak value of target ROI in t map, denoted as GRFmax, was calculated as well in Sec. 3 for comparison. Note that the multiple scenarios are considered only in performance evaluation and the method we proposed to determine the threshold is used for one tumor ROI at a time and will be discussed in detail in Secs. 2.B and 4.

2.B. FPR thresholding and cutoff

ROC curve is an excellent tool to evaluate the average performance of the proposed method in terms of sensitivity and specificity. However, in a real clinical application, we usually need to apply an optimal threshold for a specific case. The threshold could be optimized using many different error rates, and here, we use the FPR. FPR represents the probability of having mean value in GRF t (within ROI) which exceeds a given threshold u under the null hypothesis. Therefore, the statistical property of criterion T is required to estimate the FPR. We thus applied the “local average theory” to estimate the statistics of criterion T , inspiring by the local average theory of random fields that primarily used in the analysis of stochastic finite element model.³³ In Subsection 2.B.1, the 2D local average process will be briefly reviewed. As long as the variance of criterion T is computed, a proper threshold can be determined while controlling FPR.

The block diagram of FPR controlled treatment response evaluation is illustrated in Fig. 4. In addition to single pixel variance, the analysis of “local average” also requires the correlation of pixels in the target ROI. In principle, we could compute the covariance of the voxels in the targeted ROI using FIM-based approximation similar to Sec. 2.A.2. However, due to the computational cost of covariance matrix, we only calculated one column of the covariance matrix, assuming that the covariance structure of neighboring pixels is approximately the same within the ROI. We further fit this column of covariance matrix with Gaussian function. We then computed the statistics of criterion T using local average theory, given the fact that the joint distribution of pixels in a target ROI could be approximated by a joint Gaussian distribution. Once the statistics of criterion T is known, it is easy to choose proper threshold by controlling the FPR. Although we made a series approximations to simplify the computation, we believe that these approximations are reasonable and will validate them in Sec. 4.A using Monte Carlo simulation. In the simulation, spheres were simulated to mimic real tumors but the ROIs were not necessarily spherical. For the sake of simplicity, this study was carried out based on rectangular-shaped ROI.

2.B.1. 2D local average

In this subsection, we will briefly introduce the local average theory in the context of GRF, which provides an elegant approach to compute the first and second momenta

of the summation of pixel values in ROI. Given a 2D random field $R(x,y)$, the local integration over a rectangular area I_A is also a random field,

$$I_A(c_1, c_2) \equiv \int_{c_1-L_1/2}^{c_1+L_1/2} \int_{c_2-L_2/2}^{c_2+L_2/2} R(x,y) dx dy. \quad (13)$$

The rectangular window is centered at (c_1, c_2) , and its sides (having length L_1 and L_2) remain parallel to the respective coordinate axes. The 2D local average is yielded after dividing $I_A(c_1, c_2)$ by A , the area of the rectangular window,

$$R_A(c_1, c_2) \equiv \frac{1}{A} I_A(c_1, c_2). \quad (14)$$

If the random field $R(x,y)$ is homogeneous, the random variable $R_A(c_1, c_2)$ can be characterized by the statistical property of $R(x,y)$ as follows:

$$E_{R_A(c_1, c_2)} = m, \quad \text{var}_{R_A(c_1, c_2)} = \sigma^2 \gamma(L_1, L_2), \quad (15)$$

$$\gamma(L_1, L_2) = \frac{1}{L_1 L_2} \int_{-L_1}^{L_1} \int_{-L_2}^{L_2} \left(1 - \frac{|\xi|}{L_1}\right) \left(1 - \frac{|\eta|}{L_2}\right) \rho(\xi, \eta) d\xi d\eta, \quad (16)$$

where m and σ^2 are the expectation and variance of $R(x,y)$, respectively. Function $\rho(\xi, \eta)$ of the two variables ξ and η is the 2D correlation function of $R(x,y)$. If the structure of $\rho(\xi, \eta)$ is separable, or

$$\rho(\xi, \eta) = \rho_1(\xi) \rho_2(\eta), \quad (17)$$

where $\rho_1(\xi)$ and $\rho_2(\eta)$ are 1D correlation functions, and $\text{var}_{R_A(c_1, c_2)}$ mainly depends on the 1D correlation function. The Gaussian structure for 1D correlation function is as follows:

$$\rho(\xi) = e^{-\frac{\xi^2}{\theta^2}}, \quad (18)$$

where θ is the shape parameter, which can be calculated by fitting the column of the covariance matrix with the correlation functions. More mathematical details of local average theory of random field can be found in Vanmarcke's work.³⁴

2.B.2. Correlation function modeling

Based on our experience, we chose Gaussian function in Eq. (18) to model the correlation function, which can be approximated by $\text{cov}_j \approx [F(j) + \beta R(j)]^{-1} F(j) [F(j) + \beta R(j)]^{-1} e_j$.²⁸ Thus, the parameter θ in Eq. (18) can be determined by fitting the 1D correlation function with

FIG. 4. Block diagram for estimating FPR for therapy response assessment.

FIG. 5. 1D (a) and 2D (b) plots for estimated correlation functions.

Gaussian function using least square approximation. The 1D and 2D correlation functions of a specific pixel were drawn to fit the Gaussian structure, and the results demonstrate a good fit as shown in Fig. 5. Note that $R(x, y)$ and $R_A(c_1, c_2)$ refer to intermediate criterion t and final criterion T , respectively, when we apply the local average theory to our specific study. Furthermore, the correlation function within ROI of the reconstructed image $\hat{x}_{\text{post}}$ is equivalent to that of t , according to Eq. (3).

2.B.3. False positive rate estimation

In the 2D local average theory, we assumed that all the pixels in the ROI share the same correlation function. With the size of ROI become larger, the accuracy of this assumption will decrease. Therefore, instead of carrying out the multiple hypothesis test in Sec. 2.A.3, we only focused on a specific tumor ROI at one time. The equations of FPR could be modified slightly as

$$\text{FPR} = P(T > u | H_0) = P(T_{\text{null}} > u), \quad (19)$$

where T_{null} is the final score of a single tumor with null response. According to Eq. (19), given that $m = 0$, $\sigma^2 = 1$ holds for H_0 hypothesis, it is easy to obtain the statistical property of T_{null} ,

$$E_{T_{\text{null}}} = 0, \quad \text{var}_{T_{\text{null}}} = \gamma(L_1, L_2). \quad (20)$$

And the false positive rate as a function of threshold u can be estimated by

$$\begin{aligned} \text{FPR} &= P(T > u | H_0) \\ &= \int_u^{+\infty} \frac{1}{\sqrt{2\pi\text{var}_{T_{\text{null}}}}} \exp\left(-\frac{(x - E_{T_{\text{null}}})^2}{2\text{var}_{T_{\text{null}}}}\right) dx \\ &= \int_u^{+\infty} \frac{1}{\sqrt{2\pi\gamma(L_1, L_2)}} \exp\left(-\frac{x^2}{2\gamma(L_1, L_2)}\right) dx. \end{aligned} \quad (21)$$

It is intractable to obtain the integral in Eq. (21) analytically. However, this can be computed by numerical integration algorithm and was tabulated in this project.

Given that the intensity of the lesion could even increase after the treatment when the disease progresses under treatment, negative value in criterion T could be observed. In this case, an additional hypothesis test can be performed to make a decision between nonresponder and negative responder. A further discussion is included in Sec. 5. Here, we considered both negative response and null response as therapy response absent for simplicity.

3. PERFORMANCE EVALUATION

Simulations were designed to compare the traditional methods using SUV with the proposed method. We characterize their performances on predicting therapy response using ROC and corresponding AUC.

3.A. Monte Carlo simulation

To mimic the realistic therapeutic response, XCAT 2D phantoms (128^2 pixels) with lesions of various diameters (2–6 mm), multiple tumor-to-background contrasts (3–10), and different changes in intensity (6.25%–30%) were used (Fig. 1). For simplicity, negative change was not considered in this work. The multiple scenarios of therapy response applied in this simulation were summarized in Table I, corresponding to the labeled tumors in Fig. 1. Twenty two scenarios of responses were simulated (11 with response, and 11 with null response) and 1000 Poisson data sets with pseudorandom noises were generated for each scenario. A total amount of 22 000 simulated studies were generated in which 11 000

TABLE I. Lesion data sets.

Lesion #	1	2	3	4	5	6	7	8	9	10	11
Radius (mm)	2	1	3	2	2	1	3	1	2	1	1
Pretreatment ^a	7.5	7.5	7.5	7.5	12	12	12	4.5	4.5	7.5	7.5
Post-treatment ^b	5.25	6	6.75	6.75	9.75	10.5	11.25	3	3.75	5.25	6

^aLesion to background ratio of pretreatment.

^bLesion to background ratio of post-treatment.

Fig. 6. t maps for therapy response absent (left) and present (right).

studies have no effect (nonresponsive), which share the same pretreatment images as the responsive studies. The total counts are 1.3×10^6 and 1.4×10^6 for responsive and nonresponsive, respectively. Attenuation and scatter effects were not taken into account in the simulations. The ROIs used in this section were spheres containing the whole tumor objects.

3.B. Evaluation of the new criterion

By applying the proposed method to 22 000 studies, 2000 t maps were generated by Eq. (3), and each map involved 11 ROIs as shown in Fig. 6. As a result of the proposed method, 22 000 T scores were calculated by Eq. (10).

In order to see the effect of lesion-to-background contrast ratio, lesion diameter, and percentage change in activity concentration on the evaluation performance, we also applied the proposed method to the 11 simulated lesions separately. The AUC values of ROC curve for 11 scenarios are presented in Table II, with the AUC of SUVmean and SUVmax for comparison. As we can see (according to Table II), the proposed method performs well (with AUC larger than 95%) for all tumor ROIs except for lesion #2, which has both small diameter and low tumor-to-background contrast ratio. It is also worth mentioning that the performance of SUVmean and SUVmax substantially degrades in cases where size of lesion reduces (lesion #2 and #6) and the percentage change (lesion #4 and #7) or tumor-to-background contrast ratio (lesion #2 and #11) decreases, while, in general, the proposed method performs more robustly in various scenarios than SUVmean and SUVmax for treatment response evaluation.

We also evaluated the overall performance of the proposed criterion under multiple scenarios. Sensitivity and specificity of the criterion were then evaluated by Eqs. (11) and (12), respectively, for different threshold value u . ROC curves as well as AUCs were calculated for the proposed criterion T , GRFmax, SUVmean, and SUVmax (Fig. 7). We compared the performance of therapy response assessment with the above

Fig. 7. ROC curves for the proposed method (GRF), GRFmax, SUVmean, and SUVmax method.

four methods on the same scale. The results show that the proposed method offers a substantial improvement with a higher AUC (97.3%) compared with traditional SUVmean (91.4%) and SUVmax (82.0%). And the performance of the proposed method is reasonably superior to GRFmax (93.6%), because the noise was efficiently reduced by taking average of the pixels when generating criterion T , similar to SUVmean. However, both of the proposed criterion and GRFmax perform better than the traditional methods. This can be expected from the implementation of the strategy. Although we normalized the change in uptake values for both proposed and the traditional methods, the reference of normalization was different. Other than using the mean (SUVmean) or max (SUVmax) intensity of pretherapy image which is applied in the traditional method, the proposed method used the pixel-by-pixel variance derived from pretreatment GRF as the denominator which considers the random variation caused by Poisson-distributed random noise in the raw data. Consequently, this proposed normalization leads to a better performance in the sensitivity and specificity than SUVmean- and SUVmax-based methods, indicating the potential to offer a statistically significant approach to the assessment of therapy response.

4. FALSE POSITIVE RATE ESTIMATION

4.A. Statistical property study

Using local average theory, the false positive rate of the proposed method can be estimated by the approximation given in Sec. 2.B.3. To validate the accuracy of the proposed method, the standard deviation of T_{null} and FPR was estimated by Eqs. (20) and (21), respectively, and compared with simulation outcome. According to Sec. 2.A.1, the standard deviation of T_{null} primarily depends on the size of ROI. To investigate the effect of size of ROI on the estimation, we simulated 13 ROIs with different sizes and calculated T_{null} for each ROI. The result is demonstrated in Fig. 8. It is demonstrated

TABLE II. AUC evaluation for single tumor ROI.

Lesion #	1	2	3	4	5	6	7	8	9	10	11
GRF (%)	99.8	83.3	96.4	97.8	99.3	95.9	99.3	95.6	97.6	99.5	96.7
SUVmean (%)	98.3	80.9	94.4	70.6	97.8	81.1	76.7	94.7	93.7	96.3	95.6
SUVmax (%)	96.5	79.5	84.0	66.9	85.7	79.6	67.6	93.3	88.7	96.0	88.8

Fig. 8. Plots of the standard deviation of estimated and simulated T generated by different ROI sizes.

that the estimated standard deviation of T_{null} fits well with the simulation results with small ROI. However, gradually increasing disparity was observed when the ROI becomes larger, indicating the negative effect of large ROI size on the accuracy of T_{null} estimation.

Since FPR estimation was derived directly from the statistical property of T_{null} , an accurate estimation of the standard deviation T_{null} will guarantee an accurate prediction of the FPR. Figure 9 shows the estimated cumulative distribution functions of different normal distributions (solid line) and the FPR computed from simulation (dots) for 13 different ROI areas as a function of threshold. As a result, the curves of estimated and simulated FPR fitted well with each other. It indicated that the proposed therapeutic assessment achieved better performance with small tumor ROI under a given threshold than large one.

4.B. Error assessment

To quantitatively assess the accuracy of FPR estimation, metrics like root mean squared error (RMSE) and CC between the estimated and simulated FPR were evaluated under different sizes of lesion ROI. To be specific, the Euler distance

Fig. 9. Plots of the estimated and simulated FPR for different lengths of ROI area.

Fig. 10. Plots of Euler distance and correlation coefficient for FPR estimation as a function of the total number of pixels in ROI.

was computed as $\text{RMSE} = \sqrt{(1/N) \sum (\widehat{\text{FPR}}_i - \text{FPR}_i)^2}$ and the correlation coefficient was defined as $c = \frac{\sum (\widehat{\text{FPR}}_i - E(\widehat{\text{FPR}}))(\text{FPR}_i - E(\text{FPR}))}{\sqrt{\sum (\widehat{\text{FPR}}_i - E(\widehat{\text{FPR}}))^2 \sum (\text{FPR}_i - E(\text{FPR}))^2}}$, where FPR_i and $\widehat{\text{FPR}}_i$ are the simulated and estimated FPRs under the i th threshold, respectively, and N is the total number of the thresholds. $\widehat{\text{FPR}}_i$ was calculated numerically from the exact analytical expressions [Eq. (21)]. E is the average of FPR taking over all the thresholds. The results of the accuracy assessment were presented in Fig. 10, showing that our FPR estimation for the proposed method has a small root mean squared error (less than 0.04) and a very good correlation coefficient (~ 1) with the FPR generated by Monte Carlo for overall thresholds. The observed RMSE is computed from the equation and predicted RMSE is the plot fitted with the observed one, indicating the trend of RMSE with increasing or decreasing ROI size. It can be observed that RMSE and CC were monotonically increasing and decreasing functions of the number of pixels in ROI, respectively, indicating better estimations with smaller tumors.

5. DISCUSSION

We have demonstrated the statistical advantage of the proposed method that incorporates knowledge of the statistics of image and random field. More importantly, in the proposed method, threshold can be determined for a given ROI to control the FPR, while in the traditional methods using SUV, only *ad hoc* empirical threshold can be applied.

The method introduced here is a theoretical framework, which relies on two assumptions. First, it requires that the normalized Gaussian random field is homogeneous inside a lesion ROI. Second, the pretreatment and post-treatment images were assumed to have been registered before evaluation or the only difference in uptake distribution is treatment-related. For the first assumption, we are concerned with the liver lesion with uniformly distributed activity in ROI, which allowed the assumption of approximate homogeneity. For the nonuniform lesion, we could apply the mixture Gaussian

random field in same framework, which is included in our future work but not further investigated in this paper. Due to the patient motion as well as the changes caused by therapy, image registration should be conducted before treatment evaluation so that the spatial information of the tumor objects pretherapy and post-therapy coincides. Although the spatial disparity between the pretreatment and post-treatment images is not considered in this proposed theoretical method, we will account for it in the future clinical study by applying appropriate registration method. So far a number of medical image registration methods^{35,36} have been proposed for routine clinical use including treatment evaluation.^{37–39} For example, the nonrigid customized Image Registration Toolkit (IRTK) is developed and evaluated, introducing an average target registration error less than 5 mm in all directions, which is comparable to the PET spatial resolution.³⁵ Note that as a quickest and the most intuitive method, we assume that the correlation function of the data is Gaussian although it is not always the case due to the negative side-lobes in the xy plane.⁴⁰ In future work, we will improve the accuracy of variance estimation by either defining a non-Gaussian function⁴¹ or using the exact correlation function instead of Gaussian assumption to achieve better results.

Although our proposed method is designed for early response assessment while the change in the ROI is small and possibly buried in the noise, it also works for the case in which significant change of the average uptake in the given ROI is observed due to the tumor shrinkage caused by therapy.⁴² We anticipate our method will still demonstrate superior performance in this case, although the relative advantage maybe smaller.

Given the negative value in criterion T may result from tumor that does not respond or respond negatively, two hypothesis tests can be sequentially performed to take that into account. Hypothesis test I is between H_0 : therapy response absent (null and negative response) and H_1 : therapy response present (positive response). Hypothesis test II is between H'_0 : therapy response absent (null response) and H'_1 : therapy response present (positive and negative response). The first hypothesis test has already been used in this study. In the second hypothesis test, when $|z| = (|T|/\sqrt{\text{var}_{T_{\text{null}}}}) \leq z_{\alpha/2}$, H'_0 is accepted, and vice versa, where α is an acceptable significance level (usually 0.05). By performing the two hypothesis tests, the image can be categorized into three classes (1—therapy response absent, 2—response present positively, 3—response present negatively).

It is also worth mentioning that in Sec. 2.B, our purpose is to validate the theory. Therefore, we use the rectangular ROI. In the future, we will instead use appropriate sphere ROI to calculate the new criterion.

In Sec. 2.B, although the comparison of estimated and simulated T variances indicated high accuracy of the proposed Eq. (20), discrepancy between two plots can be observed in Fig. 8. This error can be attributed to multiple factors. First of all, according to Eq. (15), the function γ is estimated by correlation function, which is fitted with Gaussian distribution. However, the neighboring pixels have negative correlations and Gaussian is not accurate enough to determine the variance.

Besides, the correlation function itself is approximated by $\text{cov}_j \approx [F(j) + \beta R(j)]^{-1} F(j) [F(j) + \beta R(j)]^{-1} e_j$, in which either inaccurate detection probability matrix P or improper selection of penalty types may adversely affect the estimation. Moreover, the estimation errors of Eq. (5) would increase when the counts level is low. The quantitative analysis of the error caused by each factor will be investigated in future to help further establish the validity/improvement of the proposed approach. It is also worth noting that Eq. (21) holds for T_{null} with Gaussian distribution only. Since our approach was aimed at various clinical scenarios, for example, low dose PET scan, where the counts level is relatively low, the negative case has to be considered in our method. For the completeness of the work, we will investigate the estimation of false positive rate for truncated Gaussian in the future.

In future work, we will investigate the robustness of our estimation theory using clinical data, which pertains to binary discrimination tasks at a fixed location in the image. In addition, the actual tumors tend to have an irregular shape. Therefore, we will further extend the proposed approach to mixture Gaussian random field that works for tumors with complex shape.⁴³

6. CONCLUSION

In this work, we developed a method to evaluate therapy response using pretreatment and post-treatment PET images. This method considers the treatment response evaluation as a hypothesis test based on Gaussian random field. Furthermore, we proposed a method to compute the optimal threshold of hypothesis test to achieve a desirable FPR. Monte Carlo simulations with XCAT phantom demonstrated that the proposed method offered substantial improvement on AUC over the traditional methods with a 97.3% compared to 91.4% (SUVmean) and 82.0% (SUVmax). Moreover, the study of individual tumors indicates that traditional methods SUVmean and SUVmax are more susceptible to the lesion parameters (lesion-to-background contrast ration, lesion diameter, and percentage change) than the proposed method, which performed stably at a high level (with 95% AUC for 90.9% cases) for treatment response evaluation. It also showed the high consistency of the estimated FPR with the simulated FPR through the correlation coefficient (~ 1) and Euler distance. The focus and contribution of this work is to propose and validate a statistical framework for the treatment evaluation using PET. The limitation is that the complex shape of tumor and introtumor heterogeneity is not considered in the current calculation and simulations. In the future, we are going to extend Gaussian random field to mixture Gaussian random field in our framework to take into account the shape and heterogeneity of the tumor ROI.

ACKNOWLEDGMENTS

This work was supported in part by the scholarship from China Scholarship Council (CSC, No. 201306210270) and NIBIB No. R01EB013293.

- ^aThis research was performed at Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114.
- ^bAuthors to whom correspondence should be addressed. Electronic addresses: hzhang@mail.tsinghua.edu.cn and li.quanzheng@mgh.harvard.edu
- ¹R. Hunter, "WHO handbook for reporting results of cancer treatment," *Int. J. Radiat. Biol.* **38**(4), 481 (1980).
 - ²A. Miller, B. Hoogstraten, M. Staquet, and A. Winkler, "Reporting results of cancer treatment," *Cancer* **47**(1), 207–214 (1981).
 - ³P. Therasse, S. G. Arbuck, E. A. Eisenhauer, J. Wanders, R. S. Kaplan, L. Rubinstein, J. Verweij, M. Van Glabbeke, A. T. van Oosterom, and M. C. Christian, "New guidelines to evaluate the response to treatment in solid tumors," *J. Natl. Cancer Inst.* **92**(3), 205–216 (2000).
 - ⁴E. Eisenhauer, P. Therasse, J. Bogaerts, L. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, and M. Mooney, "New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)," *Eur. J. Cancer* **45**(2), 228–247 (2009).
 - ⁵A. Forner, C. Ayuso, M. Varela, J. Rimola, A. J. Hessheimer, C. R. de Lope, M. Reig, L. Bianchi, J. M. Llovet, and J. Bruix, "Evaluation of tumor response after locoregional therapies in hepatocellular carcinoma," *Cancer* **115**(3), 616–623 (2009).
 - ⁶M. E. Juweid, S. Stroobants, O. S. Hoekstra, F. M. Mottaghy, M. Dietlein, A. Guermazi, G. A. Wiseman, L. Kostakoglu, K. Scheidhauer, A. Buck, R. Naumann, K. Spaepen, R. J. Hicks, W. A. Weber, S. N. Reske, M. Schwaiger, L. H. Schwartz, J. M. Zijlstra, B. A. Siegel, and B. D. Cheson, "Use of positron emission tomography for response assessment of lymphoma: Consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma," *J. Clin. Oncol.* **25**(5), 571–578 (2007).
 - ⁷R. L. Wahl, H. Jacene, Y. Kasamon, and M. A. Lodge, "From RECIST to PERCIST: Evolving considerations for PET response criteria in solid tumors," *J. Nucl. Med.* **50**(Suppl. 1), 122S–150S (2009).
 - ⁸N. Avril, S. Bense, S. I. Ziegler, J. Dose, W. Weber, C. Laubenbacher, W. Römer, F. Jänicke, and M. Schwaiger, "Breast imaging with fluorine-18-FDG PET: Quantitative image analysis," *J. Nucl. Med.* **38**(8), 1186–1191 (1997).
 - ⁹C. K. Hoh, R. A. Hawkins, J. A. Glaspy, M. Dahlbom, Y. T. Nielson, E. J. Hoffman, C. Schiepers, Y. Choi, S. Rege, and E. Nitzsche, "Cancer detection with whole-body PET using 2-[¹⁸F] fluoro-2-deoxy-D-glucose," *J. Comput. Assisted Tomogr.* **17**(4), 582–589 (1993).
 - ¹⁰D. A. Pryma, J. A. O'Donoghue, J. L. Humm, A. A. Jungbluth, L. J. Old, S. M. Larson, and C. R. Divgi, "Correlation of *in vivo* and *in vitro* measures of carbonic anhydrase IX antigen expression in renal masses using antibody ¹²⁴I-cg250," *J. Nucl. Med.* **52**(4), 535–540 (2011).
 - ¹¹R. Boellaard, "Need for standardization of ¹⁸F-FDG PET/CT for treatment response assessments," *J. Nucl. Med.* **52**(Suppl. 2), 93S–100S (2011).
 - ¹²P. Price and T. Jones, "Can positron emission tomography (PET) be used to detect subclinical response to cancer therapy?," *Eur. J. Cancer* **31**(12), 1924–1927 (1995).
 - ¹³S.-P. Williams, J. E. Flores-Mercado, R. E. Port, and T. Bengtsson, "Quantitation of glucose uptake in tumors by dynamic FDG-PET has less glucose bias and lower variability when adjusted for partial saturation of glucose transport," *EJNMMI Res.* **2**(1), 1–13 (2012).
 - ¹⁴P. Cheebsumon, L. M. Velasquez, C. J. Hoekstra, W. Hayes, R. W. Kloet, N. J. Hoetjes, E. F. Smit, O. S. Hoekstra, A. A. Lammertsma, and R. Boellaard, "Measuring response to therapy using FDG PET: Semi-quantitative and full kinetic analysis," *Eur. J. Nucl. Med. Mol. Imaging* **38**(5), 832–842 (2011).
 - ¹⁵J. F. Eary, D. A. Mankoff, A. M. Spence, M. S. Berger, A. Olshen, J. M. Link, F. O'Sullivan, and K. A. Krohn, "[²-¹¹C] thymidine imaging of malignant brain tumors," *Cancer Res.* **59**(3), 615–621 (1999).
 - ¹⁶J. Qi, R. M. Leahy, S. R. Cherry, A. Chatzioannou, and T. H. Farquhar, "High-resolution 3D Bayesian image reconstruction using the microPET small-animal scanner," *Phys. Med. Biol.* **43**(4), 1001–1013 (1998).
 - ¹⁷B. Jakoby, Y. Bercier, M. Conti, M. Casey, B. Bendriem, and D. Townsend, "Physical and clinical performance of the mCT time-of-flight PET/CT scanner," *Phys. Med. Biol.* **56**(8), 2375–2389 (2011).
 - ¹⁸Y. E. Erdi, S. A. Nehmeh, T. Pan, A. Pevsner, K. E. Rosenzweig, G. Mageras, E. D. Yorke, H. Schoder, W. Hsiao, and O. D. Squire, "The CT motion quantitation of lung lesions and its impact on PET-measured SUVs," *J. Nucl. Med.* **45**(8), 1287–1292 (2004).
 - ¹⁹M. R. Benz, M. S. Allen-Auerbach, F. C. Eilber, H. J. Chen, S. Dry, M. E. Phelps, J. Czernin, and W. A. Weber, "Combined assessment of metabolic and volumetric changes for assessment of tumor response in patients with soft-tissue sarcomas," *J. Nucl. Med.* **49**(10), 1579–1584 (2008).
 - ²⁰A. Ahmadian, M. R. Ay, J. H. Bidgoli, S. Sarkar, and H. Zaidi, "Correction of oral contrast artifacts in CT-based attenuation correction of PET images using an automated segmentation algorithm," *Eur. J. Nucl. Med. Mol. Imaging* **35**(10), 1812–1823 (2008).
 - ²¹J. A. Boucek, R. J. Francis, C. G. Jones, N. Khan, B. A. Turlach, and A. J. Green, "Assessment of tumour response with ¹⁸F-fluorodeoxyglucose positron emission tomography using three-dimensional measures compared to SUVmax—A phantom study," *Phys. Med. Biol.* **53**(16), 4213–4230 (2008).
 - ²²S. Tong, A. Alessio, and P. Kinahan, "Noise and signal properties in PSF-based fully 3D PET image reconstruction: An experimental evaluation," *Phys. Med. Biol.* **55**(5), 1453–1473 (2010).
 - ²³J. Bading, P. Frankel, H. Bertilsson, G. Lopatin, and A. Raubitschek, "Statistical modeling of the PET maximum voxel," in *Presented at the Society of Nuclear Medicine Annual Meeting Abstracts* (2011).
 - ²⁴P. Khurd and G. Gindi, "Fast LROC analysis of Bayesian reconstructed emission tomographic images using model observers," *Phys. Med. Biol.* **50**(7), 1519–1532 (2005).
 - ²⁵H. H. Barrett, D. W. Wilson, and B. M. Tsui, "Noise properties of the EM algorithm. I. Theory," *Phys. Med. Biol.* **39**(5), 833–846 (1994).
 - ²⁶E. J. Soares, C. L. Byrne, and S. J. Glick, "Noise characterization of block-iterative reconstruction algorithms. I. Theory," *IEEE Trans. Med. Imaging* **19**(4), 261–270 (2000).
 - ²⁷J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Trans. Image Process.* **5**(3), 493–506 (1996).
 - ²⁸J. Qi and R. M. Leahy, "Resolution and noise properties of MAP reconstruction for fully 3-D PET," *IEEE Trans. Med. Imaging* **19**(5), 493–506 (2000).
 - ²⁹J. Qi, "A unified noise analysis for iterative image estimation," *Phys. Med. Biol.* **48**(21), 3505–3519 (2003).
 - ³⁰J. Dutta, S. Ahn, and Q. Z. Li, "Quantitative statistical methods for image quality assessment," *Theranostics* **3**(10), 741–756 (2013).
 - ³¹Z. Li, Q. Li, D. Pantazis, X. Yu, P. S. Conti, and R. M. Leahy, "Controlling familywise error rate for matched subspace detection in dynamic FDG PET," *IEEE Trans. Med. Imaging* **28**(10), 1623–1631 (2009).
 - ³²T. Nichols and S. Hayasaka, "Controlling the familywise error rate in functional neuroimaging: A comparative review," *Stat. Methods Med. Res.* **12**(5), 419–446 (2003).
 - ³³S. Lepage, "Stochastic finite element method for the modeling of thermoelastic damping in micro-resonators," Ph.D. thesis, Université de Liège, 2006, available at http://www.ltas-vis.ulg.ac.be/cmsms/uploads/File/Lepage_PhD.pdf.
 - ³⁴E. Vanmarcke, *Random Fields: Analysis and Synthesis* (World Scientific, Singapore, 2010).
 - ³⁵M. Phillips, E. Somer, P. Schleyer, D. Cash, J. Mackintosh, M. O'Doherty, S. Barrington, D. Hill, and P. Marsden, "Evaluation of 3 registration methods for pre- and post-therapy PET/CT studies," in *Presented at the Society of Nuclear Medicine Annual Meeting Abstracts* (2010).
 - ³⁶W. Zhu, R. M. Leahy, P. S. Conti, and Q. Li, "Longitudinal registration of liver PET scans using four phase CT," in *Presented at the IEEE Nuclear Science Symposium Conference Record (NSS/MIC)* (2010).
 - ³⁷J. Maintz and M. A. Viergever, "A survey of medical image registration," *Med. Image Anal.* **2**(1), 1–36 (1998).
 - ³⁸D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Phys. Med. Biol.* **46**(3), R1–R45 (2001).
 - ³⁹M. Holden, "A review of geometric transformations for nonrigid body registration," *IEEE Trans. Med. Imaging* **27**(1), 111–128 (2008).
 - ⁴⁰M.-J. Antoine, J.-M. Travière, and D. Bloyet, "Modeling of 2D PET noise autocovariance function applied to individual activation studies," in *Presented at the IEEE Nuclear Science Symposium and Medical Imaging Conference Record* (1994).
 - ⁴¹K. J. Worsley, S. Marrett, P. Neelin, and A. Evans, "Searching scale space for activation in PET images," *Hum. Brain Mapp.* **4**(1), 74–90 (1996).
 - ⁴²E. Wolsztynski, F. O'Sullivan, E. U. Conrad, J. O'Sullivan, and J. F. Eary, "A novel approach to the assessment of response to chemotherapy in sarcoma imaged with PET-FDG," in *Presented at the IEEE Nuclear Science Symposium Conference Record (NSS/MIC)* (2010).
 - ⁴³M. Soltani and P. Chen, "Effect of tumor shape and size on drug delivery to solid tumors," *J. Biol. Eng.* **6**, 4–18 (2012).