Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2015 April 13; 9416: 94161K-. doi:10.1117/12.2081286.

Objective evaluation of reconstruction methods for quantitative SPECT imaging in the absence of ground truth

Abhinav K. Jha¹, Na Song², Brian Caffo³, and Eric C. Frey¹

¹Division of Medical Imaging Physics, Department of Radiology and Radiological Sciences, Johns Hopkins University, Baltimore, MD, USA

²Division of Nuclear Medicine, Department of Radiology, Albert Einstein College of Medicine, Yeshiva University, Bronx, NY, USA

³Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Abstract

Quantitative single-photon emission computed tomography (SPECT) imaging is emerging as an important tool in clinical studies and biomedical research. There is thus a need for optimization and evaluation of systems and algorithms that are being developed for quantitative SPECT imaging. An appropriate objective method to evaluate these systems is by comparing their performance in the end task that is required in quantitative SPECT imaging, such as estimating the mean activity concentration in a volume of interest (VOI) in a patient image. This objective evaluation can be performed if the true value of the estimated parameter is known, i.e. we have a gold standard. However, very rarely is this gold standard known in human studies. Thus, no-gold-standard techniques to optimize and evaluate systems and algorithms in the absence of gold standard are required. In this work, we developed a no-gold-standard technique to objectively evaluate reconstruction methods used in quantitative SPECT when the parameter to be estimated is the mean activity concentration in a VOI. We studied the performance of the technique with realistic simulated image data generated from an object database consisting of five phantom anatomies with all possible combinations of five sets of organ uptakes, where each anatomy consisted of eight different organ VOIs. Results indicate that the method provided accurate ranking of the reconstruction methods. We also demonstrated the application of consistency checks to test the no-gold-standard output.

Keywords

No-gold-standard methods; Quantitative SPECT; Evaluating reconstruction methods

1. INTRODUCTION

Quantitative single-photon emission computed tomography (SPECT) is emerging as an important tool for clinical studies and biomedical research.^{1, 2} In quantitative SPECT, the end task is estimating a certain functional or anatomical parameter about the object of interest from a SPECT image, such as quantifying the total activity or the mean activity concentration within a certain volume of interest (VOI) in a patient image. Quantitative

SPECT has several important applications such as in targeted radionuclide therapy (TRT) for treatment planning,^{3–5} myocardial perfusion SPECT studies that quantify regional blood flow⁶ or perform dynamic analysis of myocardial perfusion to derive the kinetic parameters or the perfusion volume,⁷ dopamine transporter scan (DaTSCAN) quantification,^{8, 9} and renal Tc-99m dimercaptosuccinic acid (DMSA) SPECT studies.^{10, 11}

Several imaging systems and algorithms have been designed to provide accurate quantitative SPECT imaging. There is a requirement for objective methods to optimize and evaluate these systems and algorithms on the basis of their performance in the task of estimating the parameter of interest accurately and precisely. Such objective evaluation is greatly simplified when we know the true value of the parameter, but this is rarely the case in patient studies. For this reason, animal, physical phantom and simulation studies have become an important tool in this optimization and evaluation process. However, it is highly desirable to ultimately validate these results using clinical studies and human images where the true value, which serves as a gold standard, is known imperfectly or not at all. Thus, in order to objectively compare systems and algorithms for quantitative SPECT, there is a need for evaluation methods that can be implemented in the absence of a gold standard. In this paper, our broad objective is to design such a no-gold-standard method to evaluate quantitative SPECT systems on the task of estimating the mean activity concentration within a VOI.

The objective evaluation of imaging systems or algorithms in the absence of gold standard has been a challenging and important research problem. Henkelman et al.¹² derived a method to perform receiver operating characteristic (ROC) analysis without knowing the true diagnosis. A method to objectively compare the performance of imaging systems and algorithms for estimation tasks in the absence of ground truth was developed by Hoppin et al.¹³ and Kupinski et al.¹⁴ The method was used to compare cardiac-ejection-fraction estimation algorithms in the absence of ground truth.¹⁵ The method was extended to objectively evaluate segmentation algorithms in diffusion-weighted magnetic resonance imaging (DWMRI) when the ground truth segmentation is not known.^{16–18}

In quantitative SPECT imaging for activity estimation, most current methods to evaluate algorithms assume a gold standard that is obtained either by applying another algorithm on the image or from another imaging modality,¹⁹ and thus is not necessarily accurate. Further, determining the actual gold standard is very complicated in these applications. Thus, there is a great need for a no-gold-standard approach for evaluating systems and algorithms based on the task of activity estimation. The particular problem we consider in this paper is the evaluation of reconstruction methods for quantitative SPECT imaging when the end task is estimating the activity within a VOI.

2. METHODS

2.1 Theory

Consider the case where P patients are imaged using a SPECT imaging system. The acquired projection data are reconstructed using K different reconstruction methods, and the mean activity concentration in the VOI is estimated using the reconstructed images from all

the methods. Our objective is to develop a no-gold-standard technique to evaluate these K reconstruction methods based on the task of estimating the mean activity concentration. A schematic illustrating the problem statement is presented in Fig. 1.

The method that we propose in this paper builds up on the basic framework as proposed previously in Hoppin et al.,¹³ Kupinski et al.¹⁴ and Jha et al.¹⁷ We assume that there is some relationship between the true and estimated activity values, consisting of both deterministic and random components. Using realistic simulation studies described below, and with the aid of measures such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC), we found that a linear model, consisting of a slope, intercept, and zero-mean normally-distributed noise term, could be used to model the relationship between the true and estimated activity values. Thus, denoting the true activity value for the p^{th} patient by a_p , the estimated activity value for this patient using the k^{th} reconstruction method by $\hat{a}_k^p$, and the slope, intercept and standard deviation of the noise term for the k^{th} method by u_k , v_k and σ_k , respectively, we write this model as

$$\hat{a}_k^p = u_k a_p + v_k + \mathcal{N}(0, \sigma_k), \quad (1)$$

where $\mathcal{N}(0, \sigma_k)$ denotes a zero-mean normally distributed random variable with standard deviation σ_k . Let us denote the unknown linear model parameters for all the reconstruction methods, i.e. $\{u_k, v_k, \sigma_k, k = 1, 2 \dots K\}$ by Θ . We next assume that the true value a_p arises from some unknown distribution. Based on our empirical studies and in order to accommodate a wide variety of distribution shapes, we choose this distribution to be a four-parameter beta distribution, parameterized by the vector Ω .¹⁷ The beta distribution has the ability to model non-symmetric data, negatively skewed, uni-modal, strictly increasing, strictly decreasing, concave, convex and uniform distributions, as illustrated by some representative beta distributions in Fig. 2. Note that we do not know the values of Ω or Θ .

Let us denote the set of all estimated mean activity concentration values, i.e the set $\{\hat{a}_k^p, k = 1, 2 \dots K, p = 1, 2, \dots, P\}$, by $\mathcal{A}$. The no-gold-standard method uses a maximum-likelihood (ML) approach, i.e. it estimates the values of the unknown parameters that maximize the probability of all the observed data. Denoting the estimated values by $\{\hat{\Theta}, \hat{\Omega}\}$, the objective is to determine

$$\{\hat{\Theta}, \hat{\Omega}\} = \underset{\Theta, \Omega}{\operatorname{argmax}} \operatorname{pr}(\mathcal{A} | \Theta, \Omega) \quad (2)$$

We determined the expression for $\operatorname{pr}(\mathcal{A} | \Theta, \Omega)$ and subsequently computed the ML solution using a quasi-Newton optimization technique. We used these estimated parameters to compute two figures of merit. The first figure of merit represents the precision of the algorithm when the slope and intercepts are treated as calibration factors and corrected. The correction of the slope amplifies the noise variance, as illustrated by the schematic in Fig. 3. Therefore, the noise-to-slope ratio σ_k/a_k , was chosen as the first figure of merit. The second figure of merit is the scaled ensemble mean square error (EMSE), denoted by EMSE_{SC} , and is defined as

$$\text{EMSE}_{\text{sc}} = \frac{v_k^2 + \sigma_k^2}{u_k^2} \quad (3)$$

EMSE_{SC} is a modified EMSE metric assuming that the data are corrected for the slope factor but not the intercept. This metric measures both the bias and variance of the reconstruction algorithm.

2.2 Consistency checks

The no-gold-standard approach could yield incorrect ranking of the methods due to erroneous estimation of the model parameters. Thus, we developed checks that verify if the estimated no-gold-standard output is consistent with the observed data. These checks test the similarity between the histogram of estimated activity values and the theoretically estimated distribution computed using the no-gold-standard method. The first consistency check computes the correlation coefficient of the quantile-quantile (QQ) plot²⁰ between the two distributions, while the second check is a Pearson's χ^2 test²¹ with these two distributions.

2.3 Validating the no-gold-standard approach

We simulated a SPECT imaging system that images an I-131 radioisotope distribution modeling the uptake of an I-131 labeled anti-CD20 antibody used for radionuclide therapy of non-Hodgkins Lymphoma. The object database consisted of five patient anatomies with all possible combinations of five sets of organ uptakes. The phantoms were based on the non-uniform rational B-spline (NURBS)-based cardiac-torso (NCAT) phantom with organ sizes and uptakes based on patient data.²² The data were simulated using realistic and previously validated MC simulation methods. The acquired projection data were reconstructed using compensation for three different combinations of model-based compensation implemented using the ordered subsets expectation maximization (OSEM) algorithm. The compensations investigated included attenuation and scatter (AS) compensation, attenuation, scatter and detector response compensation (ADS), and ADS with the addition of explicit compensation for down-scatter from high-energy photons (ADS.DWN).²³ Scatter compensation used the effective scatter source estimation (ESSE) scatter model and detector response compensation included the geometric, septal penetration and septal scatter components.²⁴ The activity in eight different VOIs corresponding to the lungs, liver, heart, spleen, kidneys, pelvis, blood, and background were estimated assuming knowledge of the true organ VOIs. In our simulation study, we computed the noise-to-slope ratio and the EMSE_{SC} for the different reconstruction methods using the estimated no-gold-standard parameters. We used these estimated parameters to predict the rank of the reconstruction methods. These rankings were then compared to the true rankings obtained using the actual value of the EMSE_{SC} and noise-to-slope ratio.

3. RESULTS

The true values of the figures of merit and the values estimated using the no-gold-standard method for one of the noise realizations are shown in Table 1. It is observed that both the estimated and the true noise-to-slope ratios yield the same rankings for the three reconstruction methods. The same trend is observed with the rankings predicted using the

true and estimated $EMSE_{SC}$ values. Thus, the ranking predicted using the parameters estimated from the no-gold-standard technique is the same as the rankings from the true figures of merit.

The regression lines obtained using the estimated no-gold-standard parameters were also compared to the scatter plot of the true and estimated activity values for all the reconstruction methods. It is observed, as shown in Fig. 4, that for all the reconstruction methods, the scatter plot coincided with the estimated regression line, thus demonstrating the efficacy of the no-gold-standard technique. It is also observed that the regression line plots reflected the noise in the relation between the true and estimated mean activity concentration values.

Further, when the no-gold-standard method estimated the correct ranking, the correlation coefficient of the Q-Q plot and the p-value returned by the Pearson's χ^2 test were both very close to 1. Thus the consistency check correctly predicted that the no-gold-standard method yielded the correct rankings.

4. CONCLUSIONS

We have developed a no-gold-standard method for evaluating reconstruction methods for quantitative SPECT imaging where the end task is estimating the activity concentration in a VOI. The performance of the method has been studied with a realistic simulated image dataset, and results show that the method provides accurate estimate of the rankings of the considered reconstruction methods using two different figures of merit. We are currently repeating these experiments with multiple noise realizations to assess the reliability of the method. An issue with the approach is the amount of patient data required to obtain reliable performance, especially since much patient data might be unavailable. Currently we are also quantifying this data requirement and investigating methods for its reduction. The method can be extended to objectively evaluate quantitative SPECT systems and algorithms for other quantitative tasks such as VOI segmentation and histogram estimation.

Acknowledgments

This work was supported by National Institute of Biomedical Imaging and Bioengineering of National Institute of Health under numbers R01-EB016231, R01-CA109234 and U01-CA140204. The authors would like to thank Dr. Matthew Kupinski for helpful discussions.

A portion of the reconstruction code used in this work has been licensed to GE Healthcare for inclusion in a commercial product. Under separate licensing agreements between the General Electric Co. and the Johns Hopkins University and the University of North Carolina at Chapel Hill and GE Healthcare, Dr. Frey is entitled to a share of royalty received by the universities on sales of the software. The terms of this arrangement are being managed by the Johns Hopkins University in accordance with its conflict of interest policies.

References

1. Bailey DL, Willowson KP. An evidence-based review of quantitative SPECT imaging and potential clinical applications. *J Nucl Med.* Jan.2013 54:83–89. [PubMed: 23283563]
2. Bailey DL, Willowson KP. Quantitative SPECT/CT: SPECT joins PET as a quantitative imaging modality. *Eur J Nucl Med Mol Imaging.* Sep 14.2013

3. Dewaraja YK, Frey EC, Sgouros G, Brill AB, Roberson P, Zanzonico PB, Ljungberg M. MIRD pamphlet No. 23: quantitative SPECT for patient-specific 3-dimensional dosimetry in internal radionuclide therapy. *J Nucl Med.* Aug.2012 53:1310–1325. [PubMed: 22743252]
4. Dewaraja YK, Ljungberg M, Green AJ, Zanzonico PB, Frey EC, Bolch WE, Brill AB, Dunphy M, Fisher DR, Howell RW, Meredith RF, Sgouros G, Wessels BW. MIRD pamphlet No. 24: Guidelines for quantitative 131I SPECT in dosimetry applications. *J Nucl Med.* Dec.2013 54:2182–2188. [PubMed: 24130233]
5. Ljungberg M, Sjogreen K, Liu X, Frey E, Dewaraja Y, Strand SE. A 3-dimensional absorbed dose calculation method based on quantitative SPECT for radionuclide therapy: evaluation for (131)I using monte carlo simulation. *J Nucl Med.* Aug.2002 43:1101–1109. [PubMed: 12163637]
6. Silva AJD, Tang HR, Wong KH, Wu MC, Dae MW, Hasegawa BH. Absolute quantification of regional myocardial uptake of 99mTc-sestamibi with SPECT: experimental validation in a porcine model. *J Nucl Med.* May.2001 42:772–779. [PubMed: 11337575]
7. Iida H, Eberl S, Kim KM, Tamura Y, Ono Y, Nakazawa M, Sohlberg A, Zeniya T, Hayashi T, Watabe H. Absolute quantitation of myocardial blood flow with (201)Tl and dynamic SPECT in canine: optimisation and validation of kinetic modelling. *Eur J Nucl Med Mol Imaging.* May.2008 35:896–905. [PubMed: 18202845]
8. Zaidi H, Fakhri GE. Is absolute quantification of dopaminergic neurotransmission studies with 123I SPECT ready for clinical use? *Eur J Nucl Med Mol Imaging.* Jul.2008 35:1330–1333. [PubMed: 18523768]
9. Morton RJ, Guy MJ, Clauss R, Hinton PJ, Marshall CA, Clarke EA. Comparison of different methods of DatSCAN quantification. *Nucl Med Commun.* Dec.2005 26:1139–1146. [PubMed: 16264363]
10. Cao X, Zurakowski D, Diamond DA, Treves ST. Automatic measurement of renal volume in children using 99mTc dimercaptosuccinic acid SPECT: normal ranges with body weight. *Clin Nucl Med.* Apr.2012 37:356–361. [PubMed: 22391704]
11. Yen TC, Chen WP, Chang SL, Liu RS, Yeh SH, Lin CY. Technetium-99m-DMSA renal SPECT in diagnosing and monitoring pediatric acute pyelonephritis. *J Nucl Med.* Aug.1996 37:1349–1353. [PubMed: 8708770]
12. Henkelman RM, Kay I, Bronskill MJ. Receiver operator characteristic (ROC) analysis without truth. *Med Decis Making.* 1990; 10:24–9. [PubMed: 2325524]
13. Hoppin JW, Kupinski MA, Kastis GA, Clarkson E, Barrett HH. Objective comparison of quantitative imaging modalities without the use of a gold standard. *IEEE Trans Med Imag.* May. 2002 21:441–9.
14. Kupinski MA, Hoppin JW, Clarkson E, Barrett HH, Kastis GA. Estimation in medical imaging without a gold standard. *Acad Radiol.* Mar.2002 9:290–7. [PubMed: 11887945]
15. Kupinski MA, Hoppin JW, Krasnow J, Dahlberg S, Leppo JA, King MA, Clarkson E, Barrett HH. Comparing cardiac ejection fraction estimation algorithms without a gold standard. *Acad Radiol.* Mar.2006 13:329–37. [PubMed: 16488845]
16. Jha AK, Kupinski MA, Rodriguez JJ, Stephen RM, Stopeck AT. Evaluating segmentation algorithms for diffusion-weighted mr images: a task-based approach. *Proc SPIE Medical Imaging.* Feb.2010 7627:762701–8. Best student paper award.
17. Jha AK, Kupinski MA, Rodriguez JJ, Stephen RM, Stopeck AT. Task-based evaluation of segmentation algorithms for diffusion-weighted MRI without using a gold standard. *Phys Med Biol.* Jul.2012 57:4425–4446. [PubMed: 22713231]
18. Jha, AK. ADC Estimation in Diffusion-Weighted Images. 2009.
19. Willowson K, Bailey DL, Baldock C. Quantifying lung shunting during planning for radio-embolization. *Phys Med Biol.* Jul 7.2011 56:N145–52. [PubMed: 21628777]
20. Wilk MB, Gnanadesikan R. Probability plotting methods for the analysis of data. *Biometrika.* 1968; 55(1):1–17. [PubMed: 5661047]
21. Frieden, BR. Springer series in information sciences. 3. Springer-Verlag; 1991. Probability, Statistical Optics, and Data Testing: A Problem Solving Approach.
22. Segars WP, Tsui B, Lalush D, Frey E, King M, Manocha D. Development and application of the new dynamic Nurbs-based Cardiac-Torso (NCAT) phantom. *J Nucl Med.* 2001; 42(5):7P–7P.

23. Song N, Du Y, He B, Frey EC. Development and evaluation of a model-based down-scatter compensation method for quantitative I-131 SPECT. Med Phys. Jun.2011 38:3193–3204. [PubMed: 21815394]
24. Frey E, Tsui B. A new method for modeling the spatially-variant, object-dependent scatter response function in SPECT. IEEE Nuclear Science Symposium. Nov.1996 2:1082–1086.

Figure 1.

A schematic illustrating the problem statement for the no-gold-standard method

Figure 2.

Various shapes taken by the beta distribution as the distribution parameters are varied

Figure 3.

A demonstration of the effect of correcting the data for the bias and slope terms. In (a), we present an example plot of synthetically generated true and estimated mean activity concentration values. The estimated values are corrected for the slope and bias, and in (b), a plot of the true and corrected mean activity concentration values is presented. We observe that the noise is amplified due to the correction of the slope.

Figure 4.

Regression lines estimated using the no-gold-standard method for the three reconstruction methods. The solid line is generated using the estimated linear model parameters, and the dashed line denotes the estimated standard deviation.

Table 1

The true and estimated figures of merit for one noise realization. The rankings predicted from the estimated figures of merit are the same as the rankings using the true figures of merit.

Reconstruction method	True noise-to-slope ratio	Estimated noise-to-slope ratio	True EMSE_{SC}	Estimated EMSE_{SC}
AS	0.299	0.173	0.128	0.056
ADS	0.190	0.099	0.039	0.014
ADS-DWN	0.167	0.002	0.029	0.001