Published in final edited form as:

Acad Radiol. 2017 October ; 24(10): 1203–1215. doi:10.1016/j.acra.2017.04.007.

Measurement accuracy of atherosclerotic plaque structure on CT using phantoms to establish ground truth

Samantha St Pierre, BS¹, Jenifer Siegelman, MD, MPH², Nancy A. Obuchowski, PhD³, Xiaonan Ma, MS¹, David Paik, PhD¹, and Andrew J. Buckler, MS^{1,*}

¹Elucid Bioimaging Inc., 225 Main Street, Wenham, MA 01984

²Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115

³Quantitative Health Sciences, Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, Ohio 44195

Abstract

Rationale and objectives—The purpose of this study was to characterize analytic performance of software-aided arterial vessel structure measurements across a range of scanner settings for CT angiography (CTA) where ground truth is known. We characterized performance for measurands which may be efficiently measured for clinical cases without use of software as well as those which may be done manually but which is generally not done due to the effort level required unless software is employed.

Materials and Methods—Four measurands (lumen area, stenosis, wall area, wall thickness) were evaluated using tissue-mimicking phantoms to estimate bias, heteroscedasticity, and limits of quantitation both pooled across scanner settings and individually for eight different settings. Reproducibility across scanner settings was also estimated.

Results—Measurements of lumen area have a near constant bias of +1.3mm for measurements ranging from 3–40mm²; stenosis bias is +7% across a 30–70% range; wall area bias is +14% across a 50–450mm² range; and wall thickness bias is +1.2mm across a 3–9mm range. All measurements possess properties that make them suitable for measuring longitudinal change. Lumen area demonstrates the most sensitivity to scanner settings (bias from as low as +.1mm to as high as +2.7mm); wall thickness demonstrates negligible sensitivity.

Conclusions—Variability across scanner settings for lumen measurands was generally higher than bias for a given setting. The converse was true for the wall measurands, where variability due

*Corresponding Author: 225 Main Street Suite 15, Wenham, MA01984, andrew.buckler@elucidbio.com, Telephone: 1-978-468-0508, Fax: 1-978-468-0527.

Authorship: The authorship of this manuscript is consistent with the International Committee of Medical Journal Editors "Uniform Requirements for Manuscripts Submitted to Biomedical Journals" and conforms to specifications in the "Authorship Responsibility, Copyright Transfer and Financial Disclosure" form.

Conflicts of Interest: none.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

to scanner settings was very low. Both bias and variability due to scanner settings of vessel structure were within clinically useful levels.

INTRODUCTION

Assessment of atherosclerotic plaque is an essential diagnostic and treatment-planning tool. Extensive efforts have begun to provide quantitative information to the physician in assessing an individual patient's immediate cardiovascular risk. Sophisticated and powerful, postprocessing of Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) have been proposed [1–5], and this technology offers non-invasive alternatives to invasive catheterization procedures.

To evaluate an image-based analysis, performance should be compared to ground truth evaluated across a spectrum of disease and imaging protocols routinely used in clinical practice. In this work, we use synthetic vessel phantoms with readily measured ground truth values. The phantoms are fabricated using vessel tissue-mimicking material, as the true vessel wall thickness value in excised tissue would be difficult to ascertain, and bias is not expected to differ with clinical data. Our group has also conducted reader variability studies using clinical data to study precision based on the assumption that patient anatomy variability contributes significantly to the variability in the measurements. Our use of phantoms to estimate bias, and clinical data to estimate variability, is based on this rationale.

MATERIALS AND METHODS

The bias and precision of CTA-derived arterial vessel structure measurements was estimated using a set of precision-machined phantoms where ground-truth measurements are assessed by a hand-held micrometer. The phantoms were fabricated to specifications to mimic appropriate CT characteristics, in sizes that represented a wide clinical spectrum of vessel sizes. The phantoms used in this study were designed for the validation of CTA images of peripheral and central vessels.

The vascular phantom set consists of hollow Noryl tubes with a stenotic region in the center of the length. The density of Noryl is 1.06 g/ml, which was selected to mimic the density of vessel wall in CT analysis and dimensions were chosen to represent a range of human arteries (see Figure 1). The phantoms used in this study were manufactured at Trans Form Plastics Inc. (Danvers, MA), one to represent the aorta, two to represent femoral arteries, three to represent carotid arteries and one to represent vertebral arteries.

The phantoms were filled with contrast media utilized in routine CT angiography and scanned in a range of scanner settings representative of current clinical practice. Statistical measures of bias were estimated from these data.

Imaging

For the scans, the phantoms were filled with diluted contrast agent (Omnipaque, GE Healthcare, Chicago, IL) between 10–12 mg Iodine/ml to achieve the same contrast between vessel wall and lumen found in patient CTA scans at 100–120 kVp (based on published

relationship of iodine concentration vs. Hounsfield units for 80–120 kVp) [6]. The phantoms were suspended in a plastic cage, submerged in a box of vegetable oil (to emulate surrounding fat), and scanned simultaneously.

Imaging was performed across image acquisition protocols representative of clinical practice using a Siemens SOMATOM Definition Flash at the Brigham and Women's Hospital (Boston, MA). 120 and 100 kVp, radiation exposure levels 5–15 mGy, and various reconstruction kernels were used. Apart from parameters that were varied in Tables 2–5, the scans also had a slice thickness of 0.75 mm, single collimation width of 0.6 mm, total collimation width of 38.4 mm, spiral pitch factor of 0.8, and pixel spacing of 0.64 mm.

Target Definition and Analysis

The eight image series described in Tables 1–4 were loaded into dedicated software (vascuCAP, Elucid Bioimaging Inc., Wenham, MA) for analysis. Based on user initialization by clicking on relevant points on the image, the software generates the path lines and initial lumen and vessel segmentations. Editing, using built-in CTA-derived tools of the segmentations generated by the software, was used to remove or mitigate artifacts (e.g., bubbles) in the phantom.

Target Initialization—For each vessel, a sequence of points along each vessel centerline was manually created by clicking on a volume rendering of the CTA. 3D coordinates and an initial lumen radius were estimated at each point by ray-casting.

Lumen Segmentation—In order to avoid false positive inclusion of calcified regions of plaque as lumen, after the software attenuates detected regions of calcium, the segmentation of the lumen is performed by an algorithm which utilizes a numerical analysis of surfaces and shapes referred to as a level set model which allows numerical computations involving curves and surfaces on a fixed Cartesian grid and using a threshold determined as the Otsu threshold of the region surrounding the target initialization variable radius tube [7, 8].

Wall Segmentation—The outer wall of the vessel (serosa/adventitia) is segmented using a geodesic active contour level set evolution, initialized with the union of lumen segmentation after min/max curvature evolution and the target initialization centerline expanded by 1.2 mm to account for the wall at areas of complete stenosis [9, 10]. The level set edge potential function captures outer wall image edges using distance to Canny edge detections transformed through a logistic sigmoid function [11].

Segmentation Editing—Readers are provided tools to manually edit the lumen and/or the wall segmentation. For each mode, results are visualized in real time in both 2D and 3D views. In “Draw Mode”, the mouse cursor becomes a variable radius 3D sphere ‘paintbrush’ that can either add or subtract regions.

Figure 2 depicts examples of phantoms from our study.

Quantitative Measurements—The following measurements were made based on the segmentations. Lumen and wall area were assessed at multiple cross sections, whereas

maximum stenosis and maximum wall thickness were determined across the full set of cross sections to determine where the maximum value was found.

- **Lumen Area:** Geometric area of segmented lumen, calculated per-cross section, expressed in units of mm².
- **Maximum Stenosis:** Calculated as the (1 - ratio of minimum lumen diameter with plaque to reference lumen diameter without plaque) x100 (also calculated on an area basis but clinical practice uses diameters more commonly). Maximum Stenosis is expressed as a ratio >0.
- **Wall Area:** Geometric area of the segmented wall minus the segmented lumen, calculated per-cross section, expressed in units of mm².
- **Maximum Wall Thickness:** Calculated by measuring the maximum cross-sectional wall thickness, expressed in units of mm.

Statistical Analysis

In this work, we use metrics sanctioned by a broad-based consensus effort organized by the Radiological Society of North America [12, 13]. We plotted the CTA-derived value calculated by the software versus ground truth value. We evaluated the data for (i) associations between the magnitude of the true measurement and bias; (ii) associations between the magnitude of the true measurement and heteroscedasticity in the CTA-derived measurements; and (iii) limits of quantitation.

A linear model was fit (for all the data or for a subset, depending on the findings from the plot) where the dependent variable is the CTA-derived measurement (Y) (or its log, log(Y)) and the independent variable is the true value (X) (or its log, log(X)), as follows: $E(Y/X) = \beta_0 + \beta_1 X$. We constructed a 95% CI for the intercept β_0 using a t-statistic as the pivotal statistic. CIs not containing 0 indicate constant bias. We also constructed a 95% CI for the slope β_1 . CIs not containing 1 indicate proportional bias [14]. These calculations allowed conclusions to be drawn regarding the linearity of the measurements [12]. This is important to indicate whether the measurand may be used for accurate longitudinal change determination. For measurands that demonstrated a non-zero quadratic term, which indicates non-linearity, the effect of the degree of non-linearity was assessed using simulations.

The precision of the CTA-derived measurements was estimated by the standard deviation:

$\sqrt{\frac{1}{n-1} \sum_{i=1}^n (Y_i - X_i - \bar{d})^2}$ where $\bar{d}$ is the sample mean of the differences, $\bar{d} = \frac{1}{n} \sum_{i=1}^n (Y_i - X_i)$ [1]. We constructed a 95% CI for the standard deviation using a chi square statistic as the pivotal statistic.

We present the bias profile (bias of CTA-derived measurements for various ranges of true values) and precision profile (standard deviation of CTA-derived measurements from subjects with similar true values for various ranges of the true value) as summaries of performance [15].

We estimated these metrics both pooled and stratified over scanner settings. Detailed results may be found in the supplementary material.

Additionally, we analyzed the reproducibility coefficient (RDC), as well as %RDC, across the various scanner settings to provide insight into the relative magnitude of the difference when different acquisition parameters are used [16]. The RDC was defined as:

$$RDC = 1.96 \sqrt{2\sigma_{\varepsilon}^2} = 2.77\sigma_{\varepsilon}$$

where σ_{ε}^2 is the within-vessel across-settings variance. The range in which two measurements on the same vessel were expected to fall for 95% of measurements is given by $[-RDC, +RDC]$ [17]. Additionally, the within-vessel coefficient of variability ($wCV_{\text{inter-settings}}$) was calculated as a measure of precision for single measurements but which may be taken according to different acquisition settings [16]. It was calculated in an analogous fashion, but dividing each vessel-based σ_{ε}^2 by the square of the mean of the two measurements. Both $wCV_{\text{inter-settings}}$ and %RDC are relative measures proportional to the magnitude of the vessel's size. Since we were interested in how the metrics changed for differing vessel sizes, we plotted the percentage metrics as a function of vessel size.

RESULTS

There were 56 measurements (7 phantom vessels X 8 CT protocols): 8 in the aorta phantom (one for each of the scanner settings), 24 in the carotid artery phantoms (one for each of the scanner settings per each of the 3 carotid phantoms), 16 in the femoral artery phantoms (one for each of the scanner settings per phantom), and 8 in the vertebral artery phantoms (one for each of the scanner settings). The reason why multiple phantoms were used for some of the vessel types (carotid and femoral) was to provide a range of dimensions as seen in practice. Left and right designations were used by the software, but this designation was not relevant to the quantitative accuracy assessment.

Lumen Area

The true value of Lumen Area for each phantom was calculated as the quantity $\pi * \left(\frac{B}{2}\right)^2$ (letter designations from Figure 1): 8 in the aorta phantom (one for each of the scanner settings), 24 in the carotid artery phantoms (one for each of the scanner settings per phantom), 16 in the femoral artery phantoms (one for each of the scanner settings per phantom), and 8 in the vertebral artery phantoms (one for each of the scanner settings). Figure 3 illustrates the CTA-derived measurements by true value. There were no non-measurable CTA-derived measurements. Figure 3 (*Bottom Left*) illustrates how the precision across scanner settings, expressed as the %RDC, changed based on the vessel size.

Table 1 (*All Arteries and Scanner Settings*) summarizes the estimated mean bias and standard deviation, binned by ground truth values, pooled over all vessels. From the fitted model of the CTA-derived measurements on the true values, the 95% CI for a quadratic term was $[-0.007, +0.002]$, indicating no evidence for a non-linear fit ($p=0.345$). From the fitted

linear model, the 95% CI for the intercept was [-0.144, 1.237] and the 95% CI for the slope was [1.014, 1.086], indicating evidence for linearity but with a slope slightly greater than one.

When vessel type and an interaction term for vessel type and true value were added to the linear model, the interaction term was not significant ($p=0.379$), suggesting that the slope does not differ by vessel type. There was also no evidence for differences in constant bias between vessels ($p=0.351$).

Table 1 (*Stratified by Scanner Settings*) summarizes the results from stratified analyses. Detailed plots of the stratified analysis results for each series number may be found in supplementary material.

Maximum Stenosis

The true value of Maximum Stenosis for each phantom was calculated as the quantity $(1 - (B/A))$ (letter designations from Figure 1). Figure 4 illustrates the CTA-derived measurements by ground truth value, pooling over vessels. There were no non-measurable CTA-derived measurements, and no evidence that the CTA-derived measurements were more heterogeneous at lower or higher true values.

Table 2 (*All Arteries and Scanner Settings*) summarizes the estimated mean bias and standard deviation, binned by ground truth values, pooled over all vessels. From the fitted model of the log of the CTA-derived measurements on the log of the true values, the 95% CI for a quadratic term was [-0.75, 0.62], indicating little evidence for a non-linear fit ($p=0.851$). From the fitted linear model, the 95% CI for the intercept was [0.002, 0.146] and the 95% CI for the slope was [0.921, 1.100]; these indicate a small overall constant bias, but no evidence for proportional bias.

When vessel type and an interaction term for vessel type and log true value were added to the linear model, the interaction term was not significant ($p=0.605$), suggesting that the slope does not differ by vessel type. There is evidence, however, of differences in constant bias between vessels ($p=0.002$): mean % bias was -3.1% in the aorta, +11.5% in the carotid, +8.1% in the femoral, and +3.6% in the vertebral.

Table 2 (*Stratified by Scanner Settings*) summarizes the results from stratified analyses. Detailed plots may be found in supplementary material.

Figure 4 illustrates how the % *RDC* changed for differing vessel sizes.

Wall Area

The true value of Wall Area for each phantom was calculated as the quantity

$\pi * \left(\left(\frac{(2 * (E + F)) + A}{2} \right)^2 - \left(\frac{B}{2} \right)^2 \right)$ (letter designations from Figure 1). Figure 5 (Top Left) illustrates the CTA-derived measurements by ground truth value, pooling over vessels. There were no non-measurable CTA-derived measurements, and no evidence that the CTA-derived measurements were more heterogeneous at lower or higher true values.

Table 3 (*All Arteries and Scanner Settings*) summarizes the estimated mean bias and standard deviation, binned by ground truth values, pooled over all vessels. From the fitted model of the log of the CTA-derived measurements on the log of the true values, the 95% CI for a quadratic term was $[-0.052, 0.016]$, indicating little evidence for a non-linear fit ($p=0.298$). From the fitted linear model, the 95% CI for the intercept was $[0.053, 0.291]$ and the 95% CI for the slope was $[0.966, 1.016]$; these indicate a small overall constant bias, with slope close to one.

When vessel type and an interaction term for vessel type and log true value were added to the linear model, the interaction term was statistically significant ($p=0.039$), suggesting that the slope differs by vessel type. There is marginal evidence that the constant bias also differs between vessels ($p=0.055$). The estimated slopes and intercepts by vessel type are as follows: estimated slope was 0.76 and intercept was 1.17 in the carotid, and the estimated slope was 1.15 and intercept was -0.62 in the femoral (sample sizes are too small in the aorta and vertebral to fit the regression lines).

Table 3 (*Stratified by Scanner Settings*) summarizes the results from stratified analyses. Detailed plots may be found in supplementary material.

Figure 5 (*Bottom Left*) illustrates how the % *RDC* changed for differing vessel sizes.

Maximum Wall Thickness

The true value of Maximum Wall Thickness for each phantom was calculated as the quantity $G+E+(0.5*(A-B))$ (letter designations from Figure 1). Figure 6 illustrates the CTA-derived measurements by ground truth value. There were no non-measurable CTA-derived measurements, and no evidence that the CTA-derived measurements were more heterogeneous at lower or higher true values.

Table 4 (*All Arteries and Scanner Settings*) summarizes the estimated mean bias and standard deviation, binned by ground truth values, pooled over all vessels. From the fitted model of the CTA-derived measurements on the true values, the 95% CI for a quadratic term was $[-0.038, -0.003]$, indicating some evidence for a non-linear fit ($p=0.022$). From the fitted linear model (setting the quadratic term equal to zero), the 95% CI for the intercept was $[0.777, 1.135]$ and the 95% CI for the slope was $[1.018, 1.086]$, indicating evidence for an overall constant bias and proportional bias.

When vessel type and an interaction term for vessel type and true were added to the linear model, the interaction term was not significant ($p=0.127$), suggesting that the slope does not differ by vessel type. There is some evidence of differences in bias between vessels ($p=0.049$): mean bias was 1.32mm in the aorta, 1.18mm in the carotid, 1.29mm in the femoral, and 1.01mm in the vertebral.

Table 4 (*Stratified by Scanner Settings*) summarizes the results from stratified analyses. Detailed plots may be found in supplementary material.

Figure 6 illustrates how the % *RDC* changed for differing vessel sizes.

Discussion

The goal of this study was to estimate the bias and between-settings variance of CTA-derived measurements of vessel tissue-mimicking phantoms where ground truth is known. We use metrics endorsed by a broad-based consensus effort organized by the Radiological Society of North America [12, 13]. Bias is an estimate of systematic measurement error; it is the difference between the average (expected value) of measurements made on the same object and its true value. We also explored how different CT acquisition parameters impact bias estimation using stratified analyses. This effort fits within our overall goal of developing a systematic approach for evaluating the technical performance of quantitative imaging biomarkers for atherosclerosis. Of the published articles documenting analytic performance of vascular measurements, very few address bias relative to a reliable ground truth. Representative analytic performance studies of various modalities on the task of measuring vascular structure have been performed for ultrasound [18–20], MR [21–25], PET [26], and CT [26, 27], none of which address bias relative to a reliable ground truth but rather either make comparisons across modalities or address issues of repeatability or reproducibility under a variety of conditions. The field becomes increasingly objective as further studies are undertaken and published that rigorously measure bias using ground truth standards. We discuss the results of this study in the paragraphs which follow.

Lumen Area

CTA-derived measurements of lumen area have a low, almost constant bias in the original scale of mm^2 ; the average over-estimation relative to the true values is 1.3mm across measurements ranging from 3–40 mm^2 , resulting in as much as a 7% underestimation of stenosis. The bias of lumen area is higher at low values, and decreases for higher values. There was no evidence that the CTA-derived measurements were more heterogeneous at lower or higher true values. The measurements exhibit the property of linearity. The regression slope of the CTA-derived measurements on the true lumen area value is slightly greater than one (95% CI for the slope was [1.014, 1.086]) but this has a negligible effect on CIs for longitudinal change in lumen area for physiologically plausible levels of change.

The mean percent bias ranged from –1 to +36% across scanner settings. The analysis showed that there are relatively large differences for mean bias. Small lumen areas are particularly susceptible to this variation. We were not able to determine which settings contributed the most to the bias, and/or if other patterns gave rise to the differences, but exploring this would be important in future studies, if area based assessments of stenosis are important. The reproducibility coefficient analysis illustrates that when the lumen area is greater than 30 mm^2 , there is less variation among scanner settings.

Maximum Stenosis

CTA-derived measurements of maximum stenosis have a low constant bias in the percentage scale; the average over-estimation relative to the true values is 7% across measurements ranging from 30–70%. There was no evidence that the CTA-derived measurements were more heterogeneous at lower or higher true values. The measurements exhibit the property of linearity.

The mean % bias ranged from -5 to 12% across scanner settings. The analysis illustrates that the bias for the maximum stenosis readings are moderately sensitive to changes in scanner settings. The reproducibility coefficient analysis shows that the highest between-scanner differences occur for lower values of stenosis (<50%), which is a natural consequence of the variability manifesting itself as a higher percentage of lower measurements, even when the variance in the original units is constant.

Wall Area

CTA-derived measurements of Wall Area have a low constant bias in the percentage scale; the average over-estimation relative to the true values is 14% across measurements ranging from 50–450mm². There was no evidence that the CTA-derived measurements were more heterogeneous at lower or higher true values. The measurements exhibit the property of linearity.

The mean % bias ranged from 10 to 19% across scanner settings. The analysis shows low variability across scanner settings. This is further manifested in a low reproducibility coefficient, particularly as the Wall Area values increase. This is good support for automated outer wall detection/segmentation. The human eye has trouble determining these boundaries in CT in normal vessels [28, 29].

Maximum Wall Thickness

CTA-derived measurements of Maximum Wall Thickness have a low almost constant bias in the original scale; the average over-estimation relative to the true values is 1.2mm across measurements ranging from 3–9mm with some evidence of a non-linear relationship with the reference standard values. There was no evidence that the CTA-derived measurements were more heterogeneous at lower or higher true values. The estimate of the regression coefficient for the quadratic term is small though (0.02), and the estimate of the linear term is slightly over one (i.e., 1.3). To interpret these estimates, we evaluated their effect on the estimation of a new patient's wall thickness measurement. Specifically, we evaluated the coverage of 95% CIs for a new patient's true wall thickness. We used the following formula for constructing a 95% CI for a new patient's true wall thickness, where the quadratic effect is ignored and the linear slope is assumed to be one:

$$Y - \hat{\beta}_0 \pm 1.96 \times \widehat{wSD}$$

where Y is the CTA-derived measurement for a new patient, $\hat{\beta}_0$ is an estimate of the constant bias, and wSD is the within-subject SD. We simulated data (i.e. Y values) using the point estimates for the quadratic, linear, and intercept terms from the fitted model with the phantom data. From these data, we estimated $\hat{\beta}_0$ (ignoring non-linear effects and assuming a linear slope of one). Next, we simulated data for new patients using the same point estimates for the quadratic, linear, and intercept terms from the fitted model with the phantom data. Finally, we determined the proportion of simulated patients whose 95% CI contains the true wall thickness value. The coverage of the 95% CIs for a new patient's true wall thickness was at the nominal level, i.e. 94.8%. In a second set of simulations, we evaluated the effect of the bias on the estimation of a new patient's change in wall thickness measurements.

Specifically, we evaluated the coverage of 95% CIs for a new patient's true change in wall thickness. We used the following formula for constructing a 95% CI for the change:

$$(Y_2 - Y_1) \pm 1.96 \times \{ \sqrt{2(\overline{wSD})^2} \}$$

Note that this CI construction assumes a linear slope equal to one, and ignores non-linear effects.

When the true change in wall thickness was zero, the coverage was at the nominal level. The CI coverage remained >94.0% for true changes up to 3.0mm. At a 3.0mm change, the coverage dropped to 93.9%. We conclude that the non-zero quadratic effect and the slope being slightly greater than one for the CTA-derived measurements for wall thickness are insignificant for all practical purposes.

The analysis illustrates that the Maximum Wall Thickness bias does not vary greatly across scanner settings.

Our study has limitations, including for example that the analyses were performed with a single CT system, as well as a limited range of vessel geometries that were linear in nature. We have mitigated these limitations by testing across a heterogeneous set of scanner settings, and selecting representative geometries for the phantom designs that represent a wide range of clinical relevant presentations. Assessment of any semi-automated algorithm that allows editing inevitably includes differences in both experience and patience between algorithm operators in terms of how many edits they make.

In conclusion, we demonstrated clinically useful levels of analytic performance for semiautomated measurements of anatomic vessel structure with known ground truth on four measures, two of which may be calculated manually (stenosis and wall thickness) and two which are generally not feasible without the aid of software (lumen and wall area). We also explored the effect of different CT acquisition parameters as to their effect on bias using stratified analysis. These results suggest an area of further research to bring greater insight into why some settings result in different performance than others in order to derive and promulgate optimized protocols to minimize variation and optimize performance.

Acknowledgments

This work was supported in part by National Heart, Lung, and Blood Institute [grant number R44HL126224] and National Science Foundation [grant number 1353532].

Abbreviations

CT	Computed Tomography
MRI	Magnetic Resonance Imaging
CTA	Computed Tomography Angiography
CI	Confidence Intervals

RDC	Reproducibility Coefficient
wSD	Within-subject Standard Deviation

References

1. Zavodni AEH, et al. Carotid Artery Plaque Morphology and Composition in Relation to Incident Cardiovascular Events: The Multi-Ethnic Study of Atherosclerosis (MESA). *Radiology*. 2014; 271(2):381–389. [PubMed: 24592924]
2. Puchner SB, et al. High-Risk Plaque Detected on Coronary CT Angiography Predicts Acute Coronary Syndromes Independent of Significant Stenosis in Acute Chest Pain: Results From the ROMICAT-II Trial. *J Am Coll Cardiol*. 2014; 64(7):684–692. [PubMed: 25125300]
3. Hermann DM, et al. Coronary Artery Calcification Is an Independent Stroke Predictor in the General Population. *Stroke*. 2013
4. Joshi FR, et al. Non-invasive imaging of atherosclerosis. *European Heart Journal–Cardiovascular Imaging*. 2012; 13(3):205–218. [PubMed: 22277118]
5. Muntendam P, et al. The BioImage Study: Novel approaches to risk assessment in the primary prevention of atherosclerotic cardiovascular disease—study design and objectives. *Am Heart J*. 160(1):49–57.e1.
6. Bae KT. Intravenous contrast medium administration and scan timing at CT: considerations and approaches. *Radiology*. 2010; 256(1):32–61. [PubMed: 20574084]
7. Yushkevich PA, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. 2006; 31(3):1116–1128. [PubMed: 16545965]
8. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Sys, Man, Cyber*. 1979; 9(1):62–66.
9. Caselles V, Kimmel R, Sapiro G. Geodesic Active Contours. *International Journal on Computer Vision*. 1997; 22(1):61–97.
10. Malladi R, Sethian JA. Flows under Min/Max Curvature and Mean Curvature. *Graphical Models and Image Processing*. 1996; 58(2):127–141.
11. Canny J. A Computational Approach To Edge Detection. *IEEE Trans Pattern Analysis and Machine Intelligence*. 1986; 8(6):679–698.
12. Kessler LG, et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. *Statistical Methods in Medical Research*. 2014
13. Obuchowski NA, et al. Quantitative imaging biomarkers: a review of statistical methods for computer algorithm comparisons. *Stat Methods Med Res*. 2015; 24(1):68–106. [PubMed: 24919829]
14. Algorithm-Comparison-Working-Group. Quantitative Imaging Biomarkers: A Review of Statistical Methods for Computer Algorithm Comparisons. 2013 Submitted to SMMR.
15. Case-Example-Working-Group. Statistical Issues in the Comparison of Quantitative Imaging Biomarker Algorithms using Pulmonary Nodule Volume as an Example. 2013. Submitted to SMMR
16. QIBA-Performance-Working-Group. Review of Statistical Methods for Technical Performance Assessment. 2014. Submitted to SMMR
17. Barnhart HX, Barboriak DP. Applications of the repeatability of quantitative imaging biomarkers: A review of statistical analysis of repeat data sets. *Translational Oncology*. 2009; 2(4):231–235. [PubMed: 19956383]
18. Mac Ananey O, Mellotte G, Maher V. Comparison of Semi-Automated and Manual Measurements of Carotid Intima-Media Thickening. *BioMed Research International*. 2014; 2014:4.
19. Bianchini E, et al. Assessment of Carotid Stiffness and Intima-Media Thickness From Ultrasound Data: Comparison Between Two Methods. *Journal of Ultrasound in Medicine*. 2010; 29(8):1169–1175. [PubMed: 20660450]

20. Ludwig M, et al. Reproducibility of 3-dimensional ultrasound readings of volume of carotid atherosclerotic plaque. *Cardiovascular Ultrasound*. 2008; 6(1):42. [PubMed: 18727816]
21. Sun J, et al. Carotid magnetic resonance imaging for monitoring atherosclerotic plaque progression: a multicenter reproducibility study. *Int J Cardiovasc Imaging*. 2014;1–9. [PubMed: 24687407]
22. Alizadeh Dehnavi R, et al. Assessment of the carotid artery by MRI at 3T: A study on reproducibility. *Journal of Magnetic Resonance Imaging*. 2007; 25(5):1035–1043. [PubMed: 17457802]
23. Saam T, et al. Sample size calculation for clinical trials using magnetic resonance imaging for the quantitative assessment of carotid atherosclerosis. *J Cardiovasc Magn Reson*. 2005; 7(5):799–808. [PubMed: 16353440]
24. Duivenvoorden R, et al. In Vivo Quantification of Carotid Artery Wall Dimensions: 3.0-Tesla MRI Versus B-Mode Ultrasound Imaging. *Circulation: Cardiovascular Imaging*. 2009; 2(3):235–242. [PubMed: 19808598]
25. Wasserman BA, et al. MRI measurements of carotid plaque in the atherosclerosis risk in communities (ARIC) study: Methods, reliability and descriptive statistics. *Journal of Magnetic Resonance Imaging*. 2010; 31(2):406–415. [PubMed: 20099354]
26. Kwee RM, et al. Multimodality Imaging of Carotid Artery Plaques: 18F-Fluoro-2-Deoxyglucose Positron Emission Tomography, Computed Tomography, and Magnetic Resonance Imaging. *Stroke*. 2009; 40(12):3718–3724. [PubMed: 19875738]
27. de Weert TT, et al. In Vivo Characterization and Quantification of Atherosclerotic Carotid Plaque Components With Multidetector Computed Tomography and Histopathological Correlation. *Arterioscler Thromb Vasc Biol*. 2006; 26(10):2366–2372. [PubMed: 16902158]
28. Schepis T, et al. Quantification of non-calcified coronary atherosclerotic plaques with dual-source computed tomography: comparison with intravascular ultrasound. *Heart*. 2010; 96(8):610–615. [PubMed: 19933294]
29. Saba L, et al. Semiautomated and Automated Algorithms for Analysis of the Carotid Artery Wall on Computed Tomography and Sonography. *J Ultrasound Med*. 2013; 32:665–674. [PubMed: 23525393]

Figure 1.

Physical geometry of vascular phantoms. Dimensions designated by the letters A–G are utilized in calculations of specific ground truth values, as indicated in Results section, where A is the reference lumen, B is the stenotic lumen, C is the length of the stenosis, D is the length over which the wall is thickest, E is the difference between where wall is thickest and where it is thinnest, F is the length over which the wall is thinnest, and G is the thinnest wall. Each phantom has different dimensions so as to coincide with typical arterial presentations.

Figure 2.

Examples of vessels from our study. (a) Examples of phantoms scanned in the study. From left to right: Left carotid, left carotid, left femoral, right femoral and aorta (b) Axial image of all phantoms without lumen and wall boundaries defined by CTA-derived. (c) Axial image of all phantoms with defined lumen and wall boundaries.

Figure 3.

(Top Left): Scatter plot of CTA-derived Lumen Area measurements versus ground truth, pooled over vessels in the original scale of mm². The fit of the line indicates good linearity of the measurements. (Top Right) Plot of limit of quantitation in CTA-derived Lumen Area measurements by ground truth, pooled over vessels in the scale of percent mm². The bias is higher at low ground truth values, whereas, the bias is nearly constant over the 10–40% range. (Bottom Left): Plot of precision in terms of RDC (see Methods section) of CTA-derived Lumen Area measurements, pooled over vessels in the scale of percent mm². There is more variability in the measurements at low areas.

Figure 4.

(Top Left): Scatter plot of CTA-derived Maximum Stenosis measurements versus ground truth, pooled over vessels in the original scale of percent expressed as a ratio. The fit suggests good linearity. (Top Right) Plot of limit of quantitation in CTA-derived Maximum Stenosis measurements by ground truth, pooled over vessels in the scale of percent ratio. The data suggests constant bias in the 0.3–0.9 range. (Bottom Left): Plot of precision in terms of RDC (see Methods section) of CTA-derived Maximum Stenosis measurements, pooled over vessels in the scale of percent ratio. The precision increases over the range of 0.3 to 0.8.

Figure 5.

(Top Left): Scatter plot of CTA-derived Wall Area measurements versus ground truth, pooled over vessels in the original scale of mm², indicating linearity (Top Right) Plot of limit of quantitation in CTA-derived Wall Area measurements by ground truth, pooled over vessels in the scale of percent mm². (Bottom Left): Plot of precision, in terms of RDC, of CTA-derived Wall Area measurements, pooled over vessels in the scale of percent mm².

Figure 6.

(Top Left): Scatter plot of CTA-derived Maximum Wall Thickness measurements versus ground truth, pooled over vessels in the original scale of mm. (Top Right) Plot of limit of quantitation in CTA-derived Maximum Wall Thickness measurements by ground truth, pooled over vessels in the scale of percent mm. (Bottom Left): Plot of reproducibility in terms of RDC as defined in the methods section of CTA-derived Maximum Wall Thickness measurements, pooled over vessels in the scale of percent mm.

Table 1

(Top): Bias Profile for Lumen Area across arteries and scanner settings. (Bottom): Stratified Analysis for Lumen Area. The bins were established according to approximate consistency observed in result (the vertebral and the smallest carotid being the first bin, the other carotids and the smallest femoral in the middle bin, and the larger femoral with the aorta in the third bin).

Pooled Across Arteries and Scanner Settings						
True Value (mm ²)	3.14–5.7	7.0–13.0	28.0–39.0	Overall		
# observations	16	24	16	56		
Mean bias * [95% CI]	0.94 [0.16, 1.73]	0.94 [0.25, 1.62]	2.14 [1.04, 3.23]	1.28 [0.81,1.76]		
Mean % bias ** [95% CI]	23.6% [4.1%, 43.1%]	8.2% [1.4%, 15.0%]	6.5% [3.1%, 9.9%]	12.1% [5.9%, 18.3%]		
SD [95% CI]	1.48 [1.09, 2.29]	1.62 [1.26, 2.27]	2.06 [1.52, 3.18]	1.77 [1.49, 2.18]		
Pooled Across Arteries but Stratified By Scanner Settings						
KVP	100	100	100	120	120	120
mAs	156	156	325	156	606	606
CTDIvol ****	4.95	4.95	10.3	8.12	16.9	16.88
Filter	B30f	soft B30f	B30f	B26f	B30f	I30f
Mean bias * [95% CI]	1.10 [0.68, 2.89]	1.96 [0.25, 3.67]	0.25 [0.73, 1.22]	0.96 [−0.53, 2.46]	2.67 [1.68, 3.66]	0.53 [−0.92, 1.99]
Mean % bias ** [95% CI]	11% [14%, 36%]	15% [7%, 24%]	−2% [−11%, 7%]	7% [9%, 51%]	30% [9%, 51%]	−1% [−11%, 8%]
SD [95% CI]	1.93 [1.24, 4.25]	1.85 [1.19, 4.07]	1.05 [0.68, 2.32]	0.81 [0.52, 1.79]	2.21 [1.42, 4.87]	1.57 [1.01, 3.46]

* bias defined as (CTA-derived value - true value).

** % bias defined as (CTA-derived value - true value)/true value.

*** as indicated in the DICOM tag

Table 2

(Top): Bias Profile for Maximum Stenosis across arteries and scanner settings. (Bottom): Stratified Analysis for Maximum Stenosis. The bins were established according to approximate consistency observed in result (the vertebral and the two smallest carotids being the first bin, the other carotid and the smallest femoral in the middle bin, and the larger femoral with the aorta in the third bin).

Pooled Across Arteries and Scanner Settings						
True Value (Ratio)	0.32–0.35	0.50	0.63–0.67	Overall		
# observations	24	8	24	56		
Mean bias [*] [95% CI]	0.022 [0.007, 0.037]	0.059 [0.037, 0.081]	0.045 [0.016, 0.075]	0.037 [0.023, 0.052]		
Mean % bias ^{**} [95% CI]	6.48 [2.08, 10.88]	–11.87 [7.47, 16.26]	6.77 [2.26, 11.29]	7.38 [4.69, 10.06]		
SD [95% CI]	0.036 [0.028, 0.050]	0.026 [0.017 0.053]	0.070 [0.054, 0.098]	0.054 [0.045, 0.066]		
Pooled Across Arteries but Stratified By Scanner Settings						
KVP	100	100	100	120	120	120
mAs	156	325	325	156	606	606
CTDIvol (mGy) ^{***}	4.95	10.3	10.3	8.12	16.9	16.88
Filter	B30f	soft B30f	soft B30f	B26f	B30f	I30f
Mean bias [*] [95% CI]	0.04 [0.01, 0.07]	0.03 [0.02, 0.07]	0.04 [0.00, 0.09]	0.06 [0.02, 0.10]	0.05 [0.00, 0.09]	–0.01 [–0.10, 0.07]
Mean % bias ^{**} [95% CI]	9% [2%, 16%]	5% [–5%, 16%]	8% [3%, 14%]	12% [6%, 18%]	9% [1%, 18%]	–2% [–18%, 14%]
SD [95% CI]	0.03 [0.02, 0.06]	0.05 [0.03, 0.11]	0.04 [0.03, 0.10]	0.06 [0.04, 0.14]	0.04 [0.03, 0.09]	0.09 [0.06, 0.21]
						0.02 [0.02, 0.05]

* bias defined as (CTA-derived value - true value).

** % bias defined as (CTA-derived value - true value)/true value.

*** as indicated in the DICOM tag

Table 3

(Top): Bias Profile for Wall Area across arteries and scanner settings. (Bottom): Stratified Analysis for Wall Area. The bins were established according to approximate consistency observed in result (the vertebral and the smallest carotid being the first bin, the other carotids in the middle bin, and the two femorals with the aorta in the third bin).

Pooled Across Arteries and Scanner Settings						
True Value (mm)	50.5–67.3	89.2–98.6	140–449	Overall		
# observations	160	16	24	56		
Mean bias * [95% CI]	8.90 [6.9, 10.9]	10.4 [5.7, 15.1]	32.89 [27.3, 38.5]	19.6 [15.5, 23.7]		
Mean % bias ** [95% CI]	14.8% [12.2%, 17.5%]	11.5% [6.2%, 16.8%]	14.8% [12.6%, 16.9%]	13.9% [12.0%, 15.7%]		
SD [95% CI]	3.73 [2.76, 5.77]	8.84 [6.53, 13.69]	13.19 [10.25, 18.51]	15.3 [12.9, 18.8]		
Pooled Across Arteries but Stratified By Scanner Settings						
KVP	100	100	100	120	120	120
mAs	156	325	325	156	606	606
CTDIvol (mGy) ***	4.95	10.3	10.3	8.12	16.9	16.88
Filter	B30f	B30f	soft B30f	B26f	B30f	B30f
Mean bias * [95% CI]	19.84 [−0.37, 40.05]	17.91 [6.64, 29.18]	19.15 [8.10, 30.21]	25.87 [9.21, 42.53]	20.21 [9.82, 30.61]	14.40 [1.51, 27.29]
Mean % bias ** [95% CI]	11% [6%, 17%]	13% [8%, 19%]	15% [9%, 20%]	19% [11%, 26%]	17% [10%, 24%]	10% [4%, 16%]
SD [95% CI]	21.85 [14.08, 48.13]	12.19 [7.85, 26.83]	11.95 [7.70, 26.32]	18.02 [11.61, 39.68]	11.24 [7.24, 24.75]	13.94 [8.98, 30.69]

* bias defined as (CTA-derived value - true value).

** % bias defined as (CTA-derived value - true value)/true value.

*** as indicated in the DICOM tag

Table 4

(Top): Bias Profile for Maximum Wall Thickness across arteries and scanner settings. (Bottom): Stratified Analysis for Maximum Wall Thickness. The bins were established according to approximate consistency observed in result (the vertebral and the smallest carotid being the first bin, the other carotids and the smallest femoral in the middle bin, and the larger femoral with the aorta in the third bin).

Pooled Across Arteries and Scanner Settings						
True Value (mm)	2.88–3.05	4.3–4.43	5.88–8.96	Overall		
# observations	16	24	16	56		
Mean bias [*] [95% CI]	1.07 [0.98, 1.15]	1.17 [1.06, 1.29]	1.398 [1.27, 1.52]	1.21 [1.14, 1.28]		
Mean % bias ^{**} [95% CI]	36.0% [–33.0%, –38.9%]	26.9% [24.4%, 29.5%]	19.9% [16.7%, 23.1%]	27.5% [25.3%, 30.0%]		
SD [95% CI]	0.173 [0.127, 0.267]	0.268 [0.208, 0.376]	0.233 [0.172, 0.360]	0.264 [0.223, 0.325]		
Pooled Across Arteries but Stratified By Scanner Settings						
KVP	100	100	100	120	120	120
mAs	156	156	325	156	606	606
CTDIvol (mGy) ^{***}	4.95	4.95	10.3	8.12	16.9	16.88
Filter	B30f	soft B30f	B30f	B26f	B30f	B30f
Mean bias [*] [95% CI]	1.20 [0.93, 1.48]	1.32 [1.12, 1.53]	1.21 [1.07, 1.35]	1.39 [1.13, 1.65]	1.16 [0.86, 1.45]	1.08 [0.89, 1.28]
Mean % bias ^{**} [95% CI]	27% [21%, 33%]	30% [22%, 38%]	28% [20%, 36%]	32% [22%, 41%]	25% [17%, 36%]	25% [17%, 33%]
SD [95% CI]	0.30 [0.19, 0.65]	0.22 [0.14, 0.48]	0.15 [0.10, 0.33]	0.28 [0.18, 0.62]	0.34 [0.22, 0.75]	0.21 [0.13, 0.46]

* bias defined as (CTA-derived value - true value).

** % bias defined as (CTA-derived value - true value)/true value.

*** as indicated in the DICOM tag