

Published in final edited form as:

J Thorac Oncol. 2013 April ; 8(4): 478–486. doi:10.1097/JTO.0b013e31828354c8.

Lung Volume Measurements as a Surrogate Marker for Patient Response in Malignant Pleural Mesothelioma

Zacariah E. Labby, PhD¹, Samuel G. Armato III, PhD¹, James J. Dignam, PhD², Christopher Straus, MD¹, Hedy L. Kindler, MD³, and Anna K. Nowak, MD, PhD^{4,5}

¹Department of Radiology, The University of Chicago, Chicago, IL

²Department of Health Studies, The University of Chicago, Chicago, IL

³Department of Medicine, The University of Chicago, Chicago, IL

⁴Department of Medical Oncology, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia

⁵School of Medicine and Pharmacology, University of Western Australia, Crawley, Western Australia, Australia

Abstract

Introduction—The purpose of this study was to investigate continuous changes in three distinct response assessment methods during treatment as a marker of response for mesothelioma patients. Linear tumor thickness measurements, disease volume measurements, and lung volume measurements (a physiological correlate of disease volumes) were investigated in this study.

Methods—Serial CT scans were obtained during the course of clinically standard chemotherapy for 61 patients. For each of the 216 CT scans the aerated lung volumes were segmented using a fully automated method, and the pleural disease volume was segmented using a semi-automated method. Modified RECIST linear thickness measurements were acquired clinically. Diseased (ipsilateral) lung volumes were normalized by the respective contralateral lung volumes to account for differences in inspiration between scans for each patient. Relative changes in each metric from baseline were tracked over the course of follow-up imaging. Survival modeling was performed using Cox proportional hazards models with time-varying covariates.

Results—Median survival from pre-treatment baseline imaging was 12.7 months. A negative correlation was observed between measurements of lung volume and disease volume, and a positive correlation was observed between linear thickness measurements and disease volume. As continuous numerical parameters, all three response assessment methods were significant imaging biomarkers of patient prognosis in independent survival models.

Conclusions—Analysis of trajectories of linear thickness measurements, disease volume measurements, and lung volume measurements during chemotherapy for patients with

Corresponding Author: Zacariah E. Labby Department of Radiology The University of Chicago 5841 S Maryland Ave. MC 2026, P-104 Chicago, IL 60637 Phone: 773-834-5098 Fax: 773-702-0371 zelabby@uchicago.edu.

Conflicts of Interest:

ZEL – none.

SGA – receives royalties and licensing fees through the University of Chicago related to computer-aided diagnosis.

JJD – none. CS – none. HLK – none. AKN – 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.

mesothelioma indicates that increasing linear thickness, increasing disease volume, and decreasing lung volume are all significantly and independently associated with poor patient prognosis.

I. INTRODUCTION

For matters involving tumor response, there is only one metric that can be used to ascertain the truth: tumor burden. If tumor composition is assumed to be consistent over time, then changes in tumor volume will directly correspond to changes in the number of tumor cells. Some molecular imaging methods are moving toward proliferative cellular quantification [1–3]. However, until these methods become widespread, computed tomography (CT) imaging (with the possibility of volumetric quantification) will remain the best tool to assess tumor burden for patients with malignant pleural mesothelioma (MPM).

Advances in medical imaging and image processing methodology allow for response assessment metrics that (1) use full three-dimensional volume measurements [4–6] and (2) track continuous, rather than discretized, measurements over time [7, 8]. Disease volumes are a logical choice for tumor burden assessment of diseases such as mesothelioma, where the disease morphology is not compatible with the spherical geometry assumptions implicit in the Response Evaluation Criteria In Solid Tumors (RECIST) response assessment technique [9–11]. The segmentation and volumetric quantification of mesothelioma with any degree of automation is a challenging task. The morphology of the disease is widely variable, and its radiographic density is comparable to that of neighboring tissues [12]. While volume measurements of MPM have been shown to exhibit lower inter-observer variability than linear thickness measurements made according to the modified RECIST protocol [13, 14], the computational and manual challenges of the disease volume segmentation task are problematic.

Pleural disease volume was previously shown to be a significant predictor of MPM patient survival [3, 15, 16], but changing tumor burden affects more than just the volume of tumor. The hemithoracic space is fairly fixed so that when disease volume increases, aerated lung volume should be expected to decrease correspondingly. This physiologic correlation implies that changes in lung volume may have prognostic value for patients with MPM. Lung volume has been investigated to monitor response to surgical MPM tumor debulking [17]; changes in lung volume may also be a useful tool to assess tumor response for patients receiving chemotherapy so that instead of classifying response from declining tumor volume, response would be classified from increasing lung volume.

Both linear measurements based on modified RECIST [15] and lung volumes have certain advantages over disease volumes for response assessment. Disease volumes require substantial manual intervention. Linear thickness measurements are almost entirely manual (though some automation techniques have been suggested [18]) but require much less time than disease volume segmentation. Lung volume segmentation, on the other hand, is entirely automated. The purpose of this study was to compare the prognostic performance of changing lung volumes and linear thickness measurements (treated continuously) with changing disease volumes in survival models for patients with MPM.

II. PATIENTS AND METHODS

A. Patient Cohort

Imaging and clinical data from 61 patients were obtained from a prospective study involving FDG-PET and CT imaging of MPM [3]. All patients were over 18 years old with histologically or cytologically confirmed MPM and had not received prior chemotherapy or definitive radiotherapy. Patient accrual occurred from late 2003 to 2010, and the original

study was approved by the local institutional Human Research Ethics Committee at Sir Charles Gairdner Hospital (Nedlands, Australia) with patients providing written informed consent. The retrospective analysis of the Health Insurance Portability and Accountability Act (HIPAA)-compliant data was approved by both the originating institution's Human Research Ethics Committee and the Institutional Review Board at The University of Chicago, where the analysis was performed. Because the original study did not mandate a specific treatment protocol, patients were treated as clinically indicated. Initially, combination chemotherapy consisted of cisplatin and gemcitabine, and later, when it became available at the original study institution, cisplatin and pemetrexed. For inclusion in the present study, patients were required to have available modified RECIST tumor thickness measurements at baseline (prior to beginning chemotherapy) and one or more follow-up scans during chemotherapy. Since lung volume analysis was limited to patients with one non-diseased lung to serve as a control, the patients were also required to have unilateral disease. Finally, all patients were required to have a complete thoracic CT scan for all scan dates (and not simply scanned films) for automated lung segmentation. The summary description of the patient cohort is given in Table I.

B. Imaging

Patients were imaged using helical CT up to one month prior to the first cycle of chemotherapy and throughout their treatment regimen (typically after the first cycle, then every two cycles thereafter). CT staging was performed according to the Union for International Cancer Control (UICC) TNM staging system (2002). CT scans were staged by a thoracic radiologist or medical oncologist experienced in mesothelioma imaging, and tumor measurements were made clinically according to the modified RECIST protocol on baseline and all follow-up scans [13]. Pathologic staging was not performed. The clinical measurement protocol dictated that all imaging examinations from an individual patient be measured by the same clinician in an attempt to minimize variability.

A total of 216 CT scans were used in this study, with a median of four scans per patient. Eight patients had only a baseline scan with one follow-up scan, while 19 patients had three scans total, 27 patients had four scans total, and seven patients had five scans total. The median interval between scans was 48 days. Of the 216 scans, 150 scans had been performed on General Electric scanners (HiSpeed CT/i, n=81; LightSpeed Pro 16, n=1; or LightSpeed VCT, n=68), and 66 scans had been performed on Philips Brilliance 64-slice scanners. At least 101 of the scans were performed with iodinated contrast media.

Only one reconstructed series was required for lung and disease segmentation for each CT scan date, and this series was selected for each patient with consideration for reconstruction kernel and slice thickness. Preference was given to thinner slice thicknesses and "Standard" reconstruction kernels, but if for a given patient there was a scan date with only "Lung" kernel reconstructions, then matched kernel and slice thickness reconstructions were used for the other scan dates. Having this type of consistency across the scan dates for a given patient was considered important for segmentation of volumetric disease, since different amounts of disease might be segmented on different reconstructions due to, for instance, partial volume effects. Although linear thickness measurements were consistently acquired using 5-mm reconstructions, multiple reconstructed slice thicknesses existed for each CT scan. For the series used in the lung and disease segmentation components of this study, slice thicknesses were 0.63 mm (n=4), 1 mm (n=14), 1.25 mm (n=28), 2.5 mm (n=75), or 5 mm (n=95). In-plane voxel dimensions ranged from 0.54–0.86 mm, and all reconstructed axial images had an in-plane matrix size of 512 by 512 pixels. The kVp setting for the scans was predominantly 120 kVp (n=212), with 100 kVp (n=1) and 140 kVp (n=3) also used. Reconstruction kernels fell into two broad categories, with "Lung" kernels (including the Philips "L" and General Electric "Lung" kernels) used for 136 scans and "Standard" kernels

(including Philips “B” and General Electric “chest,” “soft,” and “standard” kernels) used for the remaining 80 scans.

C. Lung and Disease Volume Quantification

Lung region segmentation was performed using a segmentation algorithm described previously by Sensakovic et al. [19]. The lung segmentation method is fully automated and utilizes gray-level, morphological, and texture features to segment the aerated lung regions. The lung segmentation method has proven successful in other studies for patients with MPM [17, 20]. The resulting segmentations were all reviewed for accuracy and modified when necessary by an observer (ZEL) trained in thoracic anatomy. In-house software was used for this task (“Abrax”), and duration of any necessary intervention was tracked.

The pleural disease was segmented in each scan using a semi-automated method described previously [21]. Because of the considerable overlap in Hounsfield Unit (HU) values between mesothelioma tumor and pleural effusion [12], the semi-automated disease volume segmentation method produces contours of pleural disease and does not readily separate tumor from effusion. Therefore, the end goal of the disease segmentation technique used in this study was reliable volumes of pleural disease and not necessarily volumes of only mesothelioma tumor. To calculate lung volume and pleural disease volume for each patient scan, a pixel-counting technique was used [22].

As an independent validation of the lung segmentation method, lung segmentations were performed on a separate set of 44 CT scans from 22 patients with MPM (one baseline and one follow-up scan per patient). Automated lung segmentation was performed for each patient, and an attending radiologist (who was blinded to the computer results) contoured the aerated lung on three axial sections for the diseased (ipsilateral) hemithorax and healthy (contralateral) hemithorax (patients had unilateral disease). The area enclosed by both sets of contours was calculated, and the section-by-section areas were compared using Pearson's correlation coefficient and Bland-Altman analysis [23].

Lung volumes were used as a response assessment metric by normalizing the ipsilateral lung volume by the contralateral lung volume for each patient scan. While it is customary for CT scans to be acquired during patient breath-hold at full inspiration, it is possible that differences in patient respiratory phase between scan time points still exist. In patients with unilateral disease, the healthy (contralateral) lung can be used to normalize the volume of aerated lung in the diseased (ipsilateral) hemithorax, thereby controlling for any potential differences in inspiration. This normalized volume V_{norm} was calculated as

$$V_{norm}(t) = \frac{V_{ipsilateral}(t)}{V_{contralateral}(t)}. \quad (1)$$

D. Data Analysis

The different tumor response assessment methods in this study (linear thickness measurements, disease volumes, and normalized lung volumes) were compared using rank correlation statistics. An R^2 value is reported for the fit between changes in linear thickness from baseline and changes in disease volume from baseline for a spherical geometry model (the geometry implicit in the derivation of the RECIST classification criteria). For a sphere with diameter d that changes by an amount Δd , the relative volume change is given by

$$\frac{\Delta V}{V} = \left(\frac{\Delta d}{d}\right)^3 + 3\left(\frac{\Delta d}{d}\right)^2 + 3\left(\frac{\Delta d}{d}\right). \quad (2)$$

To compare the prognostic performance of the different response assessment methods, the univariate significance of all three metrics was assessed using Cox proportional hazards (PH) models with time-varying covariates [24–26]. Next, survival models were built using each response assessment method and the clinical covariates from the final multivariate prognostic model obtained by Labby et al. [21]: Eastern Cooperative Oncology Group (ECOG) performance status discretized as level 0 versus levels 1 or 2, disease histology discretized as epithelioid versus other, and presence of dyspnea. Survival was defined as the duration from baseline imaging to either patient death or censoring (some patients in the cohort remain living).

All three response assessment methods were allowed to change over time and were modeled using scaled logarithmic transforms of relative changes from baseline, known as the specific growth rate (SGR) [8]. The definition of the SGR metric is

$$SGR(t) = \frac{\ln \left[\frac{m(t)}{m(t_0)} \right]}{(t - t_0)}, \quad (3)$$

where $m(t)$ denotes the measurement (linear thickness, disease volume, or normalized lung volume) at an arbitrary time point and t_0 indicates the time of baseline scanning (times in this study were all modeled as fractional years). The clinical covariates mentioned earlier were included along with (1) linear measurement SGR, (2) disease volume SGR, or (3) normalized lung volume SGR in multivariate survival models.

The performance of the survival models was assessed using the Heagerty's C^x [27], derived from receiver operating characteristic (ROC) analysis. C^x is especially useful for survival models with time-varying covariates and is scaled from 0 to 1; $C^x = 0.5$ would indicate no prognostic ability, and $C^x = 1.0$ would indicate perfect prognostic ability. For this study, values of C^x are reported from training and testing on the same dataset, as well as leave-one-out cross-validation (LOOCV) performance values for the different models. Additionally, repeated random sub-sampling of the patient cohort was used to assess the difference in predictive ability between the three multivariate survival models for the different response assessment methods. In each of 1000 sub-sample iterations, each model was trained on two-thirds of the patient cohort and tested on the remaining one-third of the patient cohort. The training set was chosen randomly without replacement at each iteration, and the testing set was considered to be the remaining patients who had not been selected for the training set at that iteration. Each model (using the linear thickness SGR, disease volume SGR, or normalized lung volume SGR assessment metric) was trained on the training cohort then tested on the testing cohort. Therefore, for each sub-sample iteration, model performance statistics were tracked in a paired fashion, and differences between models were assessed using the histogram of paired differences between testing cohort performance values. Models were considered significantly different if the 95% central confidence interval (CI) of sub-sample paired differences did not include a difference of zero. All analyses were performed using the academic edition of Revolution R Enterprise (version 4.3, based on R version 2.12; Revolution Analytics, Palo Alto, CA) [28].

III. RESULTS

A. Patients and Overall Survival

Median survival from pre-treatment baseline imaging was 12.7 months (95% confidence interval, 10.2–15.3 months; range, 1.7–60 months). Of the 61 patients, there were 58 recorded deaths; the remaining three patients were censored after a median duration of 34 months. The Kaplan-Meier curve for overall survival is shown in Figure 1. Across all patients, the mean pleural disease volume at baseline was 1312 ± 853 mL (range 225–4449 mL). At the time of the first follow-up scan, the mean disease volume had reduced to 1232 mL, with geometric mean change from baseline of -11%. By the end of treatment, the geometric mean change in disease volume from baseline was -17%.

B. Lung Segmentation

Across all patients, the mean baseline ipsilateral lung volume was 1021 ± 574 mL, and the mean baseline contralateral lung volume was 2648 ± 639 mL. The mean normalized ipsilateral lung volume at baseline was 0.399 (range 0.058–1.262). By the first follow-up scan, the normalized ipsilateral lung volume had increased to 0.420, up by a geometric mean of 5% from baseline. By the end of treatment, the normalized ipsilateral lung volume had increased a geometric mean of 8% from baseline. Over the course of the entire treatment, the distinction between normalized ipsilateral lung volume increase and decrease was significantly associated with patient survival. Figure 2 shows the Kaplan-Meier survival curves for the two patient groups (log-rank $p = 0.0003$).

The extent of manual intervention necessary in the otherwise fully automated lung segmentation was minimal. For cases that required any intervention whatsoever (21% of all scans), the duration of manual intervention averaged approximately one minute. Only 1.9% of cases required five minutes or more of manual intervention. The predominant cause for manual editing of lung segmentations was erroneous inclusion of segmented bowel gas.

From the lung segmentation validation study (which did not allow manual intervention), there was very high agreement for area measurements of per-section lung segmentations between the manual approach and the automated method for the 132 axial sections evaluated. Pearson's correlation coefficient was calculated as 0.973 ($p < 0.0001$). Using Bland-Altman analysis (Figure 3), the mean bias indicated that automated lung area measurements were on average 1.17 cm^2 larger than manual measurements (or 1.1% larger given that the average section lung area was 102.03 cm^2). The 95% limits of agreement in the difference between manual measurements and automated measurements were -19.52–17.19 cm^2 , relatively small given the correlation and average measurement magnitude.

C. Linear and Volumetric Measurement Correlations

A plot comparing the relative change from baseline of linear thickness measurements and disease volumes for the 61 patients in this study is shown in Figure 4. Each of the 155 points on the plot represents a single paired change from baseline (i.e., if a patient has a baseline CT scan and three follow-up scans, there will be three data points comparing linear measurement change from baseline with volume measurement change from baseline for that patient). For these data, Spearman's rank correlation coefficient was estimated to be $\rho_{thickness} = 0.676$ and Pearson's linear correlation coefficient was estimated to be $r_{thickness} = 0.665$. Both correlations are positive, indicating that growth in disease linear thickness corresponds to growth in disease volume.

The relationship expected from a spherical geometric model (equation 2) is indicated in the plot with a dashed line. The quality of fit of the spherical model to the data is $R^2 = 0.35$.

Visual inspection of the plot indicates that the data do not reliably fall along the dashed line and instead appear nearly linear in some locations. While there was no theoretical reason to believe that mesothelioma would follow a spherical geometry (indeed, the shortcomings of the spherical model for this disease have already been investigated [11, 29]), Figure 4 provides the first empirical evidence for the inappropriateness of the spherical assumption implicit in the standard RECIST discretized response classification criteria for MPM.

A plot comparing the relative change from baseline of normalized ipsilateral lung volumes and disease volumes for the 61 patients in this study is shown in Figure 5. Again, each of the 155 points on the plot represents a single paired change from baseline. The non-parametric Spearman's rank correlation coefficient was estimated to be $\rho_{lung} = -0.687$. The linear trend correlation from Pearson's correlation coefficient was estimated to be $r_{lung} = -0.494$. The correlation coefficients are negative, indicating that for an increase in normalized lung volume, the disease volume decreases. Trajectories of the two measurement techniques for one particular patient are shown in Figure 6.

D. Survival Analysis

All three response assessment methods were significantly associated with patient survival in univariate Cox PH survival models. Increases in continuous time-varying linear thickness SGR measurements were associated with poor patient prognosis (HR=1.53, $p < 0.0001$), as were increases in disease volume SGR (HR=1.32, $p = 0.0003$) and decreases in normalized ipsilateral lung volume SGR (HR=0.76, $p = 0.003$).

In multivariate Cox PH survival models including disease histology, dyspnea, and ECOG performance status, all three response assessment methods remained significantly associated with patient survival. The model coefficients for the linear thickness model, disease volume model, and normalized lung volume model are shown in Table II. The hazard ratio estimates for the clinical covariates vary among the three multivariate survival models, but the variability is small compared with the 95% confidence intervals given in Table II.

Model performance was quantified using the C^{τ} statistic. The performance of the full multivariate model trained and tested on the same patient cohort was 0.692, 0.680, and 0.670 for the models using linear thickness measurements, disease volume measurements, and normalized lung volume measurements, respectively, along with the same clinical covariates. In the leave-oneout cross-validation, these scores were reduced slightly to 0.657, 0.625, and 0.630, respectively. Finally, the mean random sub-sample performance values for the three models were 0.659, 0.638, and 0.628, respectively. These values are summarized in Table III.

Paired differences in sub-sample testing cohort performance values between survival models incorporating the different response assessment methods were used to compare the utility of the different response metrics. The mean difference in paired C^{τ} performance values between the linear thickness model and the disease volume model was 0.022, with a 95% confidence interval of -0.077–0.123 and was therefore not significant (bootstrap $p = 0.30$). The mean difference in paired C^{τ} performance values between the normalized ipsilateral lung volume model and the disease volume model was -0.009, with a 95% confidence interval of -0.087–0.077, and was therefore not significant (bootstrap $p = 0.65$). The performance of the linear thickness model is on average 3.4% higher than the performance of the disease volume model; however, considerable overlap exists in the performance of the two models (the disease volume model outperformed the linear thickness model for 30% of the random sub-sample iterations). The performance of the normalized ipsilateral lung volume model is on average 1.4% lower than the performance of the disease volume model,

and even more overlap exists between the lung and disease volume models than between the linear thickness and disease volume models.

IV. DISCUSSION

In a previous study [21], it was shown for the first time that continuous and time-varying image-based measurements of pleural disease volume were significantly associated with patient survival in mesothelioma. The present study extends this previous investigation to three tumor response assessment methods: linear thickness measurements acquired using the modified RECIST protocol, semi-automated segmentations of pleural disease volume, and automated segmentations of normalized ipsilateral lung volume. These three response assessment methods are all significantly associated with patient survival, and there are no significant differences between models that incorporate the different response metrics. Practical differences, however, exist among the three measurement techniques and the resulting models.

Until recently, measurements of complete pleural disease volume for patients with MPM were time prohibitive, and linear thickness measurements remain the clinical standard for response assessment. In the past few years, several software algorithms for the segmentation of mesothelioma on CT scans have been published [14, 16, 19], and researchers are now able to explore true disease volume as a response assessment method. The novel response assessment metric in this study is lung volume; lung segmentation is a comparatively easier computational task than pleural disease segmentation, and there is reason to expect lung volumes to be generally correlated anatomically to disease volumes for patients with MPM. Although some gross anatomic changes to the affected hemithorax are possible in mesothelioma, a decrease in disease volume should result in a corresponding increase in the ipsilateral lung volume. Normalizing the ipsilateral lung volume by the contralateral lung volume corrects for differences in respiratory phase between a patient's CT scans, and changes in normalized lung volume form a useful response assessment metric.

The correlations among the three response assessment metrics reported in this study were in line with expectations. One would expect changes in linear thickness to be correlated with changes in disease volume, as was shown; however, the spherical geometric relationship between tumor thickness and tumor volume implicit in the RECIST protocol does not hold in mesothelioma, as evidenced in Figure 4. The correlation between normalized ipsilateral lung volumes and disease volumes was also as expected, since decreases in disease volume were met by increases in normalized ipsilateral lung volume. An example of this correlation is shown in Figure 6, where changes in normalized ipsilateral lung volume and changes in disease volume are seen to closely mirror one another. Because of the high correlation among the three metrics, using more than one response assessment metric in the same Cox PH model results in at least one of the metrics becoming a non-significant covariate (usually with a p -value larger than 0.20). Therefore, no more than one response assessment method at a time can be an independent significant covariate for patient prognosis.

The fact that the survival model with linear thickness measurements outperformed (although not at a significant level) the disease volume survival model was unexpected. Disease volumes are logically better able to capture changes in overall tumor bulk, but perhaps changes in tumor thickness are physiologically more predictive of eventual patient survival than overall volumetric changes. The two response assessment methods provide different information, and while it was previously assumed that disease volumes should be the ultimate goal of any response assessment technique, it is possible that the specific type of morphological change quantified by tumor thickness measurements is more representative of patient benefit. Another possibility is that human observers are able to place their baseline

tumor thickness measurements in locations that are in some sense more relevant for response assessment; volume measurements capture changes over the total extent of disease, while tumor thickness measurements only capture change in the discrete (up to six, by modified RECIST) locations at which baseline measurements were placed. Manual linear thickness measurements are often placed in areas of distinct tumor presence, whereas the disease volume measurements may incorporate pleural fluid in some patients. It may be possible to improve the performance of the survival model using disease volume measurements if pleural fluid could be more reliably excluded.

Also interesting is the nearly identical performance of the survival models using disease volumes and normalized ipsilateral lung volumes. The similar performance of the two models reinforces the expectation that changes in (normalized) lung volume and disease volume should convey roughly equivalent information due to the physiological correlation between the two structures. The correlation between paired C^{τ} values from random sub-sample testing cohorts showed high correlation ($r=0.77$) between the survival models using disease volume and normalized ipsilateral lung volume.

There are various advantages and disadvantages for each response assessment method. It was shown by Frauenfelder et al. [14] that the inter-observer variability is substantially lower for disease volume measurements than for linear thickness measurements, a fact that could become an important consideration if disease volumes were to be used clinically to assess tumor response. However, linear measurements require less manual time than semi-automated disease volume measurements, and existing techniques could potentially be used to partially automate the linear measurement process and thereby reduce time and variability [30, 31]. Lung volume measurement is an automated process, and the only manual intervention used in this study was the correction of obvious segmentation errors from contrast artifacts and bowel gas. It is therefore reasonable to believe that lung volume measurements would have almost no inter-observer variability. However, the utility of lung volume measurements for tumor response assessment is limited to patients with unilateral disease, as well as those patients who do not have frequent changes in pleural fluid volume (such as with in-dwelling pleural catheters). While unilateral disease is most common, this stipulation necessarily precludes lung-volume-based response assessment for a small number of patients.

While talc pleurodesis causes the fusion of the pleural space, there is no evidence to suggest that the procedure would affect the image-based lung volume measurement process. Furthermore, among the patients who underwent talc pleurodesis, an average of 157 days elapsed between the procedure and study entry. One patient underwent talc pleurodesis while on study, although a span of 56 days elapsed between talc pleurodesis and the next CT scan. Although talc pleurodesis induces local inflammation, this effect will likely have more of an impact on PET-based measurements of metabolic activity than on CT disease burden or lung volume measurements.

An inherent limitation of this study is the relatively small number of patients evaluated. The survival models compared in this study form the starting point for a validation in independent patient cohorts and should not be taken as definitive response models. While all the survival models in this study had statistically significant prognostic discrimination, absolute performance scores of around 0.65 are by no means perfect. Although the survival model from the linear thickness measurements outperformed the other two models on average, there is no statistical basis to conclude that any one model is better than another. It should be further cautioned that the survival models in this study may not be applicable to patients who receive biologically different treatments than the cytotoxic therapy used for the patient cohort in this study.

In summary, this study compared survival models using three different tumor response assessment methods for patients with MPM undergoing chemotherapeutic treatment. Models were fit using clinical covariates identified in a previous study and either (1) linear thickness measurements, (2) pleural disease volume measurements, or (3) normalized ipsilateral lung volume measurements. As a novel tumor response assessment technique, lung volumes exhibited the expected correlation with disease volumes. All three response assessment methods were significantly associated with patient survival. The model using linear thickness measurements performed, on average, better than the other two models, though the differences were not significant.

Acknowledgments

The authors would like to acknowledge Philip Caligiuri, M.D., for providing the manual lung contours used in the validation of the automated lung segmentation algorithm.

Sources of Support:

This work was supported by The University of Chicago Comprehensive Cancer Center; the Raine Medical Research Foundation; the US National Institutes of Health [grant numbers T32EB002103, R01CA102085]; the Simmons Mesothelioma Foundation; the Kazan Law Firm's Charitable Foundation; the National Health and Medical Research Council, Australia; and the Cancer Council Western Australia.

REFERENCES

- Francis RJ, Byrne MJ, van der Schaaf AA, Boucek JA, Nowak AK, Phillips M, et al. 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.* 2007; 48:1449–1458. [PubMed: 17704250]
- Lee HY, Hyun SH, Lee KS, Kim BT, Kim J, Shim YM, et al. Volume-based parameter of (18)F-FDG PET/CT in malignant pleural mesothelioma: Prediction of therapeutic response and prognostic implications. *Ann Surg Oncol.* 2010; 17(10):2787–2794. [PubMed: 20461469]
- Nowak AK, Francis RJ, Phillips MJ, Millward MJ, van der Schaaf AA, Boucek JA, et al. A novel prognostic model for malignant mesothelioma incorporating quantitative FDG-PET imaging with clinical parameters. *Clin Cancer Res.* 2010; 16(8):2409–2417. [PubMed: 20371686]
- Jaffe CC. Measures of response: RECIST, WHO, and new alternatives. *J Clin Oncol.* 2006; 24:3245–3251. [PubMed: 16829648]
- Prasad SR, Jhaveri KS, Saini S, Hahn PF, Halpern EF, Sumner JE. CT tumor measurement for therapeutic response assessment: Comparison of unidimensional, bidimensional, and volumetric techniques—Initial observations. *Radiology.* 2002; 225(2):416–419. [PubMed: 12409574]
- Boone JM. Radiological interpretation 2020: Toward quantitative image assessment. *Med Phys.* 2007; 34(11):4173–4179. [PubMed: 18072481]
- Michaelis LC, Ratain MJ. Measuring response in a post-RECIST world: From black and white to shades of grey. *Nat Rev Cancer.* 2006; 6(5):409–414. [PubMed: 16633367]
- Mehrara E, Forssell-Aronsson E, Bernhardt P. Objective assessment of tumour response to therapy based on tumour growth kinetics. *Br J Cancer.* 2011; 105(5):682–686. [PubMed: 21792200]
- Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst.* 2000; 92(3):205–216. [PubMed: 10655437]
- Therasse P, Eisenhauer EA, Verweij J. RECIST revisited: A review of validation studies on tumour assessment. *Eur J Cancer.* 2006; 42(8):1031–1039. [PubMed: 16616487]
- Oxnard GR, Armato SG III, Kindler HL. Modeling of mesothelioma growth demonstrates weaknesses of current response criteria. *Lung Cancer.* 2006; 52(2):141–148. [PubMed: 16530882]
- Corson N, Sensakovic WF, Straus C, Starkey A, Armato SG III. Characterization of mesothelioma and tissues present in contrast-enhanced thoracic CT scans. *Med Phys.* 2011; 38(2):942–947. [PubMed: 21452730]

13. Byrne MJ, Nowak AK. Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. *Ann Oncol.* 2004; 15(2):257–260. [PubMed: 14760119]
14. Frauenfelder T, Tutic M, Weder W, Götti RP, Stahel RA, Seifert B, et al. Volumetry: An alternative to assess therapy response for malignant pleural mesothelioma? *Eur Respir J.* 2011; 38(1):162–168. [PubMed: 21273389]
15. Pass HI, Temeck BK, Kranda K, Steinberg SM, Feuerstein IR. Preoperative tumor volume is associated with outcome in malignant pleural mesothelioma. *J Thorac Cardiovasc Surg.* 1998; 115(2):310–317. [PubMed: 9475525]
16. Liu F, Zhao B, Krug LM, Ishill NM, Lim RC, Guo P, et al. Assessment of therapy responses and prediction of survival in malignant pleural mesothelioma through computer-aided volumetric measurement on computed tomography scans. *J Thorac Oncol.* 2010; 5(6):879–884. [PubMed: 20421814]
17. Sensakovic WF, Armato SG III, Starkey A, Kindler HL, Vigneswaran WT. Quantitative measurement of lung reexpansion in malignant pleural mesothelioma patients undergoing pleurectomy/decortication. *Acad Radiol.* 2011; 18(3):294–298. [PubMed: 21145765]
18. Armato SG III, Oxnard GR, Kocherginsky M, Vogelzang NJ, Kindler HL, MacMahon H. Evaluation of semiautomated measurements of mesothelioma tumor thickness on CT scans. *Acad Radiol.* 2005; 12(10):1301–1309. [PubMed: 16179207]
19. Sensakovic WF, Armato SG III, Straus C, Roberts RY, Caligiuri P, Starkey A, et al. Computerized segmentation and measurement of malignant pleural mesothelioma. *Med Phys.* 2011; 38(1):238–244. [PubMed: 21361192]
20. Armato SG III, Sensakovic WF. Automated lung segmentation for thoracic CT: Impact on computer-aided diagnosis. *Acad Radiol.* 2004; 11(9):1011–1021. [PubMed: 15350582]
21. Labby ZE, Nowak AK, Dignam JJ, Straus C, Kindler HL, Armato SG III. Disease volumes as a marker for patient response in malignant pleural mesothelioma. *Ann Oncol.* In Press.
22. Sensakovic WF, Starkey A, Roberts RY, Armato SG III. Discrete-space versus continuous-space lesion boundary and area definitions. *Med Phys.* 2008; 35(9):4070–4078. [PubMed: 18841859]
23. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986; 327(8476):307–310. [PubMed: 2868172]
24. Klein, JP.; Moeschberger, ML. *Survival Analysis: Techniques for Censored and Truncated Data.* 2nd ed.. Springer; New York, NY: 2010.
25. Cox DR. Regression models and life tables. *J R Stat Soc B.* 1972; 34(2):187–220.
26. Zhou M. Understanding the Cox regression models with time-change covariates. *Am Stat.* 2001; 55:153–155.
27. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics.* 2005; 61(1):92–105. [PubMed: 15737082]
28. R Development Core Team. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: 2011.
29. Labby ZE, Armato SG III, Kindler HL, Dignam JJ, Hasani AA, Nowak AK. Optimization of response classification criteria for patients with malignant pleural mesothelioma. *J Thorac Oncol.* 2012; 7(11):1728–1734. [PubMed: 23059782]
30. Armato SG III, Oxnard GR, MacMahon H, Vogelzang NJ, Kindler HL, Kocherginsky M, et al. Measurement of mesothelioma on thoracic CT scans: A comparison of manual and computer-assisted techniques. *Med Phys.* 2004; 31(5):1105–1115. [PubMed: 15191298]
31. Armato SG III, Ogarek JL, Starkey A, Vogelzang NJ, Kindler HL, Kocherginsky M, et al. Variability in mesothelioma tumor response classification. *Am J Roentgenol.* 2006; 186(4):1000–1006. [PubMed: 16554570]

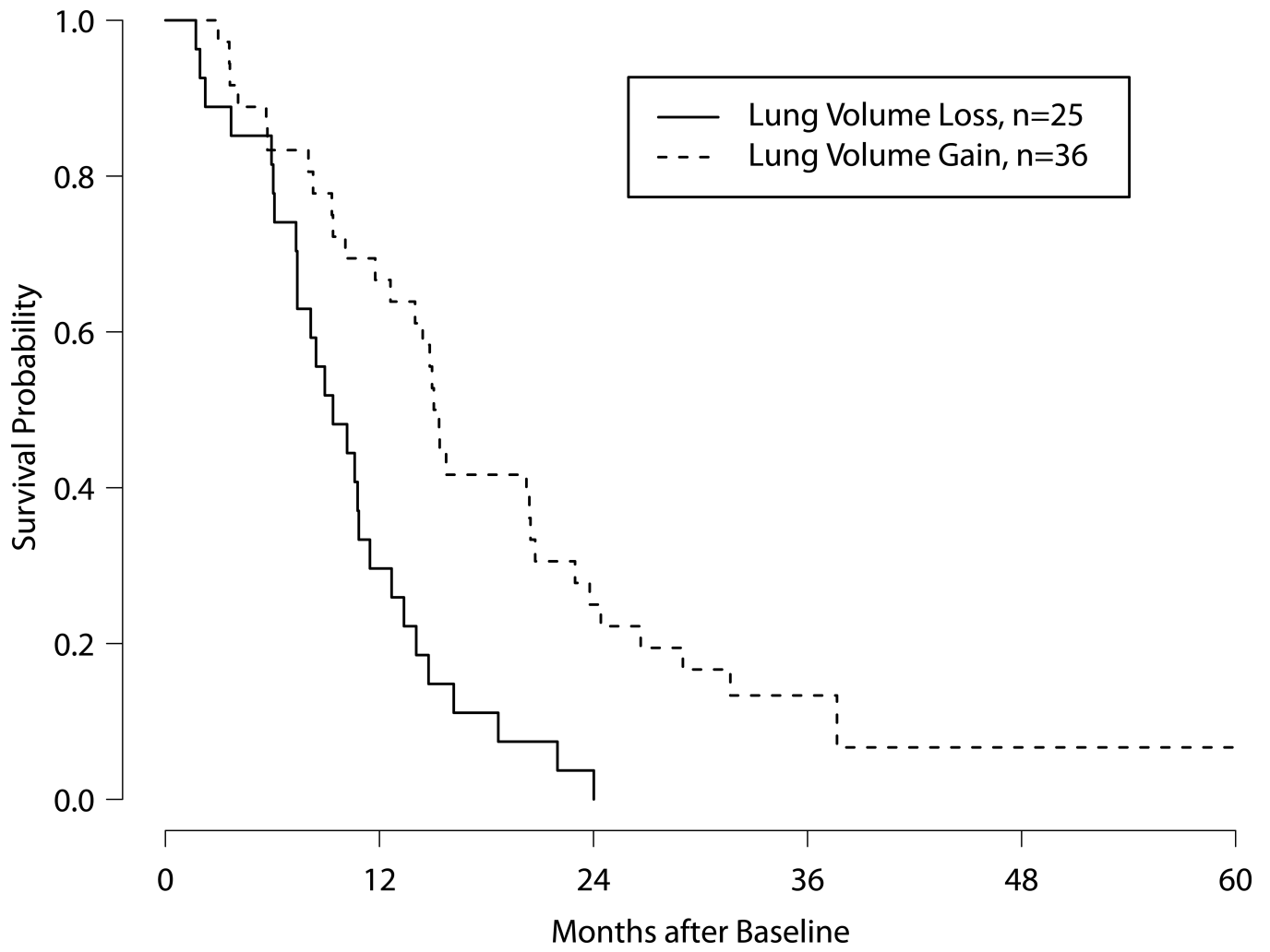
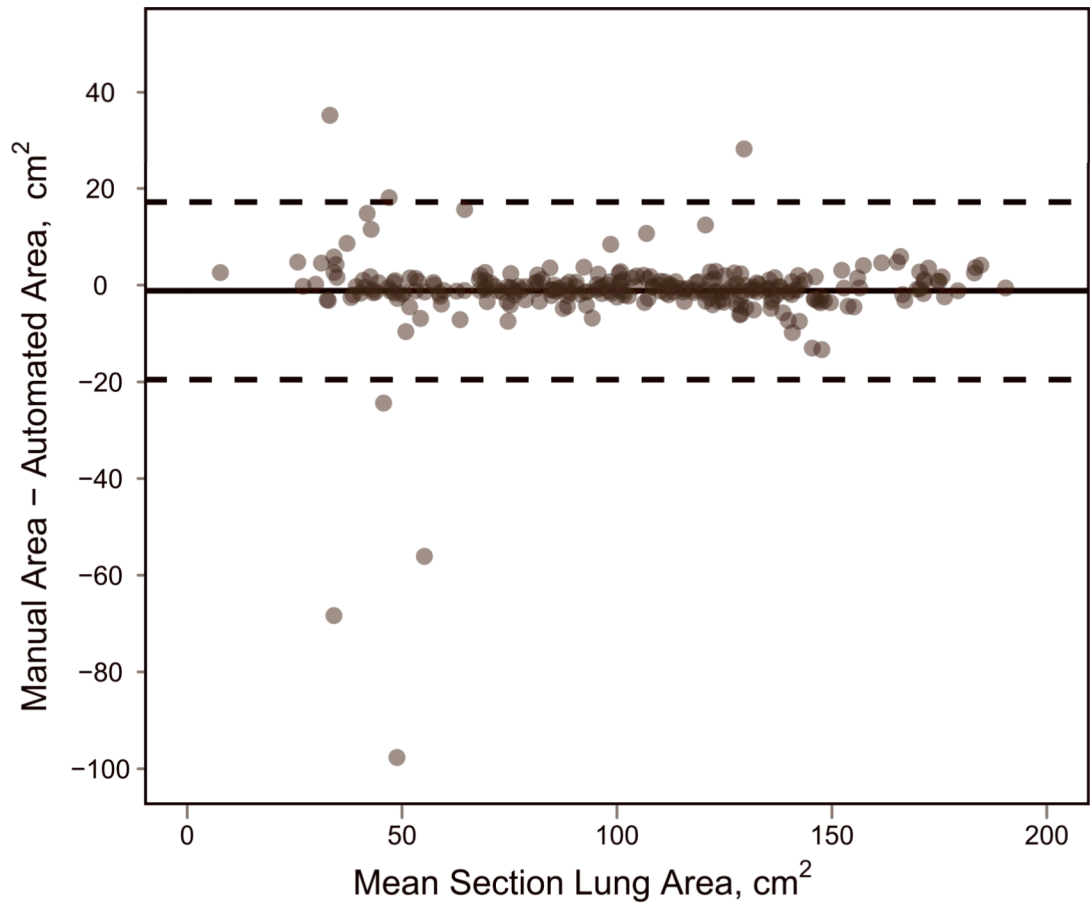


Figure 1. Kaplan-Meier survival curves for patients with and without normalized ipsilateral lung volume increase during the course of their therapy.



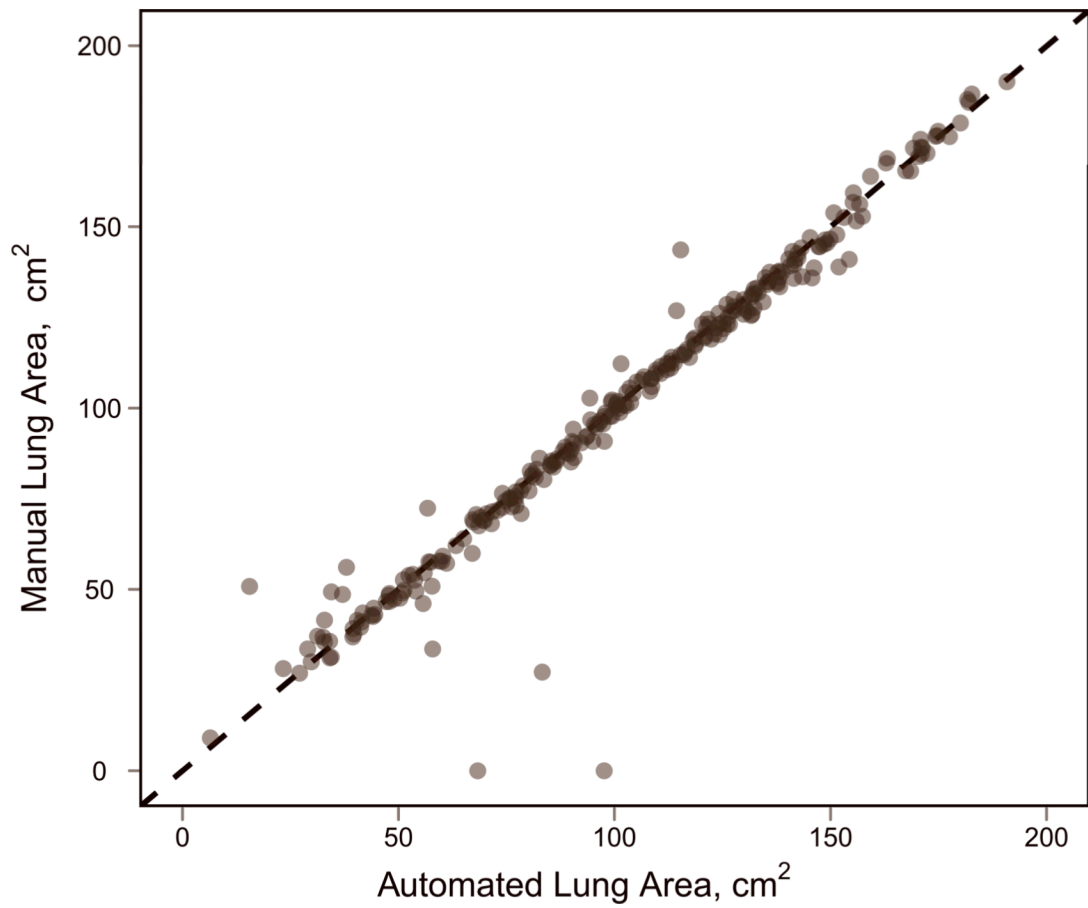


Figure 2. Validation of automated lung segmentation. (a) Bland-Altman plot, where bias is shown with a solid black line and the 95% limits of agreement are shown with dashed black lines. (b) Direct comparison between measurements, with the identity line shown.

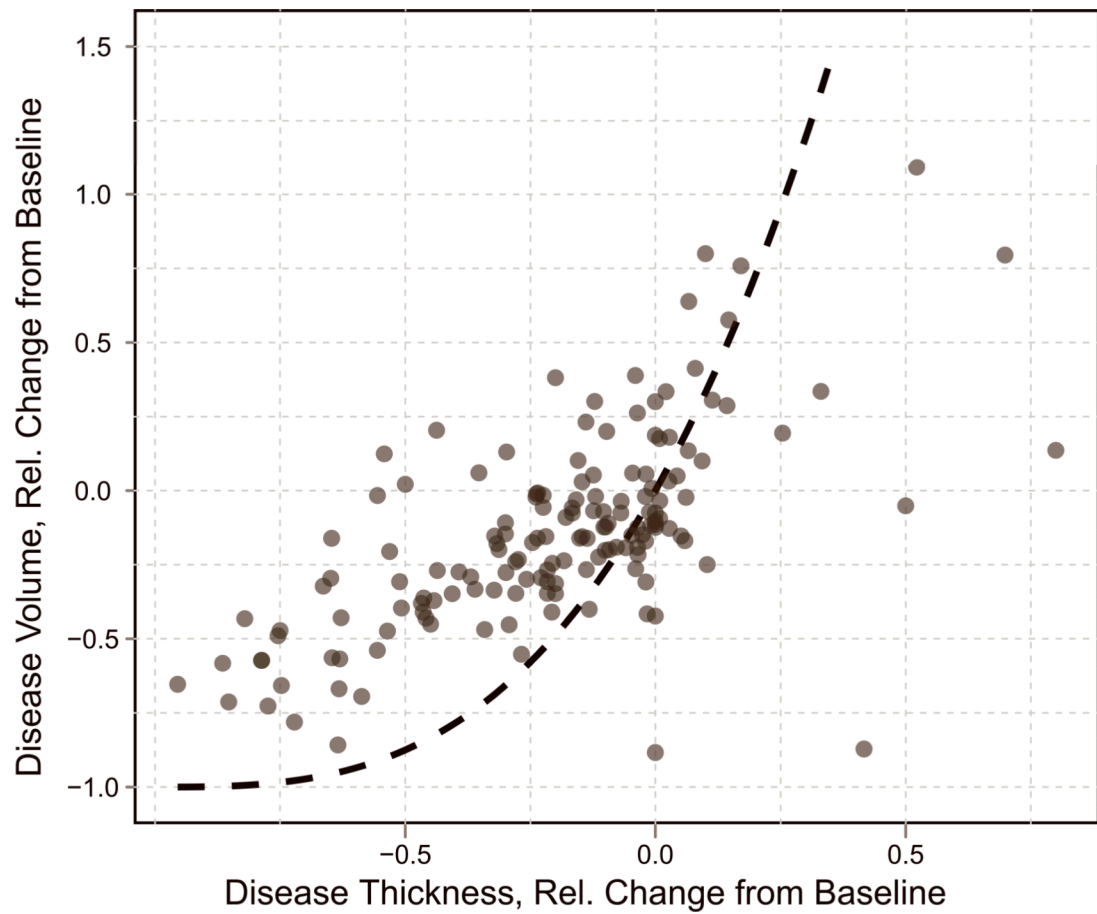


Figure 3. Relative change from baseline of disease volumes versus relative change from baseline of linear thickness measurements. The relationship expected from a spherical geometric model is indicated with a dashed black line.

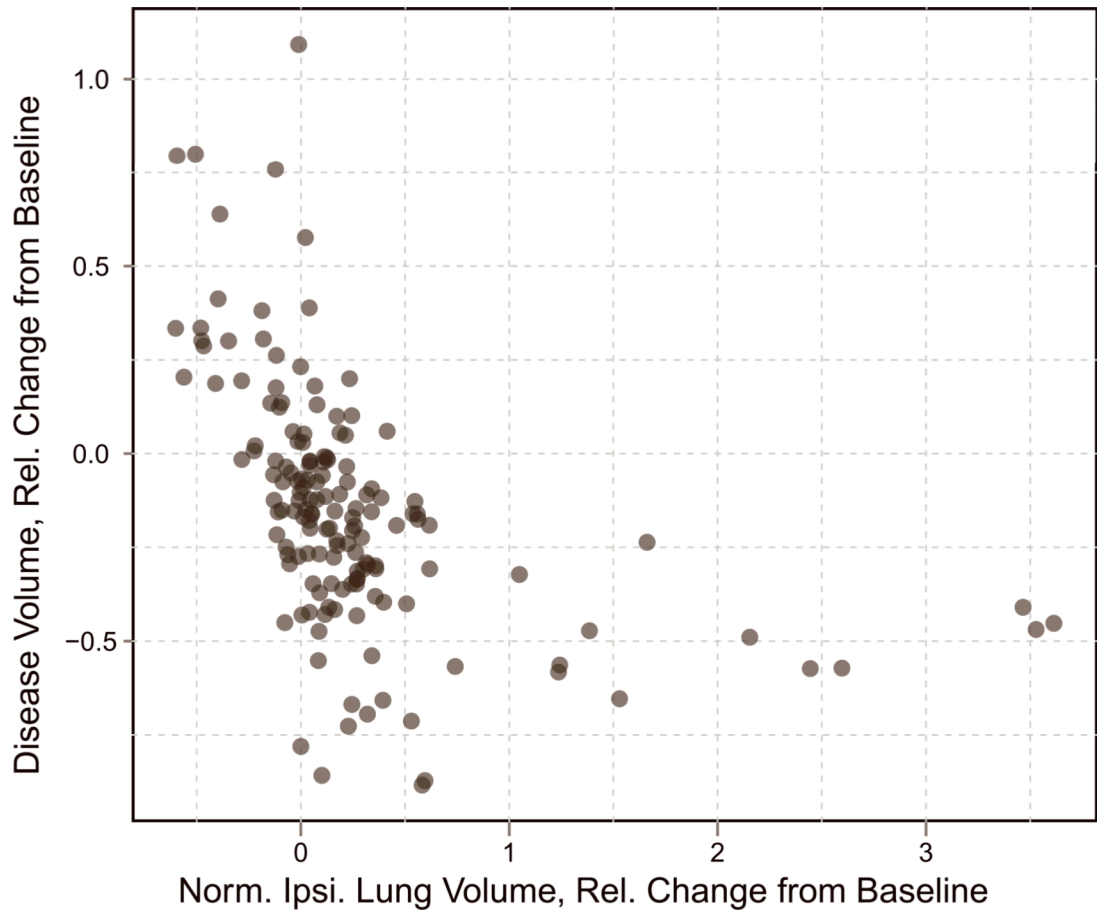


Figure 4. Relative change from baseline of disease volumes versus relative change from baseline of normalized ipsilateral lung volumes.

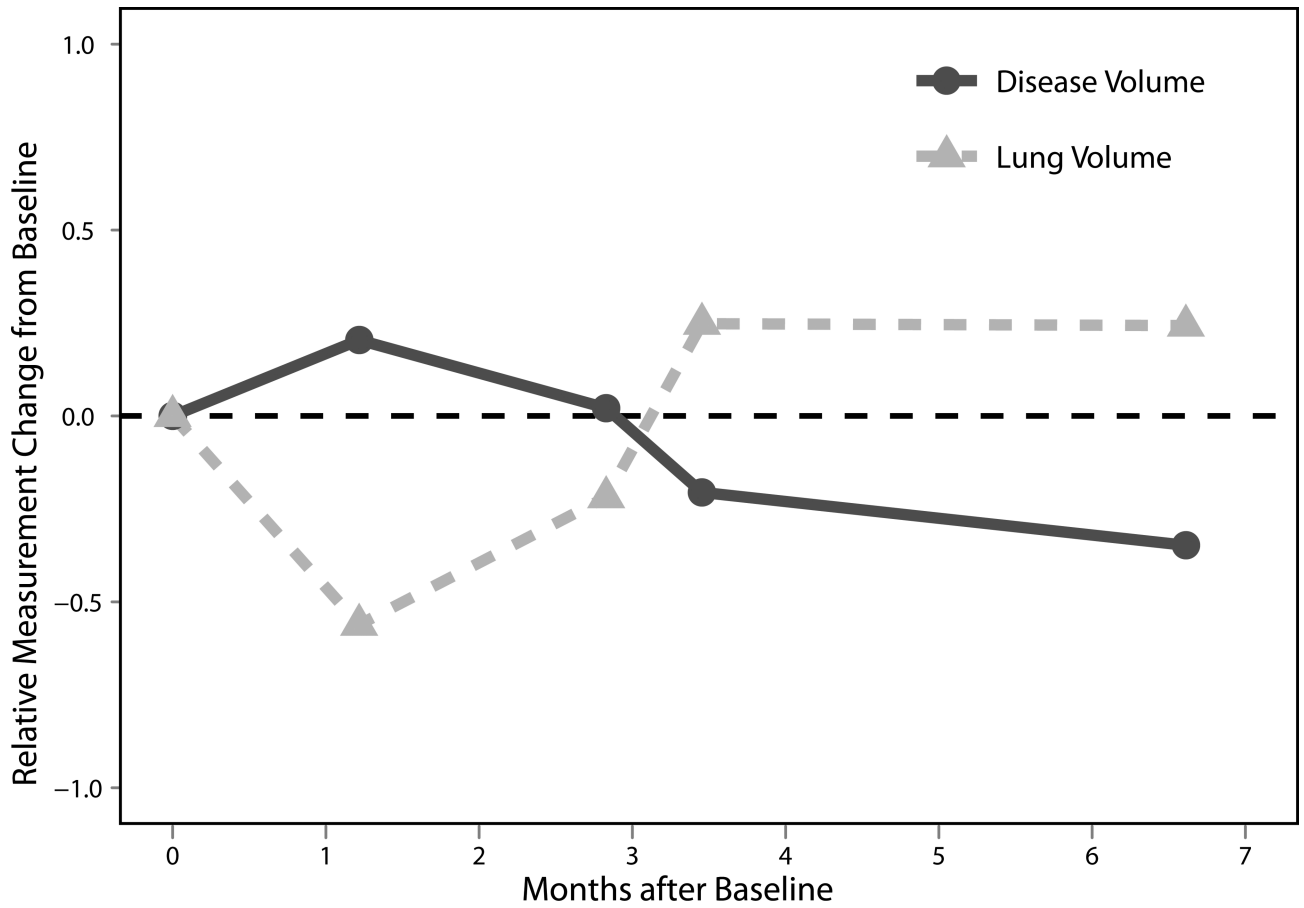


Figure 5. Relative changes from baseline scan of normalized ipsilateral lung volume and pleural disease volume for an example patient. Note the (anti-)correlation between the two curves.

Table I

Description of the patient cohort used in this study.

Sex		
	Male	n = 50
	Female	n = 11
Age at Diagnosis		
	Median	66 years
	Range	42–80 years
Chemotherapy		
	Carboplatin/Pemetrexed	n = 6
	Cisplatin/Pemetrexed	n = 31
	Cisplatin/Gemcitabine	n = 24
Histology		
	Epithelioid	n = 43
	Sarcomatoid	n = 5
	Biphasic	n = 13
T Stage		
	T1	n = 13
	T2	n = 16
	T3	n = 20
	T4	n = 12
N Stage		
	N0	n = 17
	N1	n = 2
	N2	n = 32
	N3	n = 10
M Stage		
	M0	n = 55
	M1	n = 6
IMIG Stage		
	I	n = 9
	II	n = 2
	III	n = 29
	IV	n = 21
Known Asbestos Exposure		
	Yes	n = 55
	No	n = 6
Chest Pain		

	Yes	n = 38
	No	n = 23
<hr/>		
Shortness of Breath		
	Yes	n = 50
	No	n = 11
<hr/>		
ECOG Performance Status		
	0	n = 31
	1	n = 26
	2	n = 4
<hr/>		
Talc Pleurodesis		
	Yes	n = 27
	No	n = 34
<hr/>		
Weight		
	Median	75 kg
	Range	52–121 kg
<hr/>		
Height		
	Median	171cm
	Range	155–188 cm
<hr/>		
Smoking Status		
	Never	n = 27
	Past	n = 29
	Present	n = 5
<hr/>		
Pleurectomy/Decortication		
	Yes	n=0
	No	n = 61
<hr/>		

Table II

Multivariate Cox PH model, including hazard ratios and 95% confidence intervals (CI). All tumor response assessment metrics were modeled as continuous specific growth rate (SGR) from baseline.

Variable		Hazard Ratio	95% CI	p-value
Linear Thickness Measurement Model				
Linear Thickness	(continuous, SGR)	1.47	[1.18, 1.84]	0.00053
Histology	Epithelioid	1	-	-
	Other	1.95	[1.05, 3.65]	0.036
Dyspnea	No	1	-	-
	Yes	2.46	[1.09, 5.55]	0.030
ECOG Performance Status	0	1	-	-
	1 or 2	1.47	[0.83, 2.61]	0.099
Disease Volume Measurement Model				
Disease Volume	(continuous, SGR)	1.33	[1.13, 1.58]	0.00090
Histology	Epithelioid	1	-	-
	Other	2.04	[1.10, 3.79]	0.023
Dyspnea	No	1	-	-
	Yes	2.81	[1.19, 6.61]	0.018
ECOG Performance Status	0	1	-	-
	1 or 2	1.54	[0.89, 2.67]	0.12
Normalized Lung Volume Measurement Model				
Normalized Lung Volume	(continuous, SGR)	0.76	[0.64, 0.91]	0.0033
Histology	Epithelioid	1	-	-
	Other	2.38	[1.30, 4.34]	0.0050
Dyspnea	No	1	-	-
	Yes	2.15	[0.98, 4.74]	0.056
ECOG Performance Status	0	1	-	-
	1 or 2	1.58	[0.92, 2.73]	0.099

Table III

Performance value (C^T) summary for multivariate survival models from Table II. Performance values are given for full models trained and tested on the complete cohort, from leave-one-out cross-validations, and from repeated random sub-sample simulations.

	Full Performance	LOOCV Performance	Mean Random Sub-Sample Performance	95% Random Sub-Sample Confidence Interval
Linear Thickness	0.692	0.657	0.659	[0.556, 0.760]
Disease Volume	0.680	0.625	0.638	[0.526, 0.755]
Normalized Lung Volume	0.670	0.630	0.628	[0.510, 0.744]