Observer Variability in Mesothelioma Tumor Thickness Measurements

Defining Minimally Measurable Lesions

Samuel G. Armato III, PhD, Anna K. Nowak, MD, PhD,†‡ Roslyn J. Francis, MD, PhD,†§ Masha Kocherginsky, PhD,‖ and Michael J. Byrne, MD†‡*

Introduction: Single time-point unidimensional tumor thickness measurements define measurable disease for clinical trial inclusion and also constitute a field in the International Association for the Study of Lung Cancer prospective mesothelioma staging database. The modified Response Evaluation Criteria in Solid Tumors (RECIST) guidelines for mesothelioma did not alter the 10-mm minimum tumor measurement recommendation. However, as computed tomography technology has advanced, we sought to examine whether interobserver agreement was acceptable at smaller tumor thickness in mesothelioma.

Methods: The primary observer selected 170 discrete measurement sites from 105 thoracic computed tomography scans from 50 consenting patients with pleural mesothelioma. Sites represented a range of tumor thickness, lesion morphology, and location. The outer (parietal) tumor margin was marked at each site and presented to five additional observers, who then selected the visceral margin of the tumor to create a line segment that captured tumor thickness. Relative differences among the observer measurements were estimated using a random-effects analysis of variance model to identify the smallest tumor thickness at which linear measurements could be made reliably.

Results: Systematic bias was observed, with some observers consistently measuring larger or smaller thicknesses than the thickness measured by others. The mean range across all 170 sites was 15.1% of the mean per-site measurement (SD = 9.1%; median range, 13.1%). Measurements acquired at sites with mean tumor thickness less than 7.5 mm demonstrated interobserver variability with a 75th percentile included 20% of the tumor thickness. The 95% confidence interval for relative interobserver measurement differences obtained for mean measurement lengths in the range 5 to 7.5 mm does not exceed the RECIST thresholds.

Conclusions: This study has implications for the definition of minimally measurable tumor adopted by clinical trial and staging protocols. Although the statistical findings suggest that a reduction in "minimally measurable disease" from 10 mm to 5 or 7.5 mm might be warranted, clinical factors may ultimately dictate the most appropriate definition.

Key Words: Malignant pleural mesothelioma, Response assessment, Staging, Thoracic computed tomography, Interobserver variability.

(*J Thorac Oncol.* 2014;9: 1187–1194)

Linear measurement of tumor diameter on computed tomography (CT) scans remains the standard clinical metric for the evaluation of tumor growth or response to therapy. The Response Evaluation Criteria in Solid Tumors (RECIST) guidelines[1] specify a tumor measurement approach (a single unidimensional measurement of the tumor's longest diameter) and a set of thresholds to convert numeric change in tumor measurements across temporally sequential CT scans into categories of tumor response (complete response, partial response, stable disease, and progressive disease). The modified RECIST guidelines[2] changed the tumor measurement approach, specifically for mesothelioma, from longest tumor diameter to tumor thickness perpendicular to the chest wall (or mediastinum) to accommodate the unique morphology of this disease.

Also contained within the RECIST guidelines is the specification of "measurable disease" as a tumor with a minimum diameter of 10 mm, which, for geometric and CT partial-volume–effect considerations, is a threshold that represents twice the then-state-of-the-art 5-mm thickness of CT section images. Modified RECIST did not change this threshold, which has not been challenged in the intervening years, even as CT technology has evolved. RECIST was conceptualized under assumptions of spherical tumor morphology. A 10-mm-diameter ("just-measurable") spherical tumor has a volume (i.e., tumor burden) of 523 mm^3; however, one possible morphological representation of mesothelioma tumor

Departments of *Radiology and ‖Health Studies, The University of Chicago, Chicago, Illinois; †School of Medicine and Pharmacology, University of Western Australia, Crawley, Western Australia; and Departments of ‡Medical Oncology and §Nuclear Medicine, Sir Charles Gairdner Hospital, Perth, Western Australia.

Disclosure: Dr. Armato received royalties and licensing fees from The University of Chicago related to computer-aided diagnosis. Dr. Nowak received research funding from Pfizer Australia and Boehringer Ingelheim Australia and has Advisory Board Membership for Roche Australia, Boehringer Ingelheim Australia, Verastem USA, and Roche International. The other authors declare no conflict of interest.

Address for correspondence: Samuel G. Armato III, PhD, Department of Radiology, MC 2026, The University of Chicago, 5841 S. Maryland Avenue, Chicago, IL 60637. E-mail: s-armato@uchicago.edu

ISSN: 1556-0864/14/0908-1187

with a "just-measurable" 10-mm in-plane thickness on a single 5-mm CT section encompasses a volume of 7672 mm^3 (the equivalent of a 24.5-mm-diameter spherical tumor)[3] (Fig. 1). Given that the anatomical extent of mesothelioma is rarely (if ever) constrained to a single CT section, the actual volume of a tumor with 10-mm in-plane thickness will likely be much greater than the equivalent of a 24.5-mm-diameter spherical tumor. Consequently, clinical trials that require "measurable disease" under RECIST as a criterion for enrollment may disadvantage subjects and the success of the trial through a greater baseline tumor burden. Following the rationale of RECIST that "measurable disease" should be defined as at least twice the thickness of CT section images, advances in CT technology may justify a revised definition, because state-of-the-art scanners are capable of section thicknesses less than 1 mm, and section thicknesses less than 3 mm have become more common.

Another factor that should be considered when defining "measurable disease," however, is observer measurement variability, a concept alluded to in the RECIST guidelines.[1,4] Measurements, to be a reliable quantitative tumor assessment metric on which patient management decisions are made and clinical trial efficacy is evaluated, must demonstrate an acceptable level of variability across the observers who acquire those measurements. The increase in measurement variability with decreased size of the object being measured is a well-known trend,[5] which lends credibility to the notion that some minimum tumor size should be defined below which inherent measurement variability would limit the practical utility of the acquired measurements. Although variability in mesothelioma tumor thickness measurements has been reported previously,[6] the impact of physical tumor characteristics on measurement variability has not been investigated.

The purpose of this study was to determine the dependence of mesothelioma tumor thickness measurement variability on tumor thickness, lesion morphology, and anatomical location, with the aim of informing a mesothelioma-specific definition of "measurable disease" and optimal measurement site selection.

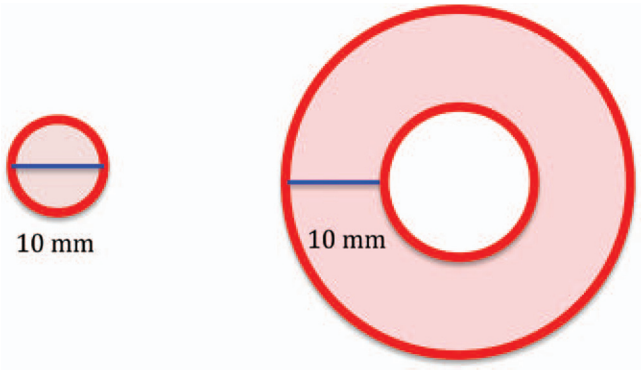

FIGURE 1. A two-dimensional example demonstrates that a 10-mm-diameter circle (representing a spherical tumor) has a much smaller area (volume) than a 10-mm-thick annulus (representing a rind of mesothelioma tumor).

MATERIALS AND METHODS

A retrospective database of 105 thoracic CT scans from 50 patients with mesothelioma was collected from Sir Charles Gairdner Hospital in Perth, Western Australia. Images were intentionally selected from heterogeneous time points throughout the disease course to obtain a range of tumor thickness and location. Scans had been performed on a GE Medical Systems (Milwaukee, WI) Hi Speed (n = 72), GE LightSpeed (n = 16), or Philips (Highland Heights, OH) Brilliance 64 (n = 17) CT scanner. Peak voltage was 120 kVp for all scans, pixel size ranged from 0.57 to 0.91 mm, and section thickness was 0.625 mm (n = 2), 1 mm (n = 1), 1.25 mm (n = 2), 2.5 mm (n = 1), 5 mm (n = 96), 7 mm (n = 2), or 10 mm (n = 1). All images had been reconstructed as 512 × 512-pixel images.

With approval from the local Human Research Ethics Committee, each scan was reviewed by a medical oncologist (AKN) (the "primary observer"), who used an in-house image visualization and manipulation software package (Abras, version 1.6) to identify 170 sites of mesothelioma tumor that represented a range of thicknesses, lesion morphologies, and anatomical locations across all scans. Through Abras, the primary observer identified a specific outer tumor margin point along the chest wall or mediastinal structures at each measurement site and created a line segment that spanned the tumor from the outer tumor margin point to an appropriate location along the inner tumor margin, in accordance with the modified RECIST tumor measurement approach.[2] The primary observer then categorized local tumor morphology (concave rind, convex rind, convex mass, or fusiform mass) (Fig. 2) and anatomical location (chest wall, mediastinum, anterior angle, or posterior angle; upper, middle, or lower zone of the thorax in the craniocaudal direction according to uniformly specified boundaries; outer tumor margin point along bone or soft tissue; and laterality). It is important to note that these 170 measurement sites were not selected to capture foci of clinical relevance but rather to represent a range of tumor thicknesses and morphologies, with anatomical location a secondary consideration.

An observer study was conducted in which Abras was used to present each of five other physicians with the specific CT section and the same preselected fixed location of the outer tumor margin point at each of the 170 primary-observer–defined tumor measurement sites. Each observer independently used Abras to create at each measurement site a line segment that spanned the tumor from the annotated predefined outer margin point to an appropriate location along the inner tumor margin (Fig. 3); the length of each observer's line segment became the tumor thickness measurement for that observer. This process was exactly the same as for the primary observer, except that the outer tumor margin point identified and recorded by the primary observer became the common fixed starting point for the measurements of the other observers; no data regarding lesion morphology or anatomical location were captured from these other observers.

Interobserver measurement variability was calculated as a function of mean tumor thickness measurement, lesion morphology, and anatomical location to identify the smallest