Published in final edited form as:

Phys Med Biol. 2015 January 21; 60(2): 671–688. doi:10.1088/0031-9155/60/2/671.

Volume estimation of low-contrast lesions with CT: a comparison of performances from a phantom study, simulations and theoretical analysis

Qin Li, Marios A Gavrielides, Rongping Zeng, Kyle J Myers, Berkman Sahiner, and Nicholas Petrick

Division of Imaging, Diagnostics, and Software Reliability, Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD 20993, USA

Abstract

Measurements of lung nodule volume with multi-detector computed tomography (MDCT) have been shown to be more accurate and precise compared to conventional lower dimensional measurements. Quantifying the size of lesions is potentially more difficult when the object-to-background contrast is low as with lesions in the liver. Physical phantom and simulation studies are often utilized to analyze the bias and variance of lesion size estimates because a ground truth or reference standard can be established. In addition, it may also be useful to derive theoretical bounds as another way of characterizing lesion sizing methods. The goal of this work was to study the performance of a MDCT system for a lesion volume estimation task with object-to-background contrast less than 50 HU, and to understand the relation among performances obtained from phantom study, simulation and theoretical analysis. We performed both phantom and simulation studies, and analyzed the bias and variance of volume measurements estimated by a matched-filter-based estimator. We further corroborated results with a theoretical analysis to estimate the achievable performance bound, which was the Cramer–Rao’s lower bound (CRLB) of minimum variance for the size estimates. Results showed that estimates of non-attached solid small lesion volumes with object-to-background contrast of 31–46 HU can be accurate and precise, with less than 10.8% in percent bias and 4.8% in standard deviation of percent error (SPE), in standard dose scans. These results are consistent with theoretical (CRLB), computational (simulation) and empirical phantom bounds. The difference between the bounds is rather small (for SPE less than 1.9%) indicating that the theoretical- and simulation-based performance bounds can be good surrogates for physical phantom studies.

Keywords

volume estimation; phantom study; Cramer–Rao’s lower bound; low-contrast lesion; computed tomography

1. Introduction

Lesion size and change in size measured from CT scans are used in clinical practice to confirm diagnosis and monitor patient response to treatment. Currently, size is primarily

measured with one-dimensional (1D) measurements, estimating the longest diameter of a lesion within a single cross-sectional slice, in accordance with the RECIST criteria (Therasse *et al* 2000, Eisenhauer *et al* 2009). The RECIST criteria replaced the WHO criteria which were based on a two-dimensional (2D) measure consisting of the product of the in-slice longest diameter and its longest perpendicular diameter (Miller *et al* 1981). However, tumor volume is an alternative size measurement that may potentially yield improved accuracy and precision over the other two sizing metrics. Volumes should more accurately represent actual tumor burden within a patient since tumors are three-dimensional (3D) and often grow asymmetrically. With modern multi-detector CT (MDCT) systems, scans can be acquired within a single breath-hold and reconstructed into a stack of thin-slice images, producing near-isotropic data and facilitating 3D tumor volume estimation. Detection of tumor size change over time is hindered by the uncertainties associated with the imaging and size measurement processes. It is therefore critical to discover and understand the sources of such uncertainties in order to facilitate the assessment, development, standardization and adoption of hardware and software for improving the accuracy and reproducibility of tumor volume estimates.

Volume estimation error depends on a number of factors including scan acquisition and reconstruction parameters, lesion characteristics, and estimation methods (Gavrielides *et al* 2009). In order to quantify the effect of these factors, a number of experimental approaches are available, each with its own benefits and limitations. Clinical studies provide realistic data but for evaluating measurement errors, they suffer from a lack of reference standard in terms of true lesion size. They also have the added risks associated with exposing patients to ionizing radiation. As such, clinical studies examining lesion sizing, when available, are typically limited to ‘coffee-break’ or zero-change experiments where patients are scanned twice in a short period of time to allow for repeatability analysis (Wormanns *et al* 2004, Reeves *et al* 2006, de Hoop *et al* 2009, Zhao *et al* 2009) but not bias analysis. Alternatively, phantom studies provide a framework where the true lesion size can be established, and they allow for studies of a wide range of imaging parameters without concerns for patient safety. A number of phantom studies have been conducted to examine the effect of different imaging factors on the error associated with volumetry measurements of lung nodules (Das *et al* 2007, Gavrielides *et al* 2009, 2010, 2012, 2013, Oda *et al* 2010, Chen *et al* 2012) and other lesions (Prionas *et al* 2010). However, a limitation of phantom studies is that they are physically limited to a finite number of lesion/background combinations, which does not match the true complexity in clinical imaging. Another alternative to clinical studies is the use of *in silico* approaches, where computer models are developed to simulate imaging systems, objects and backgrounds. This approach allows for a much wider range of object/background combinations to be evaluated within a low-cost computational setting compared with the use of physical phantoms. In previous work, we used such an approach to examine noise covariance approximations in helical CT and their effect on lung nodule estimation (Zeng *et al* 2011). A third approach that also has utility in characterizing measurement error is the derivation of theoretical bounds. Examples include the studies by Kupinski *et al* (2003), where they examined the use of the minimum variance of the ideal, unbiased estimator (calculated from the Cramer–Rao’s lower bound (CRLB)) as a figure of merit for

hardware optimization, and Li *et al* (2013b), where CRLB was derived as the limits of achievable performance.

Combining findings from clinical, phantom, and *in silico* CT studies would likely reduce the size, length, and cost of clinical trials needed to identify and thoroughly evaluate the multiple interrelated factors impacting CT-based lesion sizing measurement error. To achieve that, it would be necessary to have a better understanding of the performance bound for each approach, so that findings can be interpreted in a meaningful way. Differences among clinical, phantom, *in silico* CT studies performance and the theoretical bounds are expected because of the complexity of the imaging process. We believe it is equally crucial to discover and understand these differences in addition to trying to reduce them.

In this work, we extend our previous work (Li *et al* 2013b) on the development of a theoretical bound for a high contrast volume estimation task by further validating this bound in *in silico* simulations and phantom experiments. In particular, we investigate the performance of a matched-filter volume estimator for a low contrast (31–46 HU) volume estimation task. The volume estimator is modified from an earlier work (Gavrielides *et al* 2010) so that it is able to simultaneously estimate multiple parameters, including nuisance parameters, and by incorporating a data pre-whitening function to more accurately deal with image noise further improving the accuracy of the volume estimator. The task model incorporates synthetic lesions surrounded by a parenchyma-mimicking gelatin (73 HU) resulting in lesion/background differences in the order of 31–46 HU. When the object-to-background contrast is lower as with lesions in the liver, quantifying the size of lesions is potentially more difficult since more variation is associated with the lesion's boundary contaminated by noise. To the best of our knowledge, this is the first study attempting to compare the performance of the aforementioned three approaches for such a task.

The remaining paper is organized as follows. We first present our imaging model, volume estimator and the theoretical bound derivation for the lesion volume estimation variance, i.e. CRLB, in section 2. Then we present our implementation of the matched-filter volume estimator in section 3 and highlight its difference from the earlier version. Section 4 describes the evaluation of the CRLB, simulation and phantom performances. The physical phantom, imaging protocols and the virtual CT system are also described in this section. Finally, we make comparisons between CRLB bounds derived from the theoretical, *in silico*, and phantom performance estimates and illustrate the improvement found when using the modified matched-filter lesion volume estimator. Results are summarized in section 5, followed by a discussion of the results and the conclusion in section 6.

2. Volume estimator and theoretical performance bound

In this work, we modeled the target object as a sphere with unknown size, location and radiodensity. Let the embedded object be parameterized by vector $\mathbf{a} = (r, x, y, z, \mu)$, which corresponds to 5 parameters including radius (r), location coordinates (x, y, z) and object radiodensity (μ). We describe the imaging system by

$$\mathbf{g} = \mathcal{H} \circ (\mathbf{f}_s(\mathbf{a}) + \mathbf{f}_b) + \mathbf{n}, \quad (1)$$

where $\mathcal{H}$ represents the imaging operator mapping the objects and background from physical space to image (reconstruction) space ($\circ$ denotes a function composition), $\mathbf{f}_s(\mathbf{a})$ is the object parameterized by vector $\mathbf{a}$ superimposed on a background $\mathbf{f}_b$ (uniform in this study), $\mathbf{n}$ is the noise, and vector $\mathbf{g}$ is the observed data. In this work, $\mathcal{H}$ would be a MDCT system. Based on the work of Zeng *et al* (2011), the noise of such a system was approximated with a multivariate Gaussian with zero mean and covariance matrix $\mathbf{K}$. Then, the maximum likelihood estimator of $\mathbf{a}$ from $\mathbf{g}$, was denoted as $\hat{\mathbf{a}}(\mathbf{g})$ and defined as:

$$\hat{\mathbf{a}}(\mathbf{g}) = \underset{\mathbf{a}}{\operatorname{argmin}} (\mathcal{H} \circ (\mathbf{f}_s(\mathbf{a}) + \mathbf{f}_b) - \mathbf{g})' \mathbf{K}^{-1} (\mathcal{H} \circ (\mathbf{f}_s(\mathbf{a}) + \mathbf{f}_b) - \mathbf{g}). \quad (2)$$

From estimation theory (Van Trees 1968), $\hat{\mathbf{a}}(\mathbf{g})$ is asymptotically unbiased. Let ε_i be the i th component of the estimation error $\hat{\mathbf{a}}(\mathbf{g}) - \mathbf{a}$; then its variance satisfies

$$\sigma_{\varepsilon_i}^2 \triangleq \operatorname{Var}(\hat{a}_i(\mathbf{g}) - a_i) \geq \mathbf{F}_{ii}^{-1}(\mathbf{a}), \quad (3)$$

where the right hand side is called the Cramer–Rao’s lower bound. The matrix $\mathbf{F}$ is the Fisher information matrix, defined as: $\mathbf{F}(\mathbf{a}) = (F_{ij}(\mathbf{a}))$, where

$$F_{ij}(\mathbf{a}) = E \left\{ \frac{\partial \ln \operatorname{pr}(\mathbf{g}|\mathbf{a})}{\partial a_i} \cdot \frac{\partial \ln \operatorname{pr}(\mathbf{g}|\mathbf{a})}{\partial a_j} \right\}, \quad (4)$$

and $\mathbf{F}_{ii}^{-1}(\mathbf{a})$ is the i th diagonal element of $\mathbf{F}^{-1}$. The diagonal element of $\mathbf{F}^{-1}$ that corresponds to the radius parameter r , $\mathbf{F}_{11}^{-1}(a)$, is the bound for sizing performance. Note that we used the estimate of r for volume calculation of the spherical object. The corresponding minimum variance in volume estimation is bounded by a function of $\mathbf{F}_{11}^{-1}(\mathbf{a})$ (Van Trees 1968).

3. Implementation of the modified matched-filter estimator: effect of nuisance parameter

Instead of solving equation (2) directly, we have previously developed a matched-filter estimator (Gavrielides *et al* 2010). The algorithm minimized the cost function between the images of the target lesion and a bank of 3D templates, $\{\mathbf{K}^{-1} \mathcal{H} \circ (\mathbf{f}_s(\mathbf{a}))\}_{\mathbf{a} \in \mathcal{A}}$, where $\mathbf{K} = \mathbf{I}$ with the assumption that the noise was uncorrelated. The images of the objects $\mathcal{H} \circ (\mathbf{f}_s(\mathbf{a}))$ were generated from a simulated MDCT system with parameter vector on domain $\mathcal{A}$. In this study, we mainly modified the estimator by treating nuisance parameters and modeling the noise differently, as described below.

Since we were only interested in volume estimation, every parameter other than r is a nuisance parameter. There are several ways to deal with nuisance parameters, with each having its own advantages and disadvantages. Details are beyond the scope of this work and one can find more discussion in (Barrett and Myers 2004). In our previous study, we pre-estimated the nuisance parameters from the observer data $\mathbf{g}$ and fixed the values through the whole estimating process: the radiodensity of the object was calculated by averaging the central pixels and the lesion center was manually identified. The template bank was then generated with lesion diameter as the only variable. In this study, due to decreased object-to-background contrast and increased noise, we chose to simultaneously estimate the nuisance parameters as well as the diameter, which should improve performance by reducing the inaccuracies and variabilities associated with fixed but imperfect nuisance parameter pre-estimates. The simulated lesion template bank was again generated using a simulated MDCT system (Gavrielides *et al* 2010). For each target lesion, a bank contained templates generated with $\mathbf{a} \in \Lambda$ and Λ was defined as:

$$\Lambda = \left\{ (r=r_0 \pm k_1\delta_1, x=x_0 \pm k_2\delta_2, y=y_0 \pm k_3\delta_3, z=z_0 \pm k_4\delta_4, \mu=\mu_0 \pm k_5\delta_5) : \begin{matrix} k_i=0, 1, 2, \dots, n_i, i=1, 2, \dots, 5 \end{matrix} \right\},$$

i.e. each parameter of the object was generated in a range of $\pm n_i\delta_i$ ($i=1, \dots, 5$) around its corresponding *search center* $\cdot_0$, where n_i was manually specified. We used the target lesion size and radiodensity for the search central value r_0 and μ_0 . For the location search center (x_0, y_0, z_0) , we considered two options: use the location of the target lesion centroid (assume it is known or can be roughly estimated), or use a fixed location (for instance, the iso-center of the imaging system). The second option was more practical and saved computational cost because the set of resulting object images could be applied for the same lesion at any arbitrary location. However, the assumption behind this was that the images of the object were the same when an object shifted from (x_1, y_1, z_1) to $(x_1 + k_1I_x, y_1 + k_2I_y, z_1 + k_3I_z)$, where k_i were integers and I_x, I_y, I_z corresponded to the sizes of a voxel in x, y, z directions. In later sections, we show that with these simplifications, the estimates have additional bias. We used the CRLB from equation (3) to determine the search width (i.e. we let

$n_i\delta_i = m_i \sqrt{\mathbf{F}_{ii}^{-1}(\mathbf{a})}$, where m_i was empirically selected to be large enough so that the estimated parameters $\hat{\mathbf{a}}_i(\mathbf{g})$ would rarely be on the boundary of the search range). Once the search width was chosen, the criterion for the selection of n_i was how small a step size (δ_i) is desired in generating the template for parameter i , while keeping the computational burden reasonable. Empirically, we chose $n_i = 8$, $m_i = 6$ for $i = 1, \dots, 4$ and $n_5 = 2$, $m_5 = 2$, resulting in each bank containing $17^4 \times 5$ templates ($2n_i + 1$ for each parameter) in this study. Since it would be time consuming to scan through all possible templates, a grid-contracting strategy, was used to find the best match (Hesterman *et al* 2010).

Besides the nuisance parameters, another important component of the matched-filter estimator is the covariance matrix for pre-whitening the data, which shows up in the templates. Due to the large number of voxels in a 3D volume, it is usually difficult to achieve a stable estimate of the full covariance matrix from limited data samples. However, assumptions can be made on the noise properties to reduce the dimensionality and the

computational cost. For example, $\mathbf{K}$ can be approximated with a diagonal matrix if the noise is assumed uncorrelated; or $\mathbf{K}$ can be approximated using the noise power spectrum (NPS) if the noise is assumed stationary, i.e. uncorrelated in the Fourier domain. However, CT noise is neither uncorrelated nor globally stationary. We have previously investigated the impact of various approximations to the covariance matrix in MDCT scans on a nodule volume estimation task (Zeng *et al* 2011). In that study we found that the NPS was the most efficient approximation in terms of reducing volume estimation error relative to the amount of data required to compute the covariance, among the various approximation forms we investigated. Therefore, we chose the NPS-based pre-whitening (PW) matched-filter estimator for this study. The NPS can be estimated by averaging the spectrum of the Fourier transform of the noise (Siewerdsen *et al* 2002). More details will be described below.

4. Evaluating CRLB and performance bounds of phantom/simulation study

4.1. Calculating CRLB

In this study, we determined the CRLB for the virtual CT system we had previously developed, which approximates a real MDCT system available in our laboratory. To estimate this bound, we used the following equation for calculating Fisher information (Li *et al* 2013b),

$$\mathbf{F}(\mathbf{a}) = \left(\frac{\partial \mathbf{s}(\mathbf{a})}{\partial \mathbf{a}} \right)^T \mathbf{K}^{-1} \left(\frac{\partial \mathbf{s}(\mathbf{a})}{\partial \mathbf{a}} \right), \quad \text{where } \mathbf{s}(\mathbf{a}) = \mathcal{H} \circ \mathbf{f}_s(\mathbf{a}). \quad (5)$$

The partial derivative can be numerically approximated using a finite difference technique with the object defined as target lesions in the estimation task. To obtain the approximation of $\mathbf{K}$ by NPS method, we empirically estimated the NPS by averaging the spectrum of the Fourier transforms of many realizations or samples of noise (in this study, 100 realizations). We used the following estimation process. First, we generated 100 noisy scans of a numerical phantom using the virtual CT system. To perform comparisons among theoretical bounds, simulated results and phantom results, the objects/background of the numerical phantom were modeled with the properties of the physical phantom described in the next section. Second, for each target lesion, we calculated the difference between each of its simulated scan and the mean of all scans within a region of interest (ROI) around the lesion's center. After the mean was removed, we obtained one realization of noise. We performed this calculation in a local region because we assumed the noise was locally stationary and object dependent. The size of the ROI (specified in section 4.2.3) was selected empirically for each different imaging protocol such that the length of each dimension was about twice that of the lesion diameter. Third, we calculated the 3D discrete Fourier transform (DFT) for each of the 100 noise realizations and averaged the square of the magnitude of the 100 DFT results to obtain the NPS. Finally, under the assumption of stationarity, $\mathbf{K}$ was thus obtained implicitly through its relationship to NPS (Zeng *et al* 2011): $\mathbf{K} = \mathbf{F}^* \mathbf{W} \mathbf{F}$, where $\mathbf{F}$ is the DFT operator matrix and $\mathbf{W}$ is the NPS. Also note that to obtain the CRLB, we used 100 realizations for the NPS evaluation. As explained in later sections, the numbers of samples for NPS evaluation were 20 and 100 for phantom and

simulation studies, respectively. Discussions on how we made decisions on the number of samples and whether it was sufficient will be given in section 6.

In addition, we would like to mention that we also selected the imaging parameters of the virtual CT system to match those used in the actual CT imaging process from the phantom study. In particular, we adjusted two factors that contributed most to the noise, the reconstruction kernel (filter type) and radiation exposure (air photon flux I_0) such that the simulated noisy images had similar noise properties to that of the real scans in terms of NPS magnitude and shape. Details about the parameter selection process are given in appendix A.

4.2. Phantom study

4.2.1. Synthetic lesions and gelatin-background phantom—Three spherical synthetic lesions (CIRS Inc., Norfolk, VA, USA) were used in this study and by design, their radiodensities/diameters were: +100 HU/8 mm, +100 HU/9 mm and -10 HU/9 mm. Previously, we have obtained reference standard volumes for each lesion using high resolution micro-CT (Li *et al*/2013a). The reference standard for lesion radiodensity was the average voxel value around a $3 \times 3 \times 3$ neighborhood through the center of each lesion, extracted from 10 repeat CT scans acquired with high dose (200 mAs). The reference standard (radiodensity and size) are given in table.1. To create the low contrast, expected in the liver for example, we embedded the lesions within a gelatin-background. We used 250 g Jello-O powder (Kraft Foods Group Inc., Northfield, IL) mixed with 1500 ml water in a plastic cylindrical container (diameter 150 mm, height 300 mm). After the lesions were placed in the mixed liquid, as shown in figure 1, the mixture was refrigerated for four hours (a fourth lesion with a complex shape was also included and scanned and will be analyzed in future work due to the focus of spherical lesions for this study). Efforts were taken to keep the distances between the lesions and the phantom center constant. In cases where the lesion densities were lower than the liquid background material, a weight was attached to the lesion with a thin thread. The measured radiodensity of the resulting gelatin background was 73 HU, which made the contrast between 31–46 HU, where contrast is defined as the absolute difference between the background and the lesion.

4.2.2. Imaging protocols—The phantom was imaged with a 16-slice multi-detector helical CT scanner (IDT M $\times$ 8000, Philips Healthcare, Andover, MA). The phantom was placed such that the axis of the cylindrical container was located at the iso-center of the CT scanner. Images were acquired with a pitch value 1.2, a slice collimation of $16 \times 0.75 \text{ mm}^2$, and exposures of 20 and 100 mAs. While 100 mAs represented a standard dose in a typical clinical practice, 20 mAs represented the low dose setting. Scans were reconstructed to 0.8 and 1.5 mm thick transverse sections with 0.4 and 0.7 mm intervals, using a detail reconstruction kernel. In later sections, we use thin protocol and thick protocol to refer to these two options of reconstruction slice thickness, respectively. A field of view of 250 mm and an in-plane matrix size of 512×512 were used, resulting in a reconstructed image pixel size of $0.4883 \times 0.4883 \text{ mm}^2$. The center of the field of view was the iso-center. Acquisitions were repeated 20 times for each imaging protocol, resulting in 80 (2 slice thickness $\times$ 2 dose $\times$ 20 repeats) reconstructed scans.

4.2.3. Volume estimation and performance assessment—Volumes were estimated by the PW matched-filter estimator described in section 3. To form template banks for lesion 1–3, we obtained $\mathbf{K}$ or the NPS for each location. The NPS was estimated from ROIs centered at each spherical lesion from the 20 repeated scans. For thin and thick protocols, the size of ROI was $41 \times 41 \times 29$ and $41 \times 41 \times 21$ respectively. Let the center of the field of view be the origin (0, 0, 0). The bank of object images $\{\mathcal{H} \circ (\mathbf{f}_s(\mathbf{a}))\}_{\mathbf{a} \in A}$ were generated by the virtual CT system with objects centered at the origin (i.e. $(x_0, y_0, z_0) = (0, 0, 0)$), for the reasons stated in section 3. The template banks were then formed as the product of the corresponding $\mathbf{K}$ and the set of object images.

The performance of the volume estimates was assessed in terms of percent normalized bias (PB) and standard deviation of percent error (SPE), where PB is the expectation of percent error (PE):

$$PE = \frac{V_{\text{est}} - V_{\text{true}}}{V_{\text{true}}} \times 100\%, \text{ PB} = PE \text{ and } SPE = \text{stdev}(PE) \quad (6)$$

where V_{est} is the volume estimate and V_{true} is the reference standard.

4.3. Simulation study

4.3.1. Generation of noisy data—Noisy data were generated using the virtual CT system with the same imaging parameters as those in the actual imaging process of the phantom study, including CT geometry, pitch, slice collimation, pixel size, slice thickness (both thin and thick) and with equivalent 100 mAs exposures. Photon flux and filter type were determined as described in appendix A and were used for all simulations to appropriately match the imaging conditions. Six simulated phantoms were generated: +104 HU/8 mm sphere at the origin, +118 HU/9 mm at the origin, +27 HU/9 mm sphere at the origin, +104 HU/8 mm sphere at location of lesion 1, +118 HU/9 mm sphere at location of lesion 2, +27HU/9 mm sphere at location of lesion 3 with a simulated cylinder background (same property of physical background of the phantom). We referred to the first three lesions as simulated centered lesions. For each phantom, we generated 100 realizations of noisy images.

4.3.2. Volume estimation and performance assessment—Volumes were estimated by the PW matched-filter estimator described in section 3. To obtain $\mathbf{K}$, the NPS was estimated from ROI centered at each spherical lesion from 100 noisy realizations. The same size ROI and same object images in template banks as that of phantom study were used. Likewise, the same metrics (PB, SPE) as that applied in the phantom study were used for the performance assessment.

5. Results

5.1. Comparison among phantom, in silico CT studies performance, and theoretical bounds

The theoretical bounds on the SPE for lesions 1 to 3, evaluated as described in section 4.1, are reported in table 3 (note that the theoretical bound for SPE is simply the square root of the CRLB). For the 8 mm lesion, the lower bounds for the thin and thick protocols were 1.9% and 1.8% respectively. For the 9 mm lesion, the lower bounds for the thin and thick protocols were 1.4% and 1.3%. From these theoretical bounds, both thin and thick protocols yielded similar performance. The differences between the bounds of the 8 mm and 9 mm lesions mainly came from the differences in their contrasts. The evaluated theoretical bound on the SPE for lesion 3 is the same as that of lesion 2, which is not surprising since they have very similar contrast and the same size.

For the phantom study, PB and SPE are tabulated in table 2. The PBs were between -0.5% and 11.9% across all lesions and imaging protocols, and the maximum SPE was 5.7% . Different mAs did not appear to have much impact on PB. However, the SPE increased as mAs reduced, as expected. The reduction factor (ratio of low dose to high dose variance) ranged from 1.9 to 3.0. In theory, the reduction factor was expected to be the square root of the ratio of two exposures ($=\sqrt{5}$) (Li *et al* 2013b), which fell within the empirical range reported above. In terms of the effect of slice thickness, volume estimates with the thick protocol overestimated at a larger scale than with the thin protocol (PB range of 1.9% to 14.3% for the thick protocol compared to a range of 1.6% to 5% for the thin protocol). We investigated the source of bias and found that it came from the mismatch between the location of the objects in the template bank and that of the target lesion (appendix B). SPE seemed to be similar between thin and thick reconstructions for the same exposure. This can be due to conflicting effects: on one hand, the amount of noise increases as the slice thickness decreases; on the other hand, the number of informational slices (slices that contain the object) increases as thickness decreases. In our study, these two effects seemed to counter-balance each other. Comparing the 8 mm and 9 mm lesions volume estimates, we saw similar decrease in SPE as that of CRLB due to contrast difference. When comparing lesion 2 and 3, there was good agreement among the SPEs for each protocol as expected. However, differences in PB values were observed, possibly due to a visible air pocket found in lesion 3 (figure 2).

For simulated noisy data, PBs and SPEs are given in table 3. The template bank had no location mismatch for the simulated centered objects and the estimates were found to be almost unbiased (PB -0.3% – 0.1%). The maximum SPE was only 3.3% . For the other two non-centered objects, the volumes were all overestimated (PB 3.8% – 6.2%). The maximum SPE was 4.8% . Again, the slice thickness did not have a strong impact on the estimation and we saw SPE decrease for the 9 mm lesion compared to the 8 mm lesion due to the increase in contrast. The SPE of dataset with non-centered objects (template bank had location mismatch for the simulated objects) was overall higher than the corresponding dataset with centered objects.

We summarize the theoretical bounds, phantom and simulated performances in figure 3 using bar plots. The plots omit results for lesion 3 due to their similarity with those of lesion 2. Figure 3(a) shows the percent biases with error bar (95% of confidence interval (CI)) as given in table 3. The t-test was used to perform statistical significance testing among the simulated and phantom results for each sub-group. Figure 3(b) shows the SPE with 95% CI estimated using bootstrap resampling of the data (2000 bootstrap samples) for the simulated and phantom results. Statistical significance was evaluated using t-test as well. We did not evaluate the uncertainties on the CRLBs, but CRLBs were found to be lower than the lower limits of the 95% CI for the other three different types of results reported in figure 3. The differences, however, are small (for example, 0.8%, 0.1%, 0.6% and 0.1% respectively when compare with the lower values of the 95% CI of the centered lesions simulation results).

5.2. Improvement of modified estimator

Recall that our modified estimator estimated the nuisance parameter (location and radiodensity) simultaneously with the lesion radius and included a pre-whitening function, whereas in our old matched-filter estimator, we pre-estimated and fixed the nuisance parameters and treated the noise as uncorrelated. To illustrate the benefits of this modification, we did volume estimation on phantom data with the old estimator. We manually selected the center of each lesion and obtained the radiodensity by averaging the $3 \times 3 \times 3$ pixels around the center. We then only searched for the best match by varying the template lesion diameters. Figure 4 summarizes the comparison between our current and previous estimator designs (PBs, SPEs given in tables 2 and 4). The modified estimator generally yielded significantly smaller biases and a decreased SPE compared to those obtained by our previous estimator. For this type of lower-contrast lesion sizing task, it was found to be quite difficult to accurately locate the lesion centroid and estimate its density based on only the central pixels. Hence, errors associated with the fixed nuisance parameters for the old estimator propagated to the volume estimates.

6. Discussion/Conclusion

The empirical, simulated and theoretical results show that our matched-filter based approach for estimating the volume of solid non-attached spherical lesions with low contrast (object-to-background contrast in the order of 31–46 HU) yield both low PE and SPE. All three approaches yield estimates of PB below 12% and standard deviation of PE below 6% across all lesions and imaging protocols. Within our choices of protocols, dose was found to have a substantial impact only on the performance of the estimation in terms of standard deviation. When dose (mAs) decreased by a factor of around 5, the theoretical bounds and estimation standard deviation from both phantom and simulation studies increased by factors ranging from 1.9–3.0.

Good agreement was achieved among the volume estimation performances of the physical phantom, simulations, and theoretical bound studies. The SPE for the phantom and the theoretical bounds had differences ranging from 0.2–1.1% with two reconstruction slice thicknesses and an 100mAs exposure. The differences in the SPE measured in the phantom

and simulation studies ranged from -1.9% to 0.4% . These values are very comparable to each other.

We have carefully developed and validated the modified PW matched-filter estimator by estimating nuisance parameters additional to the size parameter. Estimates have been shown to have improved performance in terms of both accuracy and precision for this task compared to the original match-filter estimator reported in (Das *et al* 2007). One major factor was that the templates were carefully generated such that the reconstruction filter had similar frequency cutoff as that used of the real imaging system. This helped minimize the differences between the empirical results, simulation results, and theoretical bounds. However, there were still factors that needed to be accounted for, such as poly-energetic spectrum and scatter, which we have not included in our virtual CT system yet. These idealizations of the imaging system contribute to the differences among bounds.

An additional factor we investigated as a source of differences in the bounds is the uncertainty in the covariance matrix as a result of the limited number of samples available for its estimation. In the case of the phantom data, only 20 repeats scans were available to estimate the NPS. Even with this apparently small number of scans, we observed improved performance compared to the non-prewhitening (NPW) approach (appendix C) for the phantom data. For the case of the simulated data, we had more flexibility in terms of being able to test the impact of the number of samples used for the NPS estimate on the volume estimator's performance. We evaluated the PW approach with the NPS estimated from 50, 100, 200 and 500 noisy samples. For all sample sizes used to estimate the NPS, we found improved performance over the NPW approach; in addition, the results were quite similar for 100 samples or greater. We therefore chose to use 100 samples for our simulation study. We note that this evaluation is similar to the one performed in (Prionas *et al* 2010), where the authors found improvement in volume estimation through utilizing an NPS estimated from as few as 10 samples, with performance reaching a plateau after 30 samples, where the ROIs were smaller than those in this study. Finally, we note that a strategy sometimes adopted to obtain sufficient samples in the literature is to utilize ROIs extracted from multiple locations within the same scan. However, since the NPS in CT is location dependent, it can be inaccurate to do so and thus we did not use that approach here. Methods for the estimation of an accurate local NPS from only a few scans need to be investigated in future work.

One important question we wanted to address in this research besides validating the agreement among the various approaches was the role theoretical bounds might play in quantitative measurement evaluation. From this project we now anticipate that the theoretical analysis may be able to serve as a simplified way of ranking imaging protocols, imaging systems, and in determining the influence of various lesion characteristics because the bounds were found to have the same ordering of effects as the empirical results in our studies. For example, the bounds on the variability in sizing the 8 mm lesion are larger than those for the 9 mm lesion for both CRLB and the phantom results (table 3) indicating that estimation SPE decreases when size increases based on both metrics. If this can be shown to hold across a wider range of parameter and lesion characteristics, the use of the CRLB may provide an effective and efficient new tool for optimizing imaging protocols. A similar idea has also been suggested in (Kupinski *et al* 2003).

Under certain circumstances, simulation studies may be preferred to phantom studies because of logistical reasons. For instance, when the goal is to determine the minimum detectable lesion size/radiodensity change (Gavrielides *et al* 2013) synthetic lesions with very small size increments are required. In practice, it might be difficult and expensive to manufacture such a lesion set for use in a phantom study. Likewise, it may be impractical to actually image, archive and analyze all of the subtly different lesions needed for this study. From the results of our study, it appears that we can achieve reasonable agreement between the simulation and phantom experiments. While future validation is still necessary, our results provide strong initial support that the simulation approach can be used as surrogates for phantom experiments. This substantially broadens the types of questions that can be addressed in terms of optimizing and validating quantitative imaging biomarkers.

There are several limitations in our study. First, we made a strong assumption that we had full knowledge of the object shape. One way to make the estimator more general could be to include additional parameters that characterize the object shape. A second limitation was the use of non-attached solid lesions in a uniform background. We are planning to model and design phantoms representing more complicated and realistic lesions and of clinically relevant backgrounds to address this issue. We found from our experience in this study that the gelatin-based background provides an inexpensive approach for creating a semi-solid background but with the flexibility to produce a range of backgrounds with radiodensities close to blood, liver, etc. For instance, a similar approach of fabricating a gelatin-based phantom was found in (Prionas *et al* 2010), where the background they made had a radiodensity of +37.6 HU. We plan to use this basic approach but add materials such as synthetic vessels, fat-like materials, and wax powder to create more complex background to mimic the complexity and local environment found in clinical images. Lastly, the results that we presented here are estimator specific. They may not hold for other estimation approaches, e.g. segmentation-based algorithms.

In conclusion, we have derived a theoretical performance bound for a low-contrast lesion size estimation task, and compared and validated it with simulation and phantom experiments. Good agreement was found between the theoretical bounds and the volume estimation performance in the simulation and phantom experiments in terms of percent bias and standard deviation of percent error. The results of this study can be valuable toward developing and implementing methodologies for combining findings from clinical, phantom, and *in silico* CT studies regarding the sources of uncertainty for lesion volume estimation as well as other estimation tasks. Such methodologies would likely reduce the size, length, and cost of clinical trials needed to identify and thoroughly evaluate the multiple interrelated factors impacting image-based estimation tasks. Furthermore they could contribute toward the ultimate goal of accounting for sources of uncertainty, leading to more informative and meaningful quantitative biomarkers.

Appendix A

Determining photon flux and filter type

The photon flux I_0 equivalent to a 100mAs exposure was estimated in the following procedure: first, I_0 was initialized as 10^5 and the reconstruction filter was set as a Hanning filter with a variable cutoff value (no cutoff (Hann100), cutoff at 75% (Hann75) and 55% (Hann55) of the Nyquist frequency). Noisy images were then simulated using a modeled cylindrical background (same properties as the physical phantom described in section 4.2) and a spherical lesion of +104 HU, located at the same location as that of lesion 1 in the phantom study. A volume of interest (VOI) centered at the spherical lesion was then cropped from each noisy realizations. For thin and thick protocols (protocols corresponding to 0.8 mm and 1.5 mm reconstruction slice thickness), the size of the ROI was $41 \times 41 \times 29$ and $41 \times 41 \times 21$ respectively and the total number of noisy realizations was 100. The individual voxel means across the 100 realizations were then subtracted from the each VOI voxel to obtain 100 sets of noise images. The mean pixel variance and NPS at that location were then evaluated. The parameter I_0 was scaled by the ratio of the mean variances between the phantom and the simulated scans to minimize the noise level difference respectively for each of the three Hanning filters, Hann100, Hann75 and Hann55.

Figures 5 and 6 illustrate the noise properties of the simulated noisy data using Hann100, Hann75 and Hann55, compared to the phantom scans. Differences in the shape of the NPS and the appearance of the noise texture in the images can be observed for each of the Hann filter widths. Hann55 generated the best match in noise appearance compared with the other widths evaluated, and the I_0 equivalent to 100 mAs exposure was found as $I_0 = 1.3 \times 10^5$ for thin protocol and 2.0×10^5 for thick protocol. These values and Hann55 were used in the virtual CT system.

Appendix B

Influence of location and filter-type mismatch on matched-filter estimator

We examined the impact that location and filter-type mismatch between the target object and templates had on volume estimation, in terms of bias and standard deviation. For each reconstruction slice thickness (0.8 and 1.5 mm), we generated six sets of noisy scans for an 8 mm, +104 HU spherical object. Let the center of the CT volume be the origin (0, 0, 0). For the first three sets (sets 1–3), the object was placed at the origin and the CT images were reconstructed using three different filter types (Hann100, Hann75, Hann55). For the other three sets (sets 4–6), the object location was (0, 36.6, 0) mm (same as the location of lesion 1 in phantom scans) with the other conditions remaining constant. The volume of the object in each set was estimated using our template matching procedure with three template banks, where the images of objects $\mathcal{H} \circ (\mathbf{f}_s(\mathbf{a}))$ in the template banks were generated by the virtual CT system with the object centered at the origin (i.e. $(x_0, y_0, z_0) = (0, 0, 0)$) but differed in filter type (involved in $\mathcal{H}$). In short, there was no location mismatch between the template bank and the target object for sets 1–3 but there was mismatch for sets 4–6.

Table 5

Matched-filtered estimates $PB \pm SPE$ in percent. (a) No location mismatch; (b) with location mismatch. Tpl: template.

(a) Tpl Data	Thin protocol			Thick protocol		
	Hann55	Hann75	Hann100	Hann55	Hann75	Hann100
Hann55	$-0.3 \pm 3.3\%$	$-0.6 \pm 4.3\%$	$0.6 \pm 5.8\%$	$0.0 \pm 2.7\%$	$0.7 \pm 3.0\%$	$1.6 \pm 6.2\%$
Hann75	$0.4 \pm 2.0\%$	$0.3 \pm 2.0\%$	$0.1 \pm 2.2\%$	$0.0 \pm 2.0\%$	$0.0 \pm 1.9\%$	$0.0 \pm 1.5\%$
Hann100	$0.5 \pm 2.1\%$	$0.2 \pm 2.0\%$	$-0.0 \pm 1.9\%$	$0.0 \pm 1.8\%$	$0.3 \pm 1.5\%$	$0.7 \pm 1.2\%$

(b) Tpl Data	Thin protocol			Thick protocol		
	Hann55	Hann75	Hann100	Hann55	Hann75	Hann100
Hann55	$6.2 \pm 4.3\%$	$5.8 \pm 3.7\%$	$4.0 \pm 7.4\%$	$6.1 \pm 4.8\%$	$6.2 \pm 3.6\%$	$2.3 \pm 5.1\%$
Hann75	$7.5 \pm 3.1\%$	$7.1 \pm 2.7\%$	$6.8 \pm 2.7\%$	$7.1 \pm 3.3\%$	$6.9 \pm 2.5\%$	$6.0 \pm 2.6\%$
Hann100	$7.9 \pm 2.9\%$	$7.8 \pm 2.4\%$	$7.5 \pm 2.3\%$	$8.5 \pm 3.4\%$	$8.1 \pm 2.2\%$	$6.9 \pm 2.5\%$
Phantom	$4.3 \pm 2.4\%$	$3.2 \pm 2.9\%$	$2.9 \pm 4.6\%$	$10.8 \pm 2.7\%$	$9.4 \pm 3.6\%$	$9.7 \pm 5.7\%$

Volume estimation results for sets 1–3 show that the estimates are almost unbiased, ranging from -0.6 to 1.6% (table 5(A)). For sets 4–6, there is bias ranging from 4 – 7.9% for the thin slice thickness protocol and 2.3 – 8.5% for the thick slice thickness protocol (table 5(b)). The mismatch between the template bank and the images of the object due to the location difference contributes most to this error. The filter type and the reconstruction slice thickness do not appear to have a large impact on bias.

In terms of SPE, the results show that the filter type mismatch has an impact. For sets 1–3, it was observed (for all but one case) that the lowest SPEs are achieved when the filter type that produces the noisy simulated data matches the filter types used to generate the template. However, this is not as evident when location mismatch also exists. We do see increased SPE with the Hann55 filter noise data with the Hann100 generated template bank in both table 5(a) and (b). The SPE also increases when the reverse is true as well and this increase is significant ($p < 0.01$) for 6 out of 8 cases.

From these simulation results in both appendices A and B, we concluded that the Hann55 filter should be used to generate the template bank for two reasons: (1) this filter matches the real system reasonably well because the simulated noisy samples has the best match with the noise property of the real scans (figures 5–6); (2) it makes a difference whether a proper templates are used when the real scans have noise property similar to the Hann55 simulated data. To validate, for scans from the phantom study, we used three banks for the volume estimation of lesion 1 and found that SPE increased from using Hann55, Hann75 to Hann100 (table 5(b), last row). As far as location is concerned, we generated the templates centered at the origin since it is impractical to generate them for every location but also kept in mind that this could lead to increased bias.

Table 6

PW and NPW matched-filtered estimates $PB \pm SPE$ in percent for phantom study.

$PB \pm SPE$ (%) (nuisance fixed)		Thin protocol		Thick protocol	
		100 mAs	20 mAs	100 mAs	20 mAs
Lesion 1	PW	$4.3 \pm 2.4\%$	$3.6 \pm 5.6\%$	$10.8 \pm 2.7\%$	$11.9 \pm 5.6\%$
	NPW	$4.3 \pm 3.2\%$	$1.4 \pm 7.2\%$	$10.8 \pm 2.8\%$	$13.9 \pm 8.8\%$
Lesion 2	PW	$0.3 \pm 2.3\%$	$-0.5 \pm 4.4\%$	$6.8 \pm 1.5\%$	$11.1 \pm 4.8\%$
	NPW	$-1.4 \pm 2.4\%$	$-1.6 \pm 5.7\%$	$5.7 \pm 3.0\%$	$10.0 \pm 7.7\%$
Lesion 3	PW	$5.0 \pm 2.3\%$	$2.1 \pm 4.8\%$	$1.9 \pm 1.7\%$	$4.6 \pm 5.7\%$
	NPW	$5.2 \pm 2.3\%$	$1.1 \pm 5.8\%$	$4.0 \pm 2.4\%$	$5.2 \pm 7.3\%$

Appendix C

Non-prewhitening versus pre-whitening

Table 6 gives the results from both PW and NPW for the phantom study. Both approaches yield similar PB. PW reduces the SPE especially when the dose is low.

References

- Barrett, HH., Myers, KJ. Foundations of Image Science. New York: Wiley; 2004.
- Chen B, Barnhart H, Richard S, Colsher J, Amurao M, Samei E. Quantitative CT: technique dependence of volume estimation on pulmonary nodules. *Phys. Med Biol.* 2012; 57:1335. [PubMed: 22349265]
- Das M, Ley-Zaporozhan J, Gietema HA, Czech A, Mühlenbruch G, Mahnken AH, Katoh M, Bakai A, Salganicoff M, Diederich S. Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners. *Eur. Radiol.* 2007; 17:1979–84. [PubMed: 17206420]
- de Hoop B, Gietema H, van Ginneken B, Zanen P, Groenewegen G, Prokop M. A comparison of six software packages for evaluation of solid lung nodules using semi-automated volumetry: what is the minimum increase in size to detect growth in repeated CT examinations. *Eur. Radiol.* 2009; 19:800–8. [PubMed: 19018537]
- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer.* 2009; 45:228–47. [PubMed: 19097774]
- Gavrielides MA, Kinnard LM, Myers KJ, Petrick N. Noncalcified lung nodules: volumetric assessment with thoracic CT1. *Radiology.* 2009; 251:26–37. [PubMed: 19332844]
- Gavrielides MA, Li Q, Zeng R, Myers KJ, Sahiner B, Petrick N. Minimum detectable change in lung nodule volume in a phantom CT study. *Acad. Radiol.* 2013; 20:1364–70. [PubMed: 24119348]
- Gavrielides MA, Zeng R, Kinnard LM, Myers KJ, Petrick N. Information-theoretic approach for analyzing bias and variance in lung nodule size estimation with CT: a phantom study. *IEEE Trans. Med. Imag.* 2010; 29:1795–807.
- Gavrielides MA, Zeng R, Myers KJ, Sahiner B, Petrick N. Benefit of overlapping reconstruction for improving the quantitative assessment of CT lung nodule volume. *Acad. Radiol.* 2012; 20:173–80. [PubMed: 23085408]
- Hesterman JY, Caucci L, Kupinski MA, Barrett HH, Furenlid LR. Maximum-likelihood estimation with a contracting-grid search algorithm. *IEEE Trans. Nucl. Sci.* 2010; 57:1077–84. [PubMed: 20824155]

- Kupinski MA, Clarkson E, Gross K, Hoppin JW. Optimizing imaging hardware for estimation tasks. *Proc. SPIE*. 2003; 5034:309.
- Li Q, Gavrielides MA, Nagaraja S, Hagen MJ, Zeng R, Myers KJ, Sahiner B, Petrick N. A micro CT based tumor volume reference standard for phantom experiments. *Imaging Appl. Opt.* 2013a QW1G-4.
- Li Q, Zeng R, Myers KJ, Sahiner B, Gavrilides MA, Petrick N. Relating fisher information to detectability of changes in nodule characteristics with CT. *Inf. Process Med. Imaging*. 2013b; 23:584–93. [PubMed: 24684001]
- Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer*. 1981; 47:207–14. [PubMed: 7459811]
- Oda S, Awai K, Murao K, Ozawa A, Yanaga Y, Kawanaka K, Yamashita Y. Computer-aided volumetry of pulmonary nodules exhibiting ground-glass opacity at MDCT. *Am. J. Roentgenol.* 2010; 194:398–406. [PubMed: 20093602]
- Prionas ND, Ray S, Boone JM. Volume assessment accuracy in computed tomography: a phantom study. *J. Appl. Clin. Med. Phys./Am. College Med. Phys.* 2010; 11:3037.
- Reeves AP, Chan AB, Yankelevitz DF, Henschke CI, Kressler B, Kostis WJ. On measuring the change in size of pulmonary nodules. *IEEE Trans. Med. Imag.* 2006; 25:435–50.
- Siewerdsen JH, Cunningham IA, Jaffray DA. A framework for noise-power spectrum analysis of multidimensional images. *Med. Phys.* 2002; 29:2655. [PubMed: 12462733]
- Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, van Oosterom AT, Christian MC. New guidelines to evaluate the response to treatment in solid tumors. *J. Natl Cancer Inst.* 2000; 92:205–16.
- Van Trees, HL. *Detection, Estimation, and Modulation Theory, Part I*. New York: Wiley; 1968.
- Wormanns D, Kohl G, Klotz E, Marheine A, Beyer F, Heindel W, Diederich S. Volumetric measurements of pulmonary nodules at multi-row detector CT: *in vivo* reproducibility. *Eur. Radiol.* 2004; 14:86–92. [PubMed: 14615902]
- Zeng R, Petrick N, Gavrielides MA, Myers KJ. Approximations of noise covariance in multi-slice helical CT scans: impact on lung nodule size estimation. *Phys. Med. Biol.* 2011; 56:6223. [PubMed: 21896963]
- Zhao B, James LP, Moskowitz CS, Guo P, Ginsberg MS, Lefkowitz RA, Qin Y, Riely GJ, Kris MG, Schwartz LH. Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. *Radiology*. 2009; 252:263–72. [PubMed: 19561260]

Figure 1.
Left: gelatin phantom before it was refrigerated. Middle: gelatin phantom after it was refrigerated. Dense foam was squeezed on top to prevent the phantom from sliding out. Right: gelatin phantom positioned for scanning. Foam was used to make the axis of the cylinder close to the iso-center of the scanner.

Figure 2.
Central slice of lesion 1–3 from left to right. A visible air pocket was found for lesion 3.

(a)

(b)

Figure 3.

Bar plots of results from table 3. (a) Bar plot of PB with error bar (95% CI). For case 1–4, four bars correspond to CRLB, simu_c , $\text{simu}^\#$, $\text{phantom}^\#$ from left to right. (b) Bar plot of SPE with error bar estimated by using bootstrap resampling method for the simulated results and phantom results. Statistical significance among the simulated results and phantom results in each sub-group are indicated (** $p < 0.01$, * $p < 0.05$).

Figure 4.

Bar plots of PB with error bar (0.95 CI) in tables 2 and 4. Dark blue and yellow bars are when the nuisance parameters are estimated. Light blue and red bars are when the nuisance parameters are fixed. Statistical significance between the results from the two estimators for each lesion-protocol combination are indicated (** $p < 0.01$).

Figure 5. Noisy scans (central slice of object) from a phantom scan (100 mAs, 0.8×0.4 slice thickness) and simulated noisy scans with Hann100, Hann75, Hann55 reconstruction filters.

Figure 6.

Effect of reconstruction filter type on NPS shape. Central slice of 3D NPS from noisy images obtained from: (a) phantom scans (ROI of lesion 1, thin slice thickness, 20mAs); (b) simulated noisy samples with Hann100 reconstruction filter; (c) simulated noisy samples with Hann75 reconstruction filter; (d) simulated noisy samples with Hann55 reconstruction filter. The value of I_0 was tuned so that the noise levels were the same in the sense that the mean pixel variances (or the integration of the NPS) were equal.

Table 1

Summary of lesion characteristics (ref std stands for reference standard).

	Diameter (design) (mm)	Diameter (ref std) (mm)	Shape	Density (design)	Density (ref std)	Contrast (ref std)
Lesion 1	8	8.10	Spherical	100 HU	104 HU	31 HU
Lesion 2	9	8.97	Spherical	100 HU	118 HU	45 HU
Lesion 3	9	8.93	Spherical	-10 HU	27 HU	46 HU

Table 2

PBs and SPEs of PW matched-filter estimates for phantom studies. Results are reported for thin/thick, high/low mAs protocols for the three spherical lesions.

PB $\pm$ SPE (%)	Thin protocol		Thick protocol	
	100 mAs	20 mAs	100 mAs	20 mAs
Lesion 1	4.3 $\pm$ 2.4%	3.6 $\pm$ 5.6%	10.8 $\pm$ 2.7%	11.9 $\pm$ 5.6%
Lesion 2	0.3 $\pm$ 2.3%	−0.5 $\pm$ 4.4%	6.8 $\pm$ 1.5%	11.1 $\pm$ 4.8%
Lesion 3	5.0 $\pm$ 2.3%	2.1 $\pm$ 4.8%	1.9 $\pm$ 1.7%	4.6 $\pm$ 5.7%

Summary of PBs and SPEs obtained from the phantom study, simulation study and theoretical derivation.

Table 3

PB ± SPE (%)	Thin protocol, 100mAs				Thick protocol, 100 mAs			
	8 mm, 104/73 HU	9 mm, 118/73 HU	9 mm, 27/73 HU	9 mm, 27/73 HU	8 mm, 104/73 HU	9 mm, 118/73 HU	9 mm, 27/73 HU	9 mm, 27/73 HU
CRLB	±1.9%	±1.4%	±1.4%	±1.4%	±1.8%	±1.3%	±1.3%	±1.3%
Simulated ^c	-0.3 ± 3.3%	0.1 ± 1.9%	0.0 ± 1.9%	0.0 ± 1.9%	0.0 ± 2.7%	-0.0 ± 1.6%	-0.1 ± 1.7%	-0.1 ± 1.7%
Simulated [#]	6.2 ± 4.3%	3.8 ± 2.1%	4.0 ± 2.0%	4.0 ± 2.0%	6.1 ± 4.8%	5.9 ± 1.9%	5.4 ± 1.8%	5.4 ± 1.8%
Phantom [#]	4.3 ± 2.4%	0.3 ± 2.3%	5.0 ± 2.3%	5.0 ± 2.3%	10.8 ± 2.7%	6.8 ± 1.5%	1.9 ± 1.7%	1.9 ± 1.7%

^cTarget objects were centered.

[#]Target objects were not centered.

Table 4

PBs and SPEs of PW matched-filter estimates for phantom studies with nuisance parameters fixed.

PB $\pm$ SPE (%) (Nuisance fixed)	Thin protocol		Thick protocol	
	100 mAs	20 mAs	100 mAs	20 mAs
Lesion 1	$10.1 \pm 4.5\%$	$19.1 \pm 6.9\%$	$17.1 \pm 3.8\%$	$24.3 \pm 5.4\%$
Lesion 2	$8.3 \pm 2.6\%$	$25.7 \pm 5.2\%$	$14.3 \pm 3.4\%$	$27.0 \pm 5.1\%$
Lesion 3	$-3.6 \pm 2.5\%$	$-0.5 \pm 5.0\%$	$2.7 \pm 2.5\%$	$3.9 \pm 6.0\%$