European Radiology (2019) 29:2981–2988
https://doi.org/10.1007/s00330-018-5887-z

CHEST

Correlation of patient survival with clinical tumor measurements in malignant pleural mesothelioma

Feng Li[1] · Mehwish Ahmad[2] · Fawwaz Qayyum[1] · Christopher M. Straus[1] · Heber MacMahon[1] · Hedy Kindler[2] · Samuel G. Armato III[1]

Received: 21 March 2018 / Revised: 22 October 2018 / Accepted: 13 November 2018 / Published online: 7 January 2019
© European Society of Radiology 2019

Abstract

Objectives To evaluate differences in the tumor response classifications that result from clinical measurements and to compare these response classifications with overall survival for patients with malignant pleural mesothelioma (MPM).

Methods One hundred thirty-one computed tomography (CT) scans were collected from 41 MPM patients enrolled in a clinical trial. Primary measurements had been acquired by clinical radiologists at a single center during routine clinical workflow, and the variability of these measurements was investigated. Retrospective measurements were acquired by a single radiologist in compliance with the study protocol based on the modified response evaluation criteria in solid tumors (RECIST). Differences in response classification categories by the two measurement approaches were evaluated and compared with patient survival.

Results Eleven (27%) of the 41 MPM patients had primary measurements at baseline or at follow-up that deviated from the guidelines of the clinical trial protocol. Among the 41 baseline scans, no statistical difference was observed in summed tumor measurements between primary and retrospective measurements. Response classification based on primary and retrospective measurements was different in 23 (26%) of the 90 follow-up scans, and best response was the different in seven (17%) of the 41 patients. Using Harrell's *C* statistic as a measure of correlation, response based on retrospective measurements correlated better with survival ($C = 0.62$) than did response based on primary measurements ($C = 0.57$).

Conclusions Strict compliance with the measurement protocol yields tumor response classifications that may differ from those obtained in clinical practice. Response based on retrospective measurements correlated better with survival than did response based on primary measurements.

Key Points

- *Response classifications could be different between clinical primary and retrospective measurements for malignant pleural mesothelioma.*
- *Response classifications obtained by strict compliance with the trial-specific protocol correlated better with survival than the classifications based on primary measurements.*
- *Quality assurance and radiologist training measures should be used to ensure the integrity of image-based tumor measurements in mesothelioma clinical trials.*

Keywords Mesothelioma, malignant · Thorax · Tomography, X-ray computed · Response evaluation criteria in solid tumors · Survival analysis

Abbreviations

CT	Computed tomography
MPM	Malignant pleural mesothelioma
RECIST	Response evaluation criteria in solid tumors

✉ Feng Li
feng@uchicago.edu

[1] Department of Radiology, The University of Chicago, 5841 S. Maryland Avenue, Chicago, IL 60637, USA

[2] Department of Medicine, The University of Chicago, Chicago, IL, USA

Introduction

The image-based assessment of tumor growth or shrinkage requires (1) identification and reporting of individual lesions at the baseline scan and (2) tracking these lesions across all follow-up scans, as specified by the response evaluation criteria in solid tumors (RECIST) [1, 2]. Due in part to inherent differences in the workflow between clinical patients and patients who are also research subjects and differences in the reporting needs of these two patient groups, radiologists do not always assess tumor response in strict compliance with the RECIST guidelines. Also, intra- and inter-observer variability in the application of guidelines for the measurement of target lesions [3–5] could impact the assigned tumor response classification category, which could change patient management.

Malignant pleural mesothelioma (MPM) has a distinctly non-spherical growth pattern, and modified RECIST [6] adapted to this unique morphology by specifying the acquisition of unidimensional measurements of tumor thickness perpendicular to the chest wall or mediastinum. Observer variability in these thickness measurements (in controlled settings) has been shown to impact tumor response classification [7–9]; the acquisition of imaging measurements during normal clinical workflow will be more complex than the measurement tasks reported in these more idealized studies, with the expectation of increased observer variability.

The need exists to investigate actual clinical measurement of MPM tumor thickness and whether measurement differences can change patient response assessment. This study evaluated differences in the tumor response classifications that result from measurements acquired during routine clinical workflow ("primary measurements") and measurements that conformed to the clinical trial measurement protocol ("retrospective measurements"); these response classifications were compared with overall survival for MPM patients on a clinical trial.

Materials and methods

Patient population

The images and clinical data from MPM patients from a single center who had participated in a phase III clinical trial of a targeted agent, vorinostat (suberoylanilide hydroxamic acid), versus placebo (NCT 00128102) between 2006 and 2012 [10] were used in this study in a HIPAA-compliant manner with institutional review board approval. This clinical trial included 41 patients (33 men and 8 women; age at baseline computed tomography (CT) scan, range 48–84 years; mean 67.6 years). All patients had advanced MPM and had been previously treated with systemic chemotherapy. Thirty-eight of these patients had died of their disease or a related medical condition, and three patients were lost to follow-up. The mean time interval between baseline CT scan and death for the 38 patients was 390 days (range 59–1702 days).

Follow-up scans were obtained every two treatment cycles. Each patient had a baseline CT scan and between two and ten follow-up CT scans on which primary measurements had been acquired, for a total of 131 scans across all patients (41 baseline scans and 90 follow-up scans). The 131 scans included 119 contrast-enhanced and 12 non-enhanced scans. Scans were performed on Philips Brilliance 16- or 64-slice scanners. Peak voltage was 120 kVp ($n = 120$) or 140 kVp ($n = 11$) with variable exposure rate according to patient size. Section thickness was 3 mm. All images had been reconstructed as 512 × 512-pixel images.

Primary measurements

All 131 scans had been measured by radiologists in our Department of Radiology through the routine clinical workflow; these primary measurements had been saved at each time point within the picture archiving and communications system. Primary measurements were extracted from the radiology reports by a study chest radiologist (F.L., 17 years of experience). The study radiologist reviewed all measurements recorded by the clinical radiologists from all 131 CT scans.

Retrospective measurements

Prior to review of the primary measurements and reports, the same study radiologist retrospectively acquired tumor measurements from all patients enrolled in the trial based on the study protocol, which specified modified RECIST (see Fig. 1). Six linear tumor thickness measurements were acquired perpendicular to the chest wall or mediastinum adjacent to fixed structures such as the ribs and the vertebral column (two measurements on each of three axial CT sections separated by more than 1 cm) from the baseline scan of each patient in accordance with an exact implementation of modified RECIST. In total, 246 baseline scan measurements were acquired across the 41 patients. These retrospective measurements acquired from the baseline scans were replicated in all corresponding follow-up scans for each patient.

Modified RECIST specifies the acquisition of six measurements and does not explicitly set a lower limit on measurement size; in practice, however, this approach generally has been modified to the acquisition of up to six measurements of at least 10 mm each. To accommodate analyses according to both approaches, retrospective tumor measurements as small as 1 mm were obtained to achieve exactly six measurements per scan; assessments then were performed (1) with all retrospective measurements (six per scan) and (2) only with those retrospective measurements ≥ 10 mm (up to six per scan). The number of measurements ≥ 10 mm among the 246

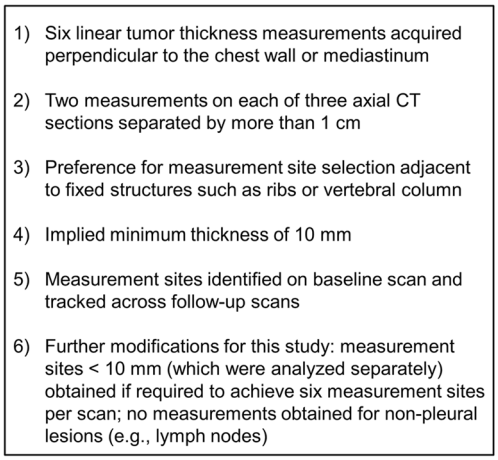

Fig. 1 Summary of the retrospective measurements per modified RECIST

retrospective measurements from the baseline scans was 114 (46%). In total, 786 retrospective measurements across all baseline and follow-up scans were obtained on CT images displayed at a mediastinal window (width 400/level 40). Five non-pleural suspicious lesions (three mediastinal lymph nodes, one non-calcified lung nodule, and one chest wall mass) were present in the baseline scans but were not included in the retrospective measurements and not extracted from the primary measurements when available; although modified RECIST allows for bi-dimensional measurements of non-pleural lesions, the purpose of this study was to evaluate variability in the implementation of the main component of modified RECIST: measurement of pleural tumor thickness.

Data analysis

Tumor response classification based on (1) the sum of primary measurements (various numbers of tumor thickness measurements per scan) or (2) the sum of retrospective measurements (six tumor thickness measurements per scan) was performed for each patient according to the RECIST guidelines: (1) a reduction of at least 30% in summed measurements between the baseline scan and each follow-up scan was classified as partial response (PR), (2) an increase of at least 20% in summed measurements was classified as progressive disease (PD), and (3) neither sufficient shrinkage to qualify as PR nor sufficient increase to qualify as PD was considered stable disease (SD).

The variability of primary measurements was investigated. Baseline scan summed primary measurements were compared with the corresponding retrospective measurements. A paired Student's t test was used to evaluate the difference in the means between the primary and retrospective summed measurements after a Kolmogorov-Smirnov test was used to test the normality of the data. Differences in response classification categories assigned to each patient based on both sets of measurements were evaluated. These analyses were repeated with just those primary and retrospective measurement sites that exceeded the RECIST 10-mm minimum measurable disease threshold at baseline; in these subsequent analyses, only the six largest primary measurements were used if the scan included more than six primary measurements.

Harrell's C statistic [11] was used to evaluate the correlation of survival (from the date of the baseline CT scan) with best response determined from all measurements and then from just those measurements ≥ 10 mm (not exceeding six per scan) among the primary and retrospective measurements. Bootstrapping was performed to compare the differences between the C statistic obtained from these different measurement approaches, and the 95% confidence intervals obtained for these comparisons were computed from 1000 bootstrap iterations. It should be noted that vorinostat proved to be ineffective in this patient cohort, so the number of responding patients was minimal [10]. In general, if response category is correlated with survival, then patients classified as PR should survive longer than patients classified as SD, and both groups should survive longer than PD patients. A value of $C = 0.5$ is equivalent to correct classification by chance alone.

Results

The primary interpretation of the 131 CT scans from 41 MPM patients involved 13 radiologists. The numbers of scans measured by the 13 radiologists [1] by themselves and [2] with a radiology resident were 57 (44%) and 74 (56%), respectively. Variability in the numbers of scans per patient and the numbers of radiologists involved in the acquisition of tumor measurements across a patient's scans is shown in Table 1. The 41 baseline scans included 177 primary measurements of tumor thickness ($n = 168$; 95%) or longest tumor diameter ($n = 9$; 5%). The median numbers of (1) measurement sites per scan and (2) sections per scan from which measurements had been acquired across the 41 baseline scans were 4.0 sites per scan (range 2–9) and 3.0 sections per scan (range 2–5), respectively, for the 41 baseline scans. The primary measurements on the baseline or follow-up scans of 11 patients deviated from modified RECIST by spanning more than three CT sections and/or capturing more than two sites per section. The minimum size of the 177 primary measurements on the baseline scans was 1 mm; the number of measurements < 10 mm among the primary measurements was 59 (33%). Of the 177 baseline scan primary measurements, 170 (96%) had corresponding primary measurements across all follow-up scans from the respective patient. In total, 573 primary measurements across all baseline and follow-up scans were obtained, with images displayed primarily with a mediastinal window

Table 1 Distributions of the number of scans per patient and the number of radiologists involved in primary measurement acquisition across a patient's scans (involving 131 total CT scans and 13 radiologists)

Scans/patient	Number of patients ($N = 41$)
2	21
3	9
4	5
5	1
6	2
7	1
8	1
9	0
10	1
Radiologists/patient	
1	6
2	22
3	11
4	2

(but the window width and level were not always the same). Figures 2 and 3 present example images that demonstrate differences between primary and retrospective tumor measurements and differences in the resulting tumor response classifications (discussed below).

No statistically significant difference was observed between the summed primary and summed retrospective measurements across the 41 baseline scans ($p = 0.78$) (Table 2). When only the retrospective measurements that satisfied the 10-mm minimum measurable disease threshold were compared with the primary measurements, summed measurements across baseline scans achieved a statistically significant difference between retrospective and primary measurements ($p \leq 0.002$).

Scan-by-scan RECIST response classification based on primary and retrospective measurements was different in 23 (26%) of the 90 follow-up scans, and RECIST best response across all follow-up scans of a patient based on primary and retrospective measurements was different in seven (17%) of the 41 patients (Table 3). Best response based on retrospective measurements correlated better with overall survival ($C = 0.62$, 95% confidence interval [0.57, 0.67]) than did best response based on primary measurements ($C = 0.57$, 95% confidence interval [0.52, 0.62]); both were statistically significantly different from random guessing because their 95% confidence intervals do not include 0.5, but the difference between retrospective measurements ($C = 0.62$) and primary measurements

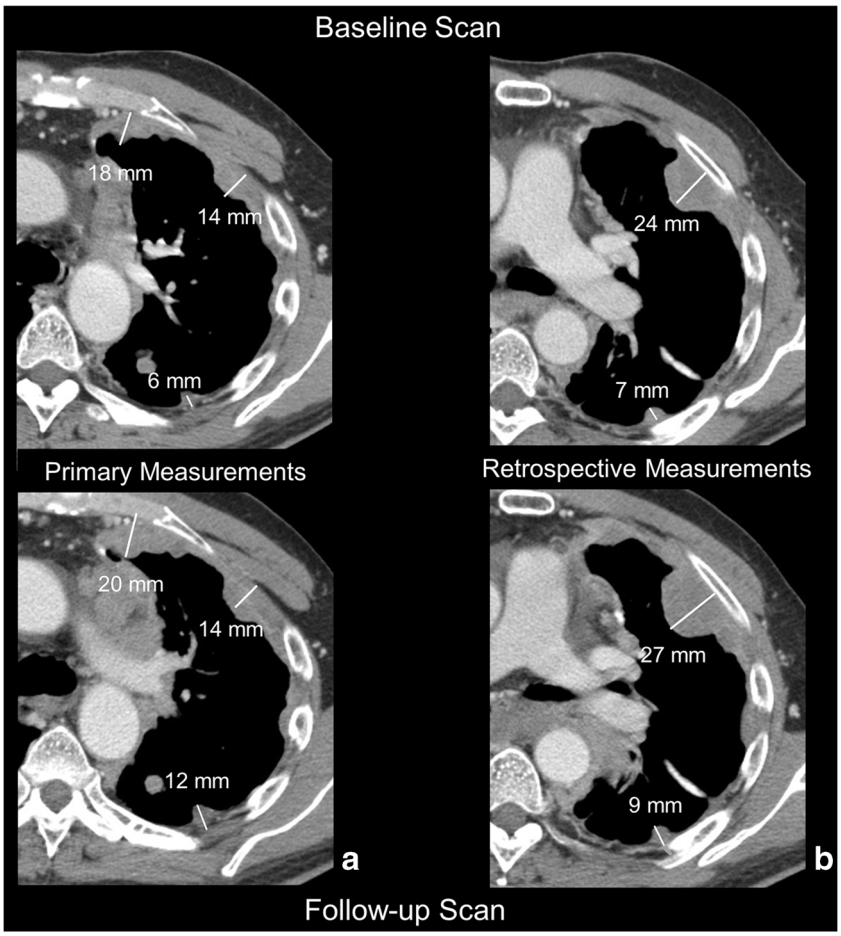

Fig. 2 Example sections from baseline scan and first follow-up scan (after two treatment cycles) of a 67-year-old man with (**a**) three primary measurement sites and (**b**) two retrospective measurement sites. Patient best response based on primary measurements and retrospective measurements was SD and PD, respectively. Note that the acquired primary measurements were not always perpendicular to fixed anatomic structures. Time between the best response CT scan and death was 36 days

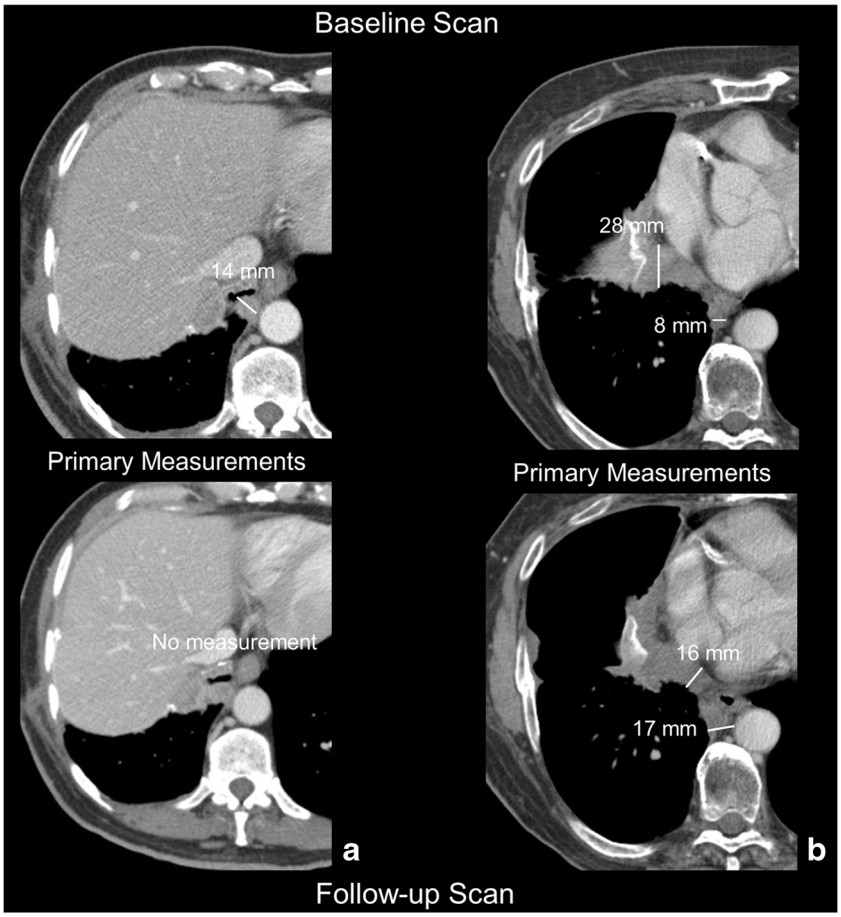

Fig. 3 Primary measurements on baseline scan and best-response follow-up scan sections from (**a**) a 66-year-old man with one primary measurement site (note that this one baseline scan site was not measured on the follow-up scan) and (**b**) a 77-year-old man with two primary measurement sites (note that these two sites were not consistently measured on the follow-up scan). Patient A best response based on primary measurements and retrospective measurements (not shown) was PR and SD, respectively. Patient B best response based on primary measurements and retrospective measurements (not shown) was SD and PD, respectively

($C = 0.57$) failed to achieve statistical significance. When limited to measurement sites that were at least 10 mm at baseline, best response correlation with survival for retrospective measurements decreased from $C = 0.62$ to $C = 0.52$ (95% confidence interval [0.46, 0.58]), and best response correlation with survival for primary measurements decreased from $C = 0.57$ to $C = 0.38$ (95% confidence interval [0.32, 0.44]). Time between the follow-up CT scan that captured best response and the CT scan that finally diagnosed disease progression or death for the seven patients with different classification by primary and retrospective measurements is shown in Table 4. Table 5 shows the 95% confidence intervals, obtained from bootstrapping, of the differences in C statistic between the measurement approaches. Likely due to the relatively small number of cases used, a statistically significant difference in C statistic (indicated by a 95% confidence interval that does not include 0) was observed for only one of these comparisons: the primary measurements using all measurements and the primary

Table 2 Analysis of primary measurements and retrospective measurements

	Baseline scans		Baseline and follow-up scans	
	Primary	Retrospective	Primary	Retrospective
Individual measurements	$N = 177$	$N = 246$	$N = 573$	$N = 786$
Mean (SD) (mm)	15.0 (11.3)	11.0 (7.6)	14.7 (12.7)	11.1 (8.5)
Median (mm)	13	9	12	9
Range (mm)	1–85	1–43	1–101	1–52
Summed measurements	$N = 41$	$N = 41$	$N = 131$	$N = 131$
Mean (SD) (mm)	64.6 (33.2)	65.7 (21.7)	64.5 (37.2)	66.9 (25.7)
Median (mm)	62	69	58	68
Range (mm)	14–166	26–118	11–196	12–146

SD standard deviation

Table 3 RECIST-based response category

	Response for individual follow-up scans ($N = 90$)			
Retrospective measurement	Primary measurement			
	PR	SD	PD	Total
PR	3	0	0	3
SD	1	51	10	62
PD	2	10	13	25
Total	6	61	23	90
	Best response for individual patients ($N = 41$)			
Retrospective measurement	Primary measurement			
	PR	SD	PD	Total
PR	1	0	0	1
SD	1	26	3	30
PD	0	3	7	10
Total	2	29	10	41

PR partial response, *PD* progressive disease, *SD* stable disease

measurements limited to measurement sites that were at least 10 mm at baseline.

Discussion

Radiologists often are expected to apply one approach to image interpretation for standard-of-care oncology patients and another approach (or, more precisely, one of a collection of other possible approaches) to image interpretation for oncology patients enrolled on clinical trials. Since clinical trials usually involve multiple institutions, protocols have been established to guide image interpretation across all patients and all sites in an attempt to ensure consistency in tumor response assessment, which is often a trial endpoint. Achieving consistency can be a challenge since radiologists less familiar with the protocol (or unaware that a particular patient is enrolled in a trial) could be involved when these cases come through the clinical workflow. The purpose of this study was to evaluate differences in the tumor response classifications that result from primary measurements (acquired during normal clinical workflow) and retrospective measurements (retrospectively acquired in accordance with uniform guidelines) for MPM patients and to compare these response classifications with overall patient survival.

Table 4 Time between best response CT scan and final PD CT scan or death for the seven patients with different classification

Patient	Primary measurement	Retrospective measurement	Final PD time (days)	Death time (days)
Case 1	PR	SD	219	NA
Case 2	SD	PD	NA	36
Case 3	SD	PD	NA	73
Case 4	SD	PD	205	453
Case 5	PD	SD	29	117
Case 6	PD	SD	121	270
Case 7	PD	SD	139	280

PR partial response, *PD* progressive disease, *SD* stable disease, *NA* not available

Table 5 95% confidence intervals for differences in *C* statistic between measurement approaches

Measurement approach comparison	95% confidence interval
Retrospective (all) vs. primary (all)	(− 0.097–0.21)
Retrospective (all) vs. retrospective (≥ 10 mm)	(− 0.098–0.28)
Primary (all) vs. primary (≥ 10 mm)	(0.065–0.34)
Retrospective (≥ 10 mm) vs. primary (≥ 10 mm)	(− 0.064–0.39)

Evaluating a large multicenter oncology trial, Thiesse et al [4] reported differences between tumor response classifications by local investigators and by a central review committee and suggested that all clinical trial imaging should undergo independent central review. While independent central review may maximize consistency, such review has its own challenges [12–14]. Tumor response (and the approach used to obtain the necessary measurements) also differs among radiologists within a single site; the present study investigated variability in local primary measurements using the chest CT scans of 41 MPM patients enrolled in a single clinical trial at our medical center. The results demonstrated that many radiologists were involved in the imaging evaluation of tumor response classification, and inconsistent measurement approaches had been implemented. Central review could overcome some of these fundamental issues of consistency, and one or several radiologists dedicated to one or several clinical trials within an institution likely would serve to reduce the variability that was observed in this single-institution study.

The findings of Muenzel et al [5] suggested that the reproducibility of tumor response classification may be improved when the imaging assessment is performed by a single observer or by the combined assessment of four observers. In the present study, an experienced chest radiologist acquired retrospective measurements in accordance with modified RECIST. Response classification based on primary and retrospective measurements differed in 23 (26%) of the 90 follow-up scans, and the retrospective best response differed from the primary best response in seven (17%) of the 41 patients. These differences may have resulted from a combination of factors affecting the primary readings of these clinical trial scans: limited reading time, lack of training on the protocol measurement guidelines, ineffective communication between radiologists and the clinical trial team, ambiguities in the modified

RECIST guidelines, and variability in the selection of baseline scan measurement sites as well as inter-observer variability in the matching of measurement sites in subsequent follow-up scans.

Jain et al [15] analyzed image-based tumor response for 570 oncology patients from 24 clinical trials and did not find a clear relationship between tumor response classification and overall survival. Another study, however, found that the correlation of patient survival with tumor response classification according to modified RECIST (as measured by a single reader) yielded a *C* value of 0.78 in 78 MPM patients who underwent standard-of-care chemotherapy [16]. In the present study, correlation of survival with best response achieved a *C* value of 0.62 based on retrospective measurements.

The acquisition of measurements from the CT scans of MPM patients has some challenges. First, it may not be possible to measure two distinct tumor sites with thickness ≥ 10 mm in the same CT section, especially in the case of previously treated patients. Consequently, the thickness of individual measurement sites from the primary measurements (mean, 15 mm) was greater than that from the retrospective measurements (mean, 11 mm), which effectively forced the selection of two measurement sites in each of three CT sections to capture the greatest extent of tumor burden but without limiting measurements to those ≥ 10 mm. Second, it may be difficult to track corresponding measurement sites across temporally sequential scans to evaluate size change. Tumor thickness can change in apparent size, for example, due to differences in patient positioning or rotation between scans or due to changes in post-treatment volume of the thoracic cavity. Third, the validity of the 10-mm threshold for "measurable disease" has already been challenged in the MPM setting [9]; the present study found better correlation of survival with best response (based on primary or retrospective measurements) when measurement sites were not limited to those that exceeded 10 mm.

This retrospective study had several limitations. First, the MPM patients were obtained from a single clinical trial at one medical center. Second, the retrospective measurements were obtained by a single study radiologist. Third, to avoid another potential source of error and variability, the actual radiology reports were used rather than the case report forms, which are extracted from the radiology reports by a local study coordinator and used to transmit measurements to the study sponsor during a clinical trial.

Computerized image-assessment systems, tumor volume quantification, and PET-CT scans have been all evaluated for MPM tumor response classification, and these advanced measurement techniques have demonstrated correlation with patient survival [17–22]. Nevertheless, the more simple linear measurement approach of modified RECIST remains the standard for the evaluation of tumor response classification in MPM. Strict adherence to the measurement protocol, however, yields response classifications that are better correlated with patient survival than less-consistent measurements obtained during routine clinical interpretation.

Quality assurance and radiologist training measures should be used to ensure the integrity of image-based tumor thickness measurements in MPM clinical trials. We believe that the quality, consistency, and reliability of image-based tumor measurements at local sites would be enhanced through (1) effective communication between the radiologists and the clinical trial team, (2) radiologist training on protocol measurement guidelines, and (3) minimization of the number of involved radiologists, ideally limited to a few protocol-trained individuals responsible for imaging measurements across multiple clinical trials.

Acknowledgements This study has been presented in part at the RSNA 2011 meeting.

Funding The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Feng Li.

Conflict of interest The authors of this manuscript declare relationships with the following companies: Dr. Armato is a consultant for Aduro Biotech, Inc. Dr. MacMahon is a shareholder of Hologic and a consultant for Riverain Technologies. Drs. Li, Armato, and MacMahon receive royalties and licensing fees for computer-aided diagnosis technology through The University of Chicago. Dr. Kindler is a consultant for Merck and receives research support from Merck.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
- Retrospective
- Cross sectional study
- Performed at one institution

References

1. Therasse P, Arbuck AG, Eisenhauer EA et al (2000) New guidelines to evaluate the response to treatment in solid tumor. J Natl Cancer Inst 92:205–216
2. Eisenhauer EA, Therasse P, Bogaerts J et al (2009) New response evaluation criteria in solid tumors: revised RECIST guideline (version 1.1). Eur J Cancer 45:228–247
3. Hopper KD, Kasales CJ, van Slyke MA, Schwartz TA, TenHave TR, Jozefiak JA (1996) Analysis of interobserver and intraobserver variability in CT tumor measurements. AJR Am J Roentgenol 167: 851–854

4. Thiesse P, Ollivier L, Di Stefano-Louineau D et al (1997) Response rate accuracy in oncology trials: reasons for interobserver variability. J Clin Oncol 15:3507–3514
5. Muenzel D, Engels HP, Bruegel M, Kehl V, Rummeny EJ, Metz S (2012) Intra- and inter-observer variability in measurement of target lesions: implication on response evaluation according to RECIST 1.1. Radiol Oncol 46:8–18
6. Byrne MJ, Nowak AK (2004) Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. Ann Oncol 15:257–260
7. Armato SG 3rd, Oxnard GR, MacMahon H et al (2004) Measurement of mesothelioma on thoracic CT scans: a comparison of manual and computer-assisted techniques. Med Phys 31:1105–1115
8. Armato SG 3rd, Ogarek JL, Starkey A et al (2006) Variability in mesothelioma tumor response classification. AJR Am J Roentgenol 186:1000–1006
9. Armato SG 3rd, Nowak AK, Francis RJ, Kocherginsky M, Byrne MJ (2014) Observer variability in mesothelioma tumor thickness measurements: defining minimally measurable lesions. J Thorac Oncol 9:1187–1194
10. Krug LM, Kindler HL, Calvert H et al (2015) Vorinosat in patients with advanced malignant pleural mesothelioma who have progressed on previous chemotherapy (VANTAGE-014): a phase 3, double-blind, randomized, placebo-controlled trial. Lancet Oncol 16:447–456
11. Harrell FE Jr, Lee KL, Mark DB (1996) Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15:361–387
12. Ford R, Schwartz L, Dancey J et al (2009) Lessons learned from independent central review. Eur J Cancer 45:268–274
13. Dodd LE, Korn EL, Freidlin B et al (2008) Blinded independent central review of progression-free survival in phase III clinical trials: important design element or unnecessary expense? J Clin Oncol 26:3791–3796
14. Amit O, Mannino F, Stone AM et al (2011) Blinded independent central review of progression in cancer clinical trials: results from a meta-analysis. Eur J Cancer 47:1772–1778
15. Jain RK, Lee JJ, Ng C et al (2012) Change in tumor size by RECIST correlates linearly with overall survival in phase I oncology studies. J Clin Oncol 30:2684–2690
16. Labby ZE, Armato SG 3rd, Kindler HL, Dignam JJ, Hasani A, Nowak AK (2012) Optimization of response classification criteria for patients with malignant pleural mesothelioma. J Thorac Oncol 7:1728–1734
17. Liu F, Zhao B, Krug LM et al (2010) Assessment of therapy response and prediction of survival in malignant pleural mesothelioma through computer-aided volumetric measurement on computed tomography scans. J Thorac Oncol 5:879–884
18. Ak G, Metintas M, Metintas S, Yildirim H, Ozkan R, Ozden H (2010) Three-dimensional evaluation of chemotherapy response in malignant pleural mesothelioma. Eur J Radiol 74:130–135
19. Gill RR, Richards WG, Yeap BY et al (2012) Epithelial malignant pleural mesothelioma after extrapleural pneumonectomy: stratification of survival with CT-derived tumor volume. AJR Am J Roentgenol 198:359–363
20. Labby ZE, Nowak AK, Dignam JJ, Straus C, Kindler HL, Armato SG 3rd (2013) Disease volumes as a marker for patient response in malignant pleural mesothelioma. Ann Oncol 24:999–1005
21. Francis RJ, Byme MJ, van der Schaaf AA et al (2007) Early prediction of response to chemotherapy and survival in malignant pleural mesothelioma using a novel semiautomated 3-dimensional volume-based analysis of serial 18F-FDG PET scans. J Nucl Med 48:1449–1458
22. Cheng L, Tunariu N, Collins DJ et al (2015) Response evaluation in mesothelioma: beyond RECIST. Lung Cancer 90:433–441