Translational Oncology
www.transonc.com

Volume 2 Number 4 December 2009 pp. 216–222

Computed Tomography Assessment of Response to Therapy: Tumor Volume Change Measurement, Truth Data, and Error[1]

Michael F. McNitt-Gray*, Luc M. Bidaut†, Samuel G. Armato III‡, Charles R. Meyer§, Marios A. Gavrielides¶, Charles Fenimore#, Geoffrey McLennan, Nicholas Petrick§, Binsheng Zhao††, Anthony P. Reeves‡‡, Reinhard Beichel§§, Hyun-Jung (Grace) Kim* and Lisa Kinnard§**

*Department of Radiology, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; †Department of Imaging Physics, Division of Diagnostic Imaging, UT-MD Anderson Cancer Center, Houston, TX, USA; ‡Department of Radiology, University of Chicago, Chicago, IL, USA; §Department of Radiology, University of Michigan, Ann Arbor, MI, USA; ¶Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD, USA; #National Institute of Standards and Technology, Gaithersburg, MD, USA; **Department of Internal Medicine, School of Medicine, University of Iowa, Iowa City, IA, USA; ††Department of Radiology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; ‡‡Biomedical Engineering, School of EECS, Cornell University, Ithaca, NY, USA; §§Department of Radiology, University of Iowa, Iowa City, IA, USA

Abstract

RATIONALE AND OBJECTIVES: This article describes issues and methods that are specific to the measurement of change in tumor volume as measured from computed tomographic (CT) images and how these would relate to the establishment of CT tumor volumetrics as a biomarker of patient response to therapy. The primary focus is on the measurement of lung tumors, but the approach should be generalizable to other anatomic regions. *MATERIALS AND METHODS:* The first issues addressed are the various sources of bias and variance in the measurement of tumor volumes, which are discussed in the context of measurement variation and its impact on the early detection of response to therapy. *RESULTS AND RESOURCES:* Research that seeks to identify the magnitude of some of these sources of error is ongoing, and several of these efforts are described herein. In addition, several resources for these investigations are being made available through the National Institutes of Health–funded Reference Image Database to Evaluate Response to therapy in cancer project, and these are described as well. Other measures derived from CT image data that might be predictive of patient response are described briefly, as well as the additional issues that each of these metrics may encounter in real-life applications. *CONCLUSIONS:* The article concludes with a brief discussion of moving from the assessment of measurement variation to the steps necessary to establish the efficacy of a metric as a biomarker for response.

Translational Oncology (2009) 2, 216–222

Address all correspondence to: Professor Michael F. McNitt-Gray, PhD, Department of Radiological Sciences, David Geffen School of Medicine at UCLA, Suite 650, 924 Westwood Blvd, Los Angeles, CA 90095-1721. E-mail: mmcnittgray@mednet.ucla.edu

[1] This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contracts N01-CO-12400 and HHSN261200800001E.
Received 16 August 2009; Revised 16 August 2009; Accepted 17 August 2009

DOI 10.1593/tlo.09226

Introduction

Lung cancer is the most common cause of cancer death in both men and women in the United States and has one of the lowest 5-year survival rates of all cancers (approximately 15%) [1]. Despite the many advances in imaging, genomics, and many other fields, this survival rate has not changed significantly in more than 20 years. While novel therapeutic agents are being introduced and evaluated for effectiveness [2], therapeutic response assessment methods have also not changed, despite significant advances in imaging technology.

Advances in computed tomography (CT) technology during the past 10 years have included increasing the number of detector rows (from typical single-detector row spiral systems in 1998 to the 64 to 320 detector row systems of 2009), decreased rotation times (down to <0.3 seconds for a full rotation), helical scanning and reconstruction techniques, and sophisticated radiation dose reduction methods, all of which have led, along with other developments, to exquisitely detailed descriptions of anatomy. The current multidetector row CT (MDCT) systems permit the acquisition of low-dose, thin-slice (<1.5 mm) image data sets of a patient's entire thorax in a single breath hold (typically now between 5 and 10 seconds). These very high resolution image data sets can also be used with advanced three-dimensional techniques (volume-rendering, maximum-intensity projection, surface-shaded display, etc) to visualize lung anatomy and pathology with unprecedented detail, using relatively noninvasive techniques. Therefore, MDCT has been used in the imaging of lung cancer, in studies investigating the early detection (e.g., screening, such as [3–6]), diagnosis (e.g., using temporal scans to detect differences over time such as [7,8]), and for the assessment of treatment response (see the review by Gavrielides et al. [9]).

Currently, the best hope for the treatment of lung cancer cure is surgical resection of a small peripheral lung nodule, which is possible in approximately 15% of patients presenting with early-stage disease (the other patients present with later-stage disease or are not surgical candidates owing to comorbidities). Those patients not able to receive surgery are often treated with chemotherapy, with or without associated external beam radiotherapy. Treatment protocols vary across the country and may include group protocol studies and clinical trials. New biologic response modifiers for lung cancer therapy have recently received increased interest. These generally less-toxic agents are designed to affect the tumor blood supply or other critical pathways in cancer cell growth, differentiation, or metastatic processes. The end point of such therapies may not be lung cancer "disappearance" but rather tumor growth cessation.

Despite advances both in CT technology and our understanding of lung cancer cellular and molecular mechanisms, the current standard method to measure tumor response to therapy using CT remains the Response Evaluation Criteria in Solid Tumors (RECIST), which is based on unidimensional, linear measurements of tumor diameter [10,11]. Because it is based on a series of linear measurements, RECIST offers a simple approach that requires minimal effort. The RECIST guidelines, however, presume that tumors are spherical and change in a uniform manner. Significant variability in the RECIST measure ments exists across different observers [12], and published work generally focuses on the surrogate of "best overall response," with only a few methods addressing other imaging end points such as "time to progression" and "disease-free survival." As a therapy response measurement procedure, RECIST maps linear data into an established set of four discrete categories: complete response, partial response, stable disease, and progressive disease. These categorical bins, however, are quite coarse, with most trial analyses critically pivoting on partial response (defined by a 30% linear sum reduction) and progressive disease (defined by a 20% increase in tumor dimension).

Because it uses only unidimensional linear measurements in its assessment, the RECIST criteria does not fully use the much higher-resolution data sets offered by CT or the advanced image segmentation and visualization methods that can be used on these data sets and that are, in fact, available on many commercial workstations. This limits the ability to accurately reflect size changes that occur in the many lesions that are not spherical in nature and may ultimately limit the ability to identify early changes in patients undergoing treatment through such an inherently flawed metric. The advances in CT technology described above had led to the development of three-dimensional methods to measure the volume of lung lesions, with the aim of developing more accurate and consistent measurements, even for nonspherical lesions, to ultimately better assess response over a shorter time interval.

This article investigates some of the issues in the use of high-resolution CT data sets for measuring tumor volumes, specifically in the problem of assessing the response to therapy involving tumors in the lung. Potential sources of both bias and variance are identified and discussed. Some investigations that are currently underway as part of the NCI funded Reference Image Database to Evaluate Response to therapy in cancer (RIDER) project [13] are described, as well as their contribution to the overall investigations into understanding bias and variance in this context. Finally, the relationship between measuring bias and variance and the establishment of CT tumor volumetrics as a biomarker are discussed.

This work was carried out as part of the RIDER project, and the intent was to develop a consensus approach to the benchmarking of software tools for the assessment of tumor response to therapy and to provide a publicly available database of images and associated metadata. The RIDER project evolved from the Lung Image Database Consortium (LIDC) project, which was funded by the National Cancer Institute to create a publicly available database of annotated thoracic CT scans as a reference standard for the development, training, and evaluation of computer-aided diagnostic methods for lung cancer detection and diagnosis [14–20]. The RIDER project is currently generating a database of temporally sequential CT, magnetic resonance imaging, and positron emission tomography scans of subjects with cancer, collected longitudinally during the course of nonsurgical cancer therapy. The database will also include phantom images of synthetic tumors and short-interval patient scans for the evaluation of the variance and bias inherent to change analysis software tools. The RIDER project was initiated in 2004 as a collaboration between the NCI's Cancer Imaging Program, the NCI's Center for Bioinformatics, the National Institute of Biomedical Imaging and Bioengineering (NIBIB), and the Prevent Cancer Foundation (formerly the Cancer Research and Prevention Foundation), with information technology support from the Radiological Society of North America. The RIDER project is composed of academic researchers, program staff at NCI, and members of the Cancer Biomedical Informatics Grid, National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, and the National Institute of Standards and Technology.

Materials and Methods

Measurement Issues Specific to CT Tumor Volumetrics

While CT images have been used as the basis for treatment response assessment using the RECIST criteria, many questions must be answered before CT tumor volumetrics can provide an accurate and precise estimate of a patient's response to therapy. As with the companion articles, these issues will first be addressed with regards to measuring bias and variance, specifically with regard to tumor volume as estimated through CT. Whereas some of these problems are somewhat straightforward to address, others are not.

Bias. Bias is typically assessed by comparing one set of measurements of a given object with another set of measurements made using some external truth standard. For example, the volume of an object on CT can be measured using the image data (e.g., counting the number of voxels that are identified as being part of the object after the application of a three-dimensional segmentation method and then multiplying the number of voxels by the physical size of each voxel). This measured value can then be compared with the estimated "true volume" of the object using some external standard such as a water displacement method.

Although this can be relatively straightforward for clearly identified and measurable test objects or phantoms, this is not at all straightforward for *in vivo* tissues. Specifically for tumors, as long as they remain *in vivo*, there is no method to reliably establish the "true volume" of the tumor with an external standard. Figure 1 shows an example of a tumor contoured by three different readers in an effort to obtain expert delineations of the "true" tumor boundary (as part of the experimental work described in Meyer et al. [20]). It should be noted that the readers were provided the same drawing instructions. As that article demonstrated, there is significant variability between readers in many different kinds of lesions. Even if the tumor is excised, there may be some distortion of the tissue (from incomplete resection or resection of additional tissue connected to the tumor, or even compression of the tissue) or just simple alteration of its physiological environment (blood loss, etc) that then prevents an accurate estimation of the volume of the tumor when it was in its *in vivo* state.

Therefore, bias is often addressed using some simulated tumors or phantom objects. These objects (described in more detail below) are meant to serve as approximations for actual tumors by providing similar x-ray attenuation properties and shapes. They also provide the significant advantage that they can be measured (repeatedly if necessary) using an external truth standard (e.g., volume displacement or other method).

Variance. Variance is typically assessed by using repeated measurements under the same measurement conditions. As described elsewhere [9], there are many potential sources of variance. In the specific problem of lung lesion measurement using CT image data, sources of variance are introduced at many different stages of the measurement. Because CT is an ionizing x-ray–based modality, repeated CT examinations of the same patient are not generally performed owing to dose

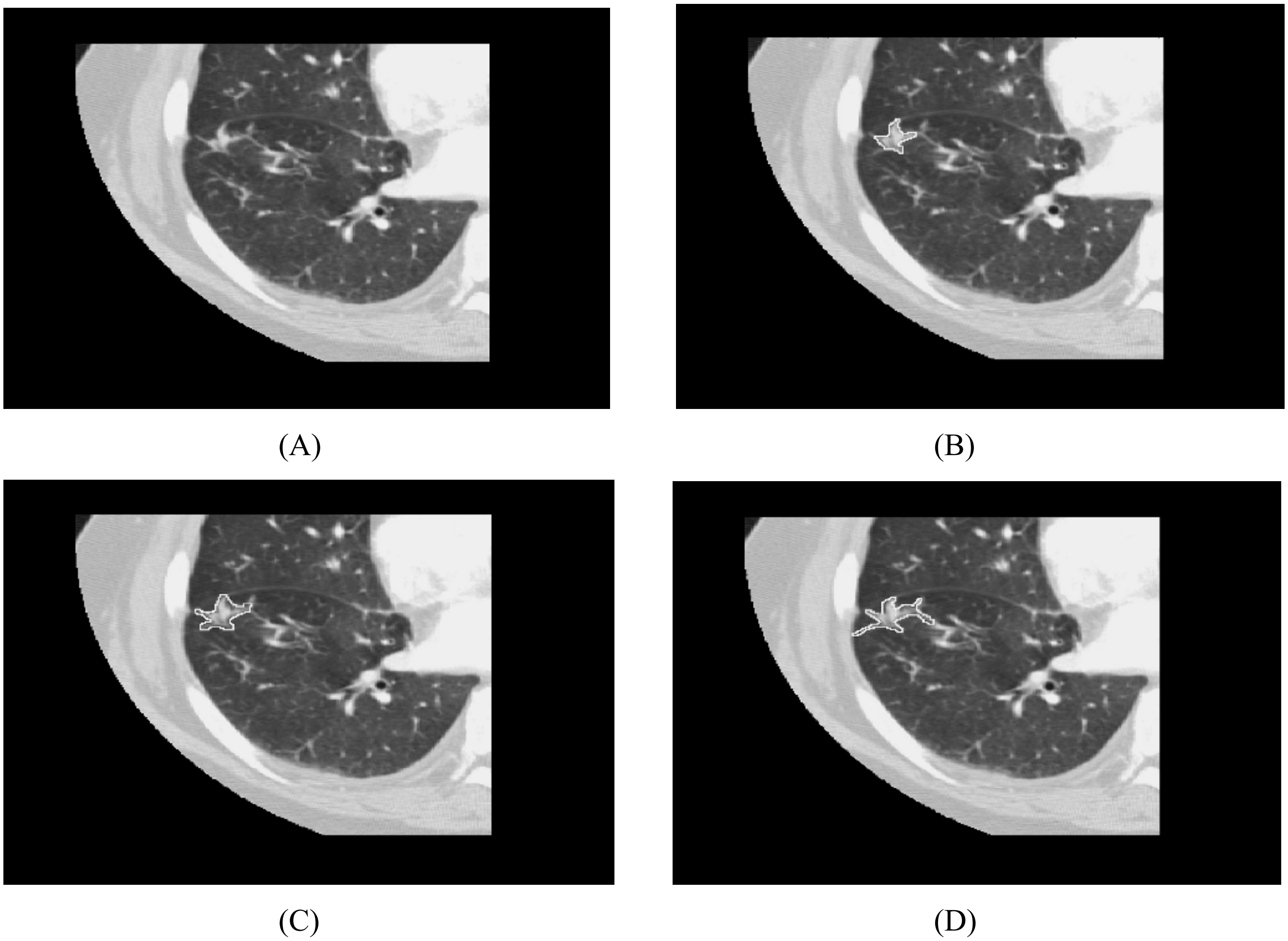

Figure 1. (A) Original image of complex lung nodule from CT. (B, C and D) White outline shows contour from three different readers. All readers were given the same drawing instructions (based on experimental work described in Meyer et al. [20]).

considerations. However, one study of repeat examinations was performed at the Memorial Sloan Kettering Cancer Center, which is described below. Whereas image acquisition and reconstruction is one source of variance, several other sources of variance (described later on) may be assessed through repeated measurements that will include the effect of the reader and of the measurement tools (e.g., algorithms) being used, as well as possible "learning" interactions between readers and cases. Therefore, variance may be investigated to some extent from phantom studies, but certain sources of variance may also be addressed by using patient image data sets.

Because one very important source of variance between different studies can be the patient image database that is used for measurements, one of the major contributions of the RIDER project is the collection of *reference* image data sets that are publicly available and therefore allow *direct* comparison of results from different measurement approaches on the exact same data sets. The purpose of these reference data sets is to allow investigations into the relative magnitudes of different sources of variance and to develop methods and best practices that will reduce variance from these various sources. Again, the ultimate goal is to improve the assessment of the response to therapy in as universal a context as achievable in practice.

Potential sources of bias and variance in CT tumor volumetrics within the lung. This section identifies and discusses some potential sources of bias and variance that are specific to the problem of measuring tumor volumes within the lung using CT. These sources of variability have been broken down into a few categories: 1) scanner-related, 2) patient-related, 3) tumor-related, 4) measurement method–related, and 5) interactions between sources of variance.

CT scanner–related factors. These describe factors that can vary within and between CT scanners and that may introduce some bias and variance into the measurement of lung volumes. In a multicenter clinical trial, investigators rarely have any choice about what scanner is to be used to image patients in the trial; similarly, in clinical practice, repeat scans of patients are often acquired with different scanners owing to scanner availability and scheduling limits. The physical hardware–based differences between scanners can be a potential source of bias and variance. Specifically, the manufacturer and model of the CT scanner as well as the number of detector rows, differences in detector material, the x-ray spectrum, and filtration materials (e.g., bow tie filter composition), as well as different available options for image reconstruction may all affect the basic image acquisition abilities and physical properties of the scanner. For example, when a scanner with only four detector rows or less is used, it can be difficult to obtain thin slices (≤1.5 mm) through the entire thorax within a single breath hold, whereas this can be routinely performed with a 64–detector row scanner.

In addition to the basic properties of the scanner itself are the differences in imaging protocol, or how the scanner is actually being used. This includes technical parameters settings such as kilovolt peak (kV[p]), milliampere-second (mA s), beam collimation, and helical pitch (when helical scanning is performed). Note that similar nominal settings on different scanners does not guarantee identical physical properties of the reconstructed images. For example, using a setting of 120 kV(p) on MDCTs from different manufacturers may not result in the same spectrum of x-ray beam energies to the patient, primarily because of differences in x-ray beam filtration, etc In addition to image acquisition parameters, the reconstructed image parameters may also play a significant role, such as reconstructed slice thickness (e.g., thin slice or thicker slices), spacing or interval between reconstructed images (e.g., creating overlap or not), reconstruction kernel (which has a significant impact on noise and resolution properties) and pixel size/ reconstructed field of view. Any of these may have a significant impact on the final image resolution (in the *x-y* plane, or *z* plane), noise, and contrast resolution. Moreover, these parameters interact with other sources of variability, such as nodule size, as will be described in the remainder of this article.

The imaging protocol may also include fundamental decisions about whether the acquisition is performed in a single helical pass through the chest or as multiple overlapped passes over various portions of the patient's body. For example, one common imaging protocol is to scan through the entire chest down to the lung bases and slightly into the abdomen, then have a slight delay (for iodinated contrast to arrive in the abdominal organs), move superiorly slightly and then scan from the top of the diaphragm through the abdomen, thereby creating a region of scanning overlap. Another common imaging protocol is to eliminate that overlap region by performing the first phase of the scan through the thorax, stopping at the top of the diaphragm, having the contrast delay, and then scanning—without moving back—from the top of the diaphragm down through the abdomen. One can also envision in the future that larger numbers of detector rows or flat-panel CTs will allow for image acquisition in a single-axial cone-beam CT acquisition.

The imaging protocol may also have some specific description of the use of intravenous (IV) or oral contrast, whereas some studies may be performed without contrast. In particular, when IV contrast is used, images will be affected by the injection rate and volume of contrast used (and whether that volume used is specified based on patient weight or some other method), the type and concentration of contrast used (there are a variety of concentrations commercially available) as well as the delay used (how long after the injection the scanning should be performed to allow the contrast to arrive at structures of interest) may all affect the characteristics and appearance of the images before subsequent analysis.

Finally, the scanner's calibration and maintenance can also be placed under the scanner-related category. Some studies may require that a quality assurance phantom be scanned at some interval (or even alongside each patient). The study quality assurance process may involve some physical measurements to determine the scanner's calibration to reference materials such as water and air. There may also be an assessment of the Hounsfield unit (HU) scale's consistency both at different points in time and even across the field of view of the reconstructed image, as well as of the scanner's contrast scale (how do CT numbers change with different reference materials). The scanner's maintenance schedule and software upgrades may affect whether parts like the x-ray tube have been changed or whether a new calibration, reconstruction, or correction algorithm is being used. Any of these items may have some effect on the scanners' intrinsic image quality in properties such as noise or resolution inherent contrast and may therefore affect the estimation of tumor volume.

Patient-, tumor-, and analysis-related factors. In addition to factors related to the CT scanner itself, there are many factors that relate to the patient and analysis method that may ultimately influence the measurement of tumor volumes. These are described in more details in [21] that describes non–modality-dependent factors.

Results and Resources

RIDER and RIDER-Related Investigations into Source of Error for CT Volumetrics

As stated above, the purpose of the RIDER project is to provide image data (and other relevant data if possible) to allow investigations into sources of error and methods to reduce them. This section will describe several data sources that are provided through RIDER and describe how they may be used to understand bias and variance in tumor volume measurements.

Patient Repeat CTs ("Coffee Break Experiment") Performed at Memorial Sloan Kettering Cancer Center

As part of a study evaluating variability in tumor measurements from same-day repeat CT scans of patients with non–small cell lung cancer [22], 32 patients were imaged twice on the same scanner at an interval less than 15 minutes apart, with the patient getting off the scanner bed between scans. This unique experiment represents repeat scans under a presumed "no change" condition. The imaging was done on one of two CT scanners (LightSpeed 16 or VCT; GE Medical Systems, Waukesha, WI), with 16 or 64 detector rows, respectively. The LightSpeed 16 was used for 28 of the subjects, and 4 of them were scanned on the VCT. Identical technical parameter settings were used for both baseline and repeat scans. The following settings were used:

i. 120 kV(p)
ii. detector configuration of 16 × 1.25 mm (64 × 0.625 mm for the VCT)
iii. pitch of 1.375:1 (0.984 for the VCT)
iv. rotation time of 0.5 second.
v. The standard-dose thoracic images were obtained without IV contrast during a single breath hold.
vi. Thin-section images of 1.25 mm were reconstructed using the lung convolution kernel.
vii. Patients were given the same breathing instructions for both studies.

As with other RIDER data sets, the repeat scans data sets are being made available to the public through the National Biomedical Imaging Archive (NBIA) Web archive (http://ncia.nci.nih.gov/). It should be noted that no annotations of the images have been provided (i.e., no reader markings or measurements).

Because these data sets were acquired during such a short time interval, they represent a nearly ideal "no change" condition during this period. However, the image data sets are not expected to be identical because of issues described above (and in the Appendix of the companion article by Meyer et al.), including differences in patient positioning, patient inspiration level, etc However, because these data sets do represent the same lesions under identical parameter settings, they should ideally translate into null bias and variance across repeat measurements and can therefore be used in investigations about minimum detectable change.

Unmarked CT Lung Studies at Different Time Intervals from UT-MD Anderson Cancer Center

In this study, many cases of patients with known tumors in the lungs (both primary and metastatic lesions are included) were submitted to NBIA under the RIDER collection. Each case had at least two image data sets from different time points; many had three or more time points. These cases were collected all from the same institution, as part of their clinical operation. Therefore, neither was there a specific attempt either to collect the data prospectively nor was their any specific attempt to standardize the imaging protocol beyond what is already done as part of that site's clinical operation. Therefore, although scans were typically performed with the same acquisition protocols on similar scanners for each exam, there is no guarantee that this was the case because this was not one of the inclusion criteria (however, it should be noted that technical parameter information is available through DICOM headers of these cases so this could be evaluated by users of the database).

Whereas neither this site nor anyone else has provided annotations of the complete tumor boundaries, as part of the original RIDER project, two radiologists (C. Jaffe and R. Gottlieb) did provide RECIST measurements on approximately 30 of the cases. These annotations are available through NBIA as well.

CT Lung Nodule Phantom from FDA Center for Devices and Radiological Health

As part of a parallel effort, colleagues at the FDA's Center for Devices and Radiological Health, Office of Science and Engineering Laboratories, Division of Imaging and Applied Mathematics have acquired and scanned an anthropomorphic phantom (Multipurpose Chest Phantom; Kyoto Kagaku Co Ltd, Kyoto, Japan). In this study, several synthetic nodules of known size, composition, and shape were placed inside the anthropomorphic phantom with and without attachment to simulated lung vasculature. This phantom has been scanned multiple times under different combinations of dose (i.e., low and high mA s), slice thickness (thin and thick slices), pitch, and reconstruction filter (smooth and sharp kernels). This provides a series of scans through this phantom with different combinations of technical parameter settings and therefore different combinations of noise and spatial resolution. These data sets are being made available through the NBIA Web site to allow various segmentation and volume calculation software techniques to be used to estimate the volume of each lesion. In addition, this phantom has also been scanned at several different institutions and similar imaging protocols have been performed using scanners from different manufacturers and different numbers of detector rows (8-, 16-, and 64-slice scanners) to investigate the variability between scans obtained on different platforms.

As described above, one of the main advantages of using phantoms is the ability to investigate bias as a difference from some externally assessed "truth" standard. Because there is no concern over radiation exposure, the phantom can be scanned repeatedly under different acquisition conditions, that is, with different technical parameter settings for a given scanner or across various models. Analysis of this data will help in determining the effects on both bias and variance of volume measurements under different technical parameter settings. In addition, because known simulated lesions having different size, shape, and composition are used, analyses can be performed to investigate the interactions between technical parameters and a nodule's size, shape, or composition, for example, as a way to assess their relative and combined impact on the final estimation of volume.

As it now stands, the database includes CT scans of the anthropomorphic chest phantom containing a vasculature insert to which synthetic nodules have been attached. Synthetic nodules in the set vary in HU density values (−630 HU, −10 HU, and +100 HU), size (5, 8, 10, 20, and 40 mm), and shape (spherical, elliptical, lobulated, and spiculated). Imaging protocols included variable exposure (20, 100, and 200 mA s), pitch (0.9 and 1.2), reconstruction kernel (detail and medium), slice collimation (16 × 0.75 and 16 × 1.5 mm), and slice thickness (0.8, 1.5, and 3.0 for the 16 × 0.75-mm collimation, and 2.0, 3.0, and 5.0 for the 16 × 1.5-mm collimation). Ten repeat scans for each protocol were acquired using a 16-row multidetector CT scanner (IDT Mx8000; Philips Medical Systems, Cleveland, OH). Scans were also collected on a 64-row multidetector CT scanner (Sensation 64; Siemens Medical Solutions, Forcheim, Germany) for a subset of the protocols described above. This database of CT scans is available to the public through the NBIA, which is located at https://imaging.nci.nih.gov/ncia/. Nonspherical nodular data sets will be released as they become available.

All human subject studies were approved by the respective institutional review board, and written informed consent was provided by each subject.

Discussion

Mitigation Measures

Although it will take some time to complete the above data acquisition and analyses to provide specific advice on bias and variance in the CT tumor volumetrics problem, in the short term, some advice may still be provided about mitigation measures that should reduce the previously detailed sources of error. An overarching advice is to keep as many potential sources of variance as possible exactly the same across interval examinations, including the following:

1) Using the same device (same scanner)
2) Using the same technical parameter settings as previous examinations for any given patient (although some thought about adjusting parameter settings owing to change in patient size, such as weight gain or weight loss, may also be appropriate).
3) Using the same patient factors (positioning, breathing instructions, etc)
4) Performing analysis with the same software (e.g., manufacturer-dependent)
5) Performing analysis with the same software settings (threshold, etc) as previous examinations.
6) Performing analysis with the same software version if updates were to add another source of variance (i.e., change in absolute measurements) without improving the change analysis (e.g., when seeking relative rather than absolute changes)
7) On the basis of results from multiple studies [23–26], that subcentimeter nodules should be measured using thin-slice (≤1 mm) imaging protocol.

Conclusions

The focus of this article has been the measurement of tumor volumes in the lung using CT and the various sources of error one can encounter in that context. Different sources of error have been described as well as investigations currently underway to address many of these issues through an expanded understanding of bias and variance. Data sets that either are or will soon be publicly available are described as well as their potential use in such investigations.

One limitation of this work is that, although it is very important to understand sources of bias and variance and to be able to measure their effects on the variable of interest (i.e., tumor volumes), this work is not sufficient in itself to establish CT tumor volumetrics as a biomarker. That is, these investigations may provide evidence for reduced variability when using volume rather than when using a linear measurement based metric such as RECIST. However, this still does not indicate whether tumor volumetrics is a good predictor of patient response to therapy. To establish that would require further information not only on tumor volumes but also on the actual overall response to therapy, which may be measured using patient outcome data or metrics such as progression-free survival. Again, while the characterization work introduced in this article is essential, it is clearly not sufficient for the establishment of a biomarker. Further investigation into alternate content-based metrics (e.g., density or even perfusion) may be required, especially in light of new therapies that do not seek to shrink tumors, but to cut off their blood supply and therefore to render the tumor biologically inactive.

While the primary focus of this article is on activities surrounding the RIDER CT tumor volumetrics effort, there are many other analyses and experiments that could be investigated with the collected data sets. These include other metrics of change beyond just volume, which could include changes in tumor density and composition. Such analyses could readily be carried out with the data sets described in this article. Other parameters, such as functional descriptions of perfusion or permeability, would require data sets acquired under different image acquisition protocols that would involve dynamic imaging of the lesions. Although these new data sets could further exploit the RIDER framework, they would also require a careful analysis of the sources of variance and covariates in a similar fashion as explored herein with volumetric analysis of tumors on CT.

Acknowledgments

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

References

[1] Jemal A, Siegel R, Ward E, Hao Y, Xu J, and Thun MJ (2008). Cancer statistics, 2008. *CA Cancer J Clin* **58**, 71–96.

[2] Ramalingam S and Belani C (2008). Systemic chemotherapy for advanced non–small cell lung cancer: recent advances and future directions. *Oncologist* **13**, 5–13.

[3] Henschke CI, Naidich DP, Yankelevitz DF, McGuinness G, McCauley DI, Smith JP, Libby D, Pasmantier M, Vazquez M, Koizumi J, et al. (2001). Early Lung Cancer Action Project: Initial findings on repeat screening. *Cancer* **92**, 153–159.

[4] Swensen SJ, Jett JR, Hartman TE, Midthun DE, Sloan JA, Sykes AM, Aughenbaugh GL, and Clemens MA (2003). Lung cancer screening with CT: Mayo Clinic experience. *Radiology* **226**, 756–761.

[5] Sone S, Li F, Yang ZG, Honda T, Maruyama Y, Takashima S, Hasegawa M, Kawakami S, Kubo K, Haniuda M, et al. (2001). Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner. *Br J Cancer* **84**, 25–32.

[6] Hillman BJ (2003). Economic, legal, and ethical rationales for the ACRIN national lung screening trial of CT screening for lung cancer. *Acad Radiol* **10**, 349–350.

[7] Kostis WJ, Reeves AP, Yankelevitz DF, and Henschke CI (2003). Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. *IEEE Trans Med Imaging* **22** (10), 1259–1274.

[8] Yankelevitz DF, Reeves AP, Kostis WJ, Zhao B, and Henschke CI (2000). Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. *Radiology* **217** (1), 251–256.

[9] Gavrielides MA, Kinnard LM, Myers KJ, and Petrick N (2009). Noncalcified lung nodules: volumetric assessment with thoracic CT. *Radiology* **251** (1), 26–37.

[10] James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Muldal A, and Therasse P (1999). Measuring response in solid tumors: unidimensional *versus* bidimensional measurement. *J Natl Cancer Inst* **91**, 523–528.

[11] Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, van Oosterom AT, Christian MC, et al. (2000). New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst* **92**, 205–216.

[12] Schwartz LH, Mazumdar M, Brown W, Smith A, and Panicek DM (2003). Variability in response assessment in solid tumors: effect of number of lesions chosen for measurement. *Clin Cancer Res* **9**, 4318–4323.

[13] Armato SG III, Meyer CR, Mcnitt-Gray MF, McLennan G, Reeves AP, Croft BY, Clarke LP, and RIDER Research Group (2008). The Reference Image Database to Evaluate Response to therapy in lung cancer (RIDER) project: a resource for the development of change-analysis software. *Clin Pharmacol Ther* **84** (4), 448–456.

[14] Clarke LP, Croft BY, Staab E, Baker H, and Sullivan DC (2001). National Cancer Institute initiative: lung image database resource for imaging research. *Acad Radiol* **8**, 447–450.

[15] Armato SG III, McLennan G, McNitt-Gray MF, Meyer CR, Yankelevitz D, Aberle DR, Henschke CI, Hoffman EA, Kazerooni EA, MacMahon H, et al. (2004). Lung Image Database Consortium: developing a resource for the medical imaging research community. *Radiology* **232**, 739–748.

[16] McNitt-Gray MF, Armato SG III, Meyer CR, Reeves AP, McLennan G, Pais RC, Freymann J, Brown MS, Engelmann RM, Bland PH, et al. (2007). The Lung Image Database Consortium (LIDC) data collection process for nodule detection and annotation. *Acad Radiol* **14**, 1464–1474.

[17] Armato SG III, McNitt-Gray MF, Reeves AP, Meyer CR, McLennan G, Aberle DR, Kazerooni EA, MacMahon H, van Beek EJ, Yankelevitz D, et al. (2007). The Lung Image Database Consortium (LIDC): an evaluation of radiologist variability in the identification of lung nodules on CT scans. *Acad Radiol* **14**, 1409–1421.

[18] Armato SG III, Roberts RY, Kocherginsky M, Aberle DR, Kazerooni EA, Macmahon H, van Beek EJ, Yankelevitz D, McLennan G, McNitt-Gray MF, et al. (2009). Assessment of radiologist performance in the detection of lung nodules: dependence on the definition of "truth". *Acad Radiol* **16** (1), 28–38.

[19] Reeves AP, Biancardi AM, Apanasovich TV, Meyer CR, MacMahon H, van Beek EJ, Kazerooni EA, Yankelevitz D, McNitt-Gray MF, McLennan G, et al. (2007). The Lung Image Database Consortium (LIDC): a comparison of different size metrics for pulmonary nodule measurements. *Acad Radiol* **14**, 1475–1485.

[20] Meyer CR, Johnson TD, McLennan G, Aberle DR, Kazerooni EA, Macmahon H, Mullan BF, Yankelevitz DF, van Beek EJ, Armato SG III, et al. (2006). Evaluation of lung MDCT nodule annotation across radiologists and methods. *Acad Radiol* **13**, 1254–1265.

[21] Meyer CR, Armato SG III, Fenimore CP, McLennan G, Bidaut LM, Barboriak DP, Gavrielides MA, Jackson EF, McNitt-Gray MF, Kinahan PE, et al. (2009). Quantitative imaging to assess tumor response to therapy: common themes of measurement, truth data, and error sources. *Transl Oncol* **2** (4), 198–210.

[22] Zhao B, James LP, Moskowitz CS, Guo P, Ginsberg MS, Lefkowitz RA, Qin Y, Riely GJ, Kris MG, and Schwartz LH (2009). Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non–small cell lung cancer. *Radiology* **252** (1), 263–272.

[23] Winer-Muram HT, Jennings SG, Meyer CA, Liang Y, Aisen AM, Tarver RD, and McGarry RC (2003). Effect of varying CT section width on volumetric measurement of lung tumors and application of compensatory equations. *Radiology* **229**, 184–194.

[24] Petrou M, Quint LE, Nan B, and Baker LH (2007). Pulmonary nodule volumetric measurement variability as a function of CT slice thickness and nodule morphology. *AJR Am J Roentgenol* **188**, 306–312.

[25] Kuhnigk JM, Dicken V, Bornemann L, Bakai A, Wormanns D, Krass S, and Peitgen HO (2006). Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. *IEEE Trans Med Imaging* **25**, 417–434.

[26] Zhao B, Schwartz LH, Moskowitz CS, Wang L, Ginsberg MS, Cooper CA, Jiang L, and Kalaigian JP (2005). Pulmonary metastases: effect of CT section thickness on measurement—initial experience. *Radiology* **234**, 934–939.