Measurement Variability in Treatment Response Determination for Non-Small Cell Lung Cancer: Improvements using Radiomics

Journal of Thoracic Imaging

Measurement Variability in Treatment Response Determination for Non-Small Cell Lung Cancer: Improvements using Radiomics

--Manuscript Draft--

Manuscript Number:	JTI-18-113R2
Full Title:	Measurement Variability in Treatment Response Determination for Non-Small Cell Lung Cancer: Improvements using Radiomics
Article Type:	Review Article (Invited)
Section/Category:	Pulmonary/thoracic
Keywords:	Molecular Targeted Therapy; Immunotherapy; Medical oncology; Phenotype; Lung cancer response; Image; Radiomics
Corresponding Author:	Ho Yun Lee Sung Kyun Kwan University School of Medicine at Samsung Medical Center Cancer Center Seoul, Seoul KOREA, DEMOCRATIC PEOPLE'S REPUBLIC OF
Corresponding Author's Institution:	Sung Kyun Kwan University School of Medicine at Samsung Medical Center Cancer Center
Order of Authors:	Geewon Lee
	So Hyeon Bak
	Ho Yun Lee
	Joon Young Choi
	Hyunjin Park
	Seung-Hak Lee
	Yoshiharu Ohno
	Mizuki Nishino
	Edwin J.R. van Beek
	Kyung Soo Lee
Manuscript Region of Origin:	KOREA, REPUBLIC OF
Abstract:	Multimodality imaging measurements of treatment response are critical for clinical practice, oncology trials and the evaluation of new treatment modalities. The current standard for determining treatment response in non-small cell lung cancer (NSCLC) is based on tumor size using the RECIST criteria. Molecular targeted agents and immunotherapies often cause morphological change without reduction of tumor size. Therefore, it is difficult to evaluate therapeutic response by conventional methods. Radiomics is the study of cancer imaging features that are extracted using machine learning and other semantic features. This method can provide comprehensive information on tumor phenotypes and can be used to assess therapeutic response in this new age of immunotherapy. Delta radiomics, which evaluates the longitudinal changes in radiomics features, shows potential in gauging treatment response in NSCLC. It is well known that quantitative measurement methods may be subject to substantial variability due to differences in technical factors and require standardization. In this review, we describe measurement variability in the evaluation of non-small cell lung cancer and the emerging role of radiomics.

August 13th, 2018

Journal of Thoracic Imaging Editorial Office

Dear Dr. Mark L. Schiebler:

Thank you for your considerations for our manuscript entitled “Measurement Variability in Treatment Response Determination for Non-Small Cell Lung Cancer: Improvements using Radiomics” and for giving us a chance to revise the manuscript. We tried to do our best in revising the manuscript by accepting your remarkable comments thus in enhancing the quality of our study.
In the revised version of the manuscript, we answered to your queries by assigning a number to each query and by answering to each question and thus each number.
We upload a revised version of the manuscript and corresponding figures to figure captions. Our specific responses are as follows:

Memo 1. I do not understand this sentence. Do you mean to say that diameter measurement varied by 11% but the volume measurements did not vary at all? Please re-write to better highlight what you mean.
→ Done. We have changed the sentence as follows:
According to a recent review article, although limit of agreement for both manual diameter measurements and semi-automated volume measurements lies in the same range in terms of absolute percentages, the percentage of lung nodules in which an actual inter-reader difference found was with 11% far lower for semi-automated nodule volume measurements compared to manual diameter measurements, where inter-reader variability occurs commonly.

Memo 2. Why wouldn’t this be more information needed to fill a 2D histogram than a 1D histogram?
→ We have changed the sentence a little bit for clarification. In general, the samples in a given bin for a 2D histogram is less than those in an equivalent 1D histogram as the voxels need to fill the bins spanning the 2D histogram compared to filling the 1D histogram.
In a hypothetical extreme example, a 5,000 voxel ROI can fill a uniform 10 bin 1D histogram with 500 counts in each bin. The same ROI would fill a uniform 10x10 bin 2D histogram with 50 counts in each bin.

Memo 3. Trim to 80 references.
→ We have reduced 52 references, thus, the total number of 144 references to 92 references. We have excluded simple references and kept major, repeating references. We hope this is acceptable for publication in the journal.

Thank you for your attention. We look forward to listening to your favorable review results soon.

With best wishes,

The Authors

Uploaded:
A copy of manuscript and matched figures

Measurement Variability in Treatment Response Determination for Non-Small Cell Lung Cancer: Improvements using Radiomics

Geewon Lee*[1,2], So Hyeon Bak*[1,3], Ho Yun Lee[1], Joon Young Choi[4], Hyunjin Park[5,6], Seung-Hak Lee[7], Yoshiharu Ohno[8, 9], Mizuki Nishino[10], Edwin J.R. van Beek[11], Kyung Soo Lee[1]

*Both authors contributed equally to this work.

[1]Department of Radiology and Center for Imaging Science, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

[2]Department of Radiology and Medical Research Institute, Pusan National University Hospital, Pusan National University School of Medicine, Busan, Korea

[3]Department of Radiology, Kangwon National University Hospital, Kangwon National University School of Medicine, Chuncheon, Korea

[4]Departments of Nuclear Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

[5]School of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon, Korea

[6]Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon, Korea

[7]Department of Electronic Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Korea

[8]Division of Functional and Diagnostic Imaging Research, Department of Radiology, Kobe University Graduate School of Medicine, Kobe, Hyogo, Japan

[9]Advanced Biomedical Imaging Research Center, Kobe University Graduate School of Medicine, Kobe, Hyogo, Japan

[10]Department of Radiology, Dana-Farber Cancer Institute and Brigham a nd Women’s Hospital, Boston, MA, USA.

[11]Edinburgh Imaging, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, U.K.

Correspondence to:

Prof. Ho Yun Lee

Department of Radiology

Samsung Medical Center, Sungkyunkwan University School of Medicine

50, Ilwon-Dong, Gangnam-Gu, Seoul 135-710, Korea

Tel) 822-3410-2502

Fax) 822-3410-0049

Email) hoyunlee96@gmail.com

Financial support: This research was supported by the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, which was funded by the Ministry of Health & Welfare (HI17C0086) and by a National Research Foundation of Korea grant funded by the Korean government (Ministry of Science, ICT, & Future Planning) (No. NRF-2016R1A2B4013046 and NRF-2017M2A2A7A02018568).

Financial/nonfinancial disclosures:

Nishino is consultant to Toshiba Medical Systems, WorldCare Clinical, Daiichi Sankyo; Research grant to the institution from Merck Investigator Studies Program, Toshiba Medical Systems, AstraZeneca; Honorarium

from Bayer, Roche

EvB is owner/founder of Quantitative Clinical Trials Imaging Services Ltd.

Dr. Yoshiharu Ohno has a research grant from Canon Medical Systems Corporation.

~~Impact of~~ Measurement Variability in Treatment Response Determination for Non-Small Cell Lung Cancer: ~~Fruitful Attention to~~Improvements using Radiomic~~s~~ ~~Approach~~

Abstract

Multimodality imaging measurements of treatment response are critical for clinical practice, oncology trials and the evaluation of new treatment modalities. The current standard ~~method in the setting~~for determining treatment response in~~of~~ non-small cell lung cancer (NSCLC) is based on tumor size using the RECIST criteria. Molecular targeted agents and immunotherapies often cause morphological change without reduction of tumor size. Therefore, it is difficult to evaluate therapeutic response by conventional methods. Radiomics is the study of cancer imaging features that are extracted using machine learning and other semantic features. This method can provide comprehensive information on tumor phenotypes and can be used to assess therapeutic response in this new age of immunotherapy. ~~Furthermore, although in its early steps,~~ D~~d~~elta radiomics, which evaluates the longitudinal changes in radiomics features~~, between interval studies may~~ show~~s~~ potential in ~~oncology studies~~gauging treatment response in NSCLC. It is well known that quantitative measurement methods may be subject to substantial variability due to differences in technical factors and require standardization. In this review, we describe measurement variability in the evaluation of non-small cell lung cancer and the emerging role of radiomics.

Key words: Molecular Targeted Therapy; Immunotherapy; Medical oncology; Phenotype; Lung cancer response; Image; Radiomics

Abbreviations:

ADC : apparent diffusion coefficient

AIF : arterial input function

CT : computed tomography

DCE : dynamic contrast enhanced

DWI : diffusion-weighted

EES : extravascular extracellular space

EGFR : epidermal growth factor receptor

FDG : fluoro-deoxyglucose

GGO : ground-glass opacity

GLCM : gray-level co-occurrence matrix

GRE : gradient-recalled echo

ISZM : intensity size zone matrix

LoG : Laplacian of Gaussian

MRI : magnetic resonance imaging

MTT : mean transit time

MTV : metabolic tumor volume

NEMA : National Electrical Manufacturers Association

NSCLC : non-small cell lung cancer

PBF : pulmonary blood flow

PBV :pulmonary blood volume

PET : positron emission tomography

RECIST : Response Evaluation Criteria In Solid Tumors

ROI : region of interest

RQS : radiomics quality score

SNR : signal-to-noise ratio

STIR : Short T1 inversion recovery

SUV : standardized uptake value

SUV_{bw} : normalization of standardized uptake value for patient body weight

SUV_{bsa} : normalization of standardized uptake value for body surface area

SUV_{lbm} : normalization of standardized uptake value for lean body mass

TLG : total lesion glycolysis

TSE : turbo spin-echo

UTE : ultrashort echo time

VOI : volume of interest

1 Introduction

Assessment of anti-tumor activity of cancer therapies is generally determined by an anatomical measurement of tumor burden. Since its first introduction in 2000 and subsequent revision in 2009, Response Evaluation Criteria In Solid Tumors (RECIST) has served as the reference standard for measuring tumor burden and confirming tumor response.[1,2] According to RECIST criteria, measurement is the maximal axial (in-plane) unidimensional measurement of a tumor's diameter. However, this conventional tumor size analysis is imperfect due to inter and intra reader measurement variability, heterogeneous tumor morphology, and different technical parameters at the time of scanning.[2,3] All of these factors contribute to measurement variability, which can lead to an erroneous determination of treatment response/progression and thereby misinform inappropriate treatment decisons.[2,3,4,5]

During the past decade, due to an improved understanding of cancer biology, a vast collection of targeted molecular therapies have been developed. This has led to a paradigm shift in the local and systemic treatment of non-small cell lung cancer (NSCLC). While conventional chemotherapy is focused on destroying rapidly dividing tumor cells, molecular targeted therapy aims at transmembraneous receptors and intracellular molecules that are responsible for the survival and proliferation of tumor cells. Molecular targeted therapy has been shown to be effective in tumors with specific genomic driver mutations. This, has which opened a new era of tumor response evaluation, where the limitations of RECIST-based approaches are increasingly being found noted.[6,4] Cancer immunotherapy with immune-checkpoint blockade attempts for utilization of the helps to activate the cancer pateints own cellular immune system to kill fight against the cancer cells.[7,8,5] It has become increasingly known Cancer immunotherapy is associated with unconventional response patterns that are not may not be accurately characterized by conventional response criteria such as the RECIST criteria.[7,9,10,4,5] In addition, morphological changes such as tumor necrosis and cavitation without concurrent tumor size reduction are frequently observed in the setting of anti-angiogenic therapy.[11-13,2,4]

Formatted: Superscript

Field Code Changed

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

~~Unfortunately, RECIST criteria do not reflect these changes in the post treatment morphology of the target lesion.~~

Radiomics is the process of extracting large amounts of advanced quantitative information embedded within radiological images. This approach to image analysis aids the field of oncology by providing a more quantitative approach for tumor response assessment.[6,14] Although state-of-the-art methods have been shown to work well for measuring tumor volume, a great deal of variability exists, and radiologists should be familiar with the technical variations, benefits, and drawbacks of radiomics regarding measurement variability. Furthermore, although computed tomography (CT) continues to play an important role, additional imaging modalities such as positron emission tomography (PET), magnetic resonance imaging (MRI) dynamic contrast enhanced (DCE) perfusion images and MRI diffusion-weighted (DWI) MRI allow for ~~the~~ multiparametric assessment of tumor biology (e.g. glucose metabolism, tumor perfusion, and tumor hypoxia) of tumor biology.[15,6] Thus, by combining detailed functional and metabolic information, these protocols ~~may~~ provide a more comprehensive depiction of the tumor microenvironment and may allow for an earlier determination of tumor response. *The purpose of the review is to focus on the technical issues regarding NSCLC tumor measurement variability and how radiomics is emerging for the early assessment of tumor response.*

Field Code Changed

Formatted: Superscript

Formatted: Font: Italic

2. Technical aspects of measurement

1) Measuring tumor volume

A. Segmentation

Precise NSCLC tumor measurement between interval studies is the current basis for tumor response assessment. ~~Although~~ Currently the long axis diameter, a unidimensional measurement, remains the standard RECIST criteria for assessment of

whether a tumor is growing or shrinking, discordant tumor response between primary reviewers and secondary reviewers has been reported.[3,4] In a previous study, ~~according to~~using RECIST for ~~lung~~ NSCLC~~cancers~~, there was a significant difference between~~among~~ readers for unidimensional measurements of tumor size. ~~Probability of~~The misclassification rates for progressive disease were 30% and 10% for interobserver and intraobserver measurements, respectively.[3,4] The primary reason for this variability is related to inter-reader differences in the manual measurements of the primary tumor.[3,7,4,16] For instance, in the case of a single unidimensional largest diameter measurement for RECIST, each reviewer may measure the tumor at different image slices. As a solution to this problem, most radiologists now agree that measuring the entire tumor volume is more accurate than a single unidimensional RECIST measurement.[~~17-22~~,8,9] Many recent publications have shown that volumetric measurements demonstrate better reproducibility and repeatability. According to a recent review article, although limit of agreement for both manual diameter measurements and semi-automated volume measurements lies in the same range in terms of absolute percentages, the percentage of lung nodules in which an actual inter-reader difference found was with 11% far lower for semi-automated nodule volume measurements compared to manual diameter measurements, where inter-reader variability occurs commonly. ~~Due to more common inter-reader variability of manual diameter measurements, the reported percentage of lung nodules in which an actual inter-reader difference found was 11% far lower for semi-automated nodule volume measurements compared to manual diameter measurements.~~[10,23] Second, volumetric measurement is more sensitive in detecting even small changes than is unidimensional measurement.[11,24] For example, in a 10 mm spherical nodule, a 1 mm increase of unidimensional diameter corresponds to a 10% increase in cross-sectional diameter and a 33% increase in volume.[12,25] Finally, as lung CT post-processing computer software is becoming widespread, tumor volumetric measurements are gaining in popularity. This has now become the standard for oncological trials are beginning as their clinical response endpoint[11,24]. We will now discuss the factors a~~e~~ffecting the variability of tumor volume measurement.[11,24,26,27]

Field Code Changed

Field Code Changed

Field Code Changed

Formatted: Superscript

Commented [mls1]: I do not understand this sentence. Do you mean to say that diameter measurement varied by 11% but the volume measurements did not vary at all? Please re-write to better highlight what you mean.

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Formatted: Superscript

Segmentation is the process by which humans (manual segmentation) and machines delineate tumor boundaries from the surrounding lung. Generally, the whole tumor is selected as the volume of interest (VOI), which is usually feasible, but in certain cases may be hampered due to indistinct tumor margins.[28,29,13] For example, when lung cancer is surrounded by a pathological abnormality such as post-obstructive pneumonia or radiation-induced lung injury, the tumor boundary is frequently obscured. Tissue reorganization and post radiation therapy, scar formation disrupts accurate tumor segmentation, leading to variability in tumor measurement.

Among various methods of segmentation, automatic and semi-automatic methods using volumetric software have been shown to be more reproducible than manual segmentation.[28,30,13] Although the current "gold standard" is considered to be manual segmentation drawn by experts, this method has major drawbacks: (1) it is a time-consuming, (2) labor-intensive task and (3) has inter and intra-reader variability. In a study comparing manual and semi-automatic segmentation, the radiomics features derived from the latter demonstrated significantly higher reproducibility ($p=0.0009$; intra-class correlation coefficient values of 0.85 and 0.77 for semi-automatic segmentation and manual segmentation, respectively) and were more robust compared to those derived from manual contouring.[14,31] When comparing repeatability (intra-algorithm comparisons) and reproducibility (inter-algorithm comparisons) of segmentation algorithms, repeatability was significantly higher than the reproducibility ($p<0.007$; average Dice score of 0.95 and 0.81 for repeatability and reproducibility, respectively), recommending that the same software be used at all time points in longitudinal studies.[15,32] However, in cases of part-solid adenocarcinomas, which have a ground-glass opacity (GGO) component, fully automatic segmentation is also be problematic due to the reduced contrast between the GGO component and surrounding lung parenchyma.[16,33] Thus, as of today, for part-solid adenocarcinomas, semi-automatic segmentation with tumor margin editing based on subjective decision by an experienced expert remains the optimal choice for accurate volumetric assessments of NSCLC (Figure

Formatted: Superscript

Formatted: Superscript

Field Code Changed

Field Code Changed

Field Code Changed

1).[34,35]17 Likewise, advanced NSCLC lung cancer patients with large tumors having irregular margins, ~~heterogenous~~heterogeneous intra-tumoral texture and surrounding atelectasis or effusions often require semi-automated approach with expert radiologist manual editing of the segmentation.[9]21 In the setting of molecular targeted therapy for NSCLC, tumor volumes obtained by a semi-automated approach have been shown to be a prognostic marker for improved ~~survival ,~~survival, solidifying its value in this era of precision medicine.[8,18]20,36

In terms of rapid and accurate tumor segmentation, fully automatic segmentation methods based on deep learning may be the solution. Several investigators have trained convolutional neural networks and demonstrated that deep learning is capable of performing accurate localization and segmentation of tumors in multiple organs.[37-40]19,20 Although most of these articles were based on MRI scans, such as brain, prostate, and rectum, deep learning technologies have shown potential to improve accuracy and robustness of tumor segmentation.

Another point that needs to be highlighted is the usage of different vendor volumetric software platforms. Studies comparing multiple volumetric software packages found considerable variation in nodule volume. This shows that that the results of software packages should not be used interchangeably.[41-44]21,22 Next we discuss the impact of technical factors at the time of CT acquisition and CT reconstruction such as radiation dose, iterative reconstruction, inspiration, and slice thickness on the variability in volumetric measurement.

2) Technical issues according to particular imaging modality

A. CT

Chest CT is the modality of choice in routine lung cancer imaging, and iterative

Formatted: Superscript

Field Code Changed

Field Code Changed

Formatted: Superscript

Formatted: Superscript

reconstruction techniques have allowed for a significant reduction of radiation dose with overall similar image quality (Figure 2).[2345] Quantitative analysis of CT data can provide accurate anatomical information about NSCLC and the surrounding lung. In principle, lung nodule volumetric measurement and comparison across interval CT scans is relatively easy and reproducible on the same scanner hardware. However, imagers and oncologists should keep in mind that there is variability in these "objective" CT metrics that are introduced by individual reviewer RECIST measurements, manual segmentation of tumor volume, and technical factors (e.g. choice of reconstruction kernel, slice-thickness, and inter-scanner differences)[46-48][24] Any combination of these, factors may cause considerable measurement variability of the tumor burden, making the task more challenging for radiologists. We next carefully discuss the various technical factors that may impact tumor measurement accuracy.

Field Code Changed

Formatted: Superscript

a. CT Reconstruction algorithms and radiation dose

Previous studies have investigated the influence of the reconstruction kernel and radiation dose on lung nodule volume using chest phantoms.[49-52][25,26] The vast majority of those studies demonstrated that various iterative reconstructions (e.g. adaptive statistical iterative reconstruction, iDose, and model based iterative reconstruction) showed no significant variability in nodule diameter or volume measurement when compared to filtered back projection.[49-52][25,26] In fact, some studies reported that iterative reconstructions demonstrated better measurement accuracy at a reduced radiation dose. They suggested that reduced noise or increased image quality from iterative reconstruction helped reduce measurement errors.[25,26][49,51,53,54] In a study comparing lung cancer screening individuals who underwent low dose CT and ultra-low dose CT with iterative reconstruction, there was no significant difference in nodule size and volume measurement between the two protocols.[27][55] In a recent study comparing subsolid nodules between model-based iterative reconstruction and filtered back

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Field Code Changed

projection, Cohen et al. demonstrated that semi-automatic measurements of diameter, volume, and solid components of the subsolid nodules were within the range of measurement variability.[28,56] Thus, lung nodule volumetric measurements acquired from scans with different reconstruction techniques can be reliably compared.

b. Slice thickness and reconstruction kernel

Prior studies have investigated the impact of slice thickness on tumor measurement for cancer screening or tumor response evaluation.[57–61,29,30] Significant differences in volume according to CT slice thickness variation were noted for smaller lung nodules, where thicker slices introduced greater measurement variability.[57,58,61,29,30] The reason for this is related to partial volume effects. Given that a thicker CT slice contains larger partial volume artifacts than a thinner image, the margin of the tumor is blurred on thicker images. The lack of isotropic voxels for Lung CT influences NSCLC nodule segmentation and any extracted radiomics features.[31,62] In cases of subcentimeter nodules, which have very small VOI, partial volume artifacts substantially influences the volume measurement.[12,25]

Similarly, when employing radiomics, recent studies have shown that thin-slice images were better than thick-slice images for radiomics features.[31,32,62,63] In patients with lung cancer, He et al. reported that a radiomics signature based on thin slices (1.25 mm) demonstrated better diagnostic performance than when applied to thick slices (5 mm).[32,63] In a chest phantom study, thinner (1.25 mm and 2.5 mm) slice thickness was found to be better for radiomics features (e.g. quantifying tumor size, shape, and density).[31,62] To minimize measurement variability thinner slice images are recommend, slice thickness should be the same and slices of different thickness should not be mixed together for analysis (Figure 3).

Studies comparing tumor volumes at different reconstruction kernels are scarce and have conflicting results. One study reported that, compared to sharp kernels, soft

Field Code Changed

Formatted: Superscript

Formatted: Superscript

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

tissue reconstructions demonstrated more repeatable volumetric measurements.[33,64] Another study reported that, compared to high frequency bone algorithms, low frequency soft algorithms demonstrated larger volumes.[43,65,34]

Field Code Changed

Formatted: Superscript

c. Effects of respiration and intravenous contrast

Differences in lung inflation should not be underestimated when measuring lung nodules. For example, collapse of the alveoli at expiration may bring over and underestimation of tumor size, whereas stretching of the tumor parenchyma and blurring of the tumor margins could be responsible for an apparent larger tumor size at inspiration. Interestingly, results of significant changes in apparent tumor volume during the respiratory cycle have been previously reported.[12,25] Furthermore, motion artefacts during respiration can significantly affect the ability to segment lung nodules, rendering their outline and volume assessment unreliable. In addition, the presence of a pleural effusion or pneumothorax may also have a large influence on the apparent tumor volume.[66] For radiomics, Oliver et al. suggested that approximately 75% of the current dictionary of CT radiomics features are susceptible to respiration.[35,67]

Field Code Changed

Field Code Changed

Another interesting point is the impact of intravenous contrast material on lung nodule volume. Due to increased attenuation of the peripheral portion of a nodule at post-contrast scans (more vascular and viable region of the tumor nodule), the contrast difference between the parenchyma and the nodule increases; thus, volume segmentation may include a greater area of the peripheral lung nodule.[22,43] Results from two studies showed that, although the precise increase in nodule volume was small, radiologists should be aware of this artifact on the contrast enhanced exams.[36,37,68,69]

Field Code Changed

Field Code Changed

B. MRI

Owing to the ability of MRI to gather multiparametric data from NSCLC, MRI may play an increasing role in categorizing the therapeutic response in lung cancer.[38,39,70,71]

Field Code Changed

MRI is more reproducible in the identification of NSCLC and has superior soft tissue contrast in comparison to CT.[72] MRI lung nodule volumes are smaller than CT lung nodule volumes due to the higher resolution of CT and the magnetic susceptibility of air surrounding.[73,74,40] Ideally, for quantitative analysis, MRI images should all have the same field of view and acquisition matrix, field strength, and slice thickness. Each one of these acquisition parameters which have a strong effect on signal-to-noise ratio.[39,71] However, the many choices for acquisition in Lung MRI complicates comparison between studies of the many features extracted from the images.[41,75] Standardization of Lung MRI protocols in the setting of gathering radiomic features from NSCLC will be very helpful.

a. Magnetic field strengths

As MR field strength increases, the signal-to-noise (SNR) increases. This increase in the SNR can be utilized to for an increase in the number of phase encoded steps for better spatial resolution and improved anatomical identification.[40,42,74,76] The use of higher MR field strength improves the ability to contour tumor masses and reduces the measurement variability.[40,74] However, as MR field strength increases, it is accompanied by Bo and B1 inhomogeneity, an increased number of image artifacts due to changes in tissue magnetic susceptibility and increase in chemical shift,[76,77,42] B1 inhomogeneity results in systematic error for T1 measurement.[43,78] In oncology practice, tumor necrosis from response to anticancer therapies leads to increased water diffusion. This results in higher signal intensity on the higher-b-value images, and the apparent diffusion coefficient (ADC) value of the corresponding region will typically increase as there is no restriction to the diffusion of water with a destruction of the closely packed cell membranes. Changes in ADC on DWI have been shown to be effective for monitoring therapeutic response in solid tumore.[79-81,44] Therefore, the field strength should be considered to evaluate the therapeutic response with ADC value.

There are field strength-related changes on the relaxivity of MR contrast media. The relaxivity of gadolinium based MR contrast media increases 5% to 10% when

Formatted: Superscript

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Formatted: Superscript

Field Code Changed

Formatted: Superscript

changing from 1.5T to 3T.[42,76] The individual dependencies of relaxivities on field strength for the types of MR contrast media were significantly different (Table 1).[45,46,82,83] Successful treatment leads to decreased magnitude of enhancement. For detecting change of enhancement of tumor, dosing for contrast media would need to be modified according to field strength.

b. MRI Acquisition parameters for staging of lung cancer

In patients with lung cancer, MRIs have been widely used to evaluate invasion of mediastinum and organs because of superior soft tissue contrast. With advancement of MR techniques, size threshold for nodules have increased, and MR has potential role for assessing indeterminate lung nodules.[47,84] With short echo times, fast spin echo sequences have enabled to assess NSCLC, however, T2 blurring affected the reproducibility of the evaluation of NSCLC.[48,85] Ultrashort echo time (UTE) with the exceedingly short T2 and T2* relaxation times of the lungs has been used in nodule detection and nodule type classification[47,49,84,86]. However, UTE sequences have the disadvantages of long scan duration because of inefficient k-space coverage and are sensitive to motion artifacts.[50,87] 3D UTE provide isotropic spatial resolution with full chest coverage and is less sensitive to motion artifacts.[48,85] The use of limited field-of-view excitation, variable readout gradient, and radial oversampling improves image quality on 3D UTE.[48,85] Most 3D UTE sequences have acquired images using radial-based trajectories.[50,87] A recent study by Ohno et al. showed that UTE images with radial acquisition was useful in detection and classification of pulmonary nodules larger than 4 mm, and interobserver agreement for nodule classification was excellent (κ = 0.95).[49,86] In addition, UTEs with spiral trajectories over a radial readout are reported to have the advantage of high k-space coverage speed while preserving image quality.[50,87] A study reported that there was a 100% detection rate for nodules 5 mm or larger and 76.7% for 2-5 mm nodules on 3D UTE with stack-of-spirals trajectory.[50,87]

Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed

Short T1 inversion recovery (STIR) turbo spin-echo imaging sequence, which is very sensitive to change in T1 and T2, has been known as an important sequence in pulmonary MR imaging.[8851] The specificity (60.6%) and accuracy (74.5%) of STIR were higher than T1 (37.9% and 67.9%) and T2 (48.5% and 67.9%) in distinguishing malignant from benign nodules.[5189] Therefore, STIR sequence could be used to characterize lung nodule, and assess clinical stage of NSCLC. DWI could be useful in assessment of lung nodules, the staging, and early detection and prediction of treatment response of NSCLC. With high lesion-to-background ratio on high b-value images, DWI is useful in the detection of lung nodules. In addition, DWI allows for the characterization of lung nodules using a quantitative assessment of diffusion of water molecules by calculating the ADC.[4784] Thus, ADC is widely used a quantitative imaging biomarker in evaluating NSCLC. A study reported that ADC value increased by 25% after one cycle of chemotherapy due to tumor necrosis and apoptosis, and this suggest that early response of treatment can be predicted by means of ADC change.[5290] However, the DWI-based evaluation of lung nodule can show the difference in value depending on the quantitative evaluation method and b value selection.[5290] Due to the impact of b value selection on DWI, quantitative parameter values should be changed depend on b value selection. Susceptibility artifacts is one of the reasons for the lower ADC differentiation of lung nodule. As b value increased, the change of distortion and susceptibility artifacts increased, and results in poor SNR.[5391]. The interobserver coefficient of variation of ADC in nodules less than 2 cm was relatively poor.[5492]. Contrast-enhanced T1 sequence can be used to characterize lung nodules according to contrast enhancement patterns as well as difference in signal intensity before and after injection of contrast agent.[5290] Contrast enhancement with gadolinium contrast agent on T1-weighted gradient-recalled echo (GRE) or turbo GRE sequences is superior to those on spin-echo and turbo spin-echo (TSE) sequences.[5593].

Quantitative features than can be derived from medical images helps to evaluate of lung cancer.[614] Entropy was known as the most reproducible MR parameter reflecting

Formatted: Superscript

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

tumor heterogeneity.[56,94] A recent study showed that histogram and texture parameters varied after contrast agent injection on DCE MRI, and the 120-150 second after contrast agent infection was optimal for analysis of MR texture parameter.[56,94] The effects of acquisition parameter variations on pixel signal intensities are masked because of blurring and partial volume effects, thus reducing the effect on the radiomics features. Repeatability of MR quantitative parameters is better for global features such as first-order statistical histogram and model-based fractal features than for local-regional texture parameters.[38,70]

Field Code Changed

Field Code Changed

Field Code Changed

c. Compensating for respiratory motion

There is an artificial increase in the volume of a solitary NSCLC during inspiration because of stretching of the tumor and the surrounding peritumoral lung parenchyma [12,25] Moreover, breathing-related motion can decrease the signal intensity on MRI particularly in areas of dynamic air trapping. During inspiration, lung volume is larger; thus, tissue density and MR signal are lower.[57,95] Perfusion could be evaluated qualitatively and quantitatively, and perfusion MRI is performed during breath hold to minimize artifacts from respiration motion because of the fast transit time of contrast agent. However, measurement of perfusion depends strongly on the level of inspiration. During inspiration, pulmonary vascular resistance is increased, while right atrial filling is increased due to the drop in intrapleural pressure. One consequence of this change in physiology is that perfusion at during a breath hold MR angiography exam performed at full inspiration (total lung capacity) is lower than perfusion performed at full expiration (residual volume). It is difficult to control the degree of inspiration during breath hold.[58,96] Therefore, measurements of perfusion performed with have relatively poor reproducibility.[57,95] Some authors have suggested that measurement of perfusion during quiet free breathing can be assessed more reproducibly because free breathing offers better patient compliance.[57,95]

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

d. Functional MR analysis

Angiogenesis is one of the important factors in the evaluation of lung cancer related to tumor survival and growth. DCE-MRI provides information of tumor angiogenesis such as blood flow, vascular volume, and permeability. Once multiple images can be acquired during the first transit or recirculation and washout of contrast medium, quantitative evaluation of contrast passage kinetics can be made.[97,98][59] Pulmonary blood flow (PBF), pulmonary blood volume (PBV) and mean transit time (MTT) can be generated by means of pixel-by-pixel analysis.[60][99] In addition, an arterial input function (AIF), rate of change in the concentration of contrast medium in the plasma with time, is quantified in the larger arteries including main pulmonary.[59,61][97,100] An accurate AIF is necessary for quantitative analysis.[61][100] AIF allows conversion of signal time curves to concentration time curves from assumption of a linear relation between the signal intensity and the concentration of contrast agent.[97,101][59]

Quantitative evaluation of DCE-MRI is based on many pharmacokinetic models, and the Toft's and Kermode model (~~Tofts~~Toft's model, Table 2 and figure 4) are the most frequently used in DCE –MRI analysis.[61][100] The ~~Tofts~~Toft's model was originally constructed with ignoring the effect of intravascular tracer[62][102]:

$$C(t) = K^{trans}\, e^{-tkep} * Ca(t) \qquad [1]$$

Where "*" is convolution, and $C(t)$ and $Ca(t)$ are concentration-time curves in the tissues in interest and in the plasma of a feeding artery, respectively. The standardized terms are presented in Table 2. The parameters of K^{trans} and k_{ep} are defined as follows[102-104][62]:

$$K^{trans} = EF_p \qquad K_{ep} = EF_p / v_e \qquad [2]$$

Where v_e is the fractional volume of the extravascular extracellular space (EES), F_p is the flow of plasma in the capillary bed, and E is related to F_p and the permeability-surface are permeability surface area product of the endothelial wall. The assumption of negligible plasma volume is invalid, particularly tumor tissues. To overcome the

Formatted: Superscript

Field Code Changed

Field Code Changed

Field Code Changed

Formatted: Superscript

Field Code Changed

Field Code Changed

Formatted: Superscript

limitation, the Tofts model has been formulated to allow for an intravascular contribution, which is referred as the extended Tofts model.[62]

$$C(t) = v_p Ca(t) + K^{trans} e^{-tkep} * Ca(t) \quad [3]$$

v_p is the fractional volume of the plasma space.

Various factors could affect the reliability of results in DCE-MRI. The accuracy and precision of pharmacokinetic parameter estimates are strongly influenced by SNR, and temporal resolution.[61,100] Using a theoretical AIF, differences in injection rate and cardiac output are ignored, which may differ between subjects and for a single subject over time.[63,105] Measurement of suboptimal AIF results in worse reproducibility than if a standardized AIF is used, although the AIF might be not so important for evaluating treatment response.[64,106] In addition, blood supply of lung takes places through both the dual pulmonary and bronchial arterial systems.[98] Primary lung cancers are supplied by dual blood supply and the bronchial circulation plays an important role in lung cancer, especially when the size of tumor is larger.[65,107] The single input perfusion analysis according to the maximum slope method calculated the dominant circulation, and ignored the secondary circulation, therefore the result of perfusion of lung cancer are likely to be underestimated.[66,108] To overcome underestimation of perfusion, the dual-input perfusion analysis technique is employed in lung cancer perfusion analysis, and a study reported that dual-input perfusion analysis is helpful for predicting the treatment effect of multi-arterial infusion chemotherapy.[66,108]

C. PET

Due to its quantitative ability and ability to target cellular biology the use of PET has continuously increased for the assessment of therapeutic response in lung cancer. The most commonly used variable is the standardized uptake value (SUV) of 18F-

Formatted: Superscript

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

deoxyglucose (FDG)-based quantitative PET parameters are used as radiomics features and therapeutic response criteria. Many biological and technical factors affect the measurement of SUV, which are described below.[67,109]

a. Normalization method for SUV calculation

SUV is calculated by activity concentration in tissue adjusted by the administered dose of radiopharmaceutical, background SUV of the blood pool and body size. Body size usually corresponds to the body weight of the patient (SUV_{bw}). However, other indexes such as lean body mass (SUV_{lbm}) or body surface area (SUV_{bsa}) can also be used. The choice of how to normalize SUV affects the measurement of SUV and how this value can be used in comparing other studies for therapeutic response to specific agent. One disadvantage of SUV_{bw} is its known overestimation in obese patients. Both SUV_{bw} and SUV_{bsa} are less sensitive to patient weight.[68,110,111]

b. PET/CT scanner models and image acquisition/reconstruction protocol

PET/CT hardware models, image acquisition and reconstruction protocols also affect the quantitative measurement of SUV. For the performance of PET/CT scanners, the most important factors are the intrinsic resolution and detector sensitivity. These key parameters directly affect in-plane resolution and voxel size which determines the amount of partial volume artifact and, SUV variability. This is magnified in the lung bases where nodules move with respiration further adding to volume averaging artifacts and image misregistration with respect to the CT used for attenuation correction.

In the image acquisition protocol, one of most important factors is uptake time. Uptake time is defined as the time interval between the injection of the PET radiopharmaceutical and start of PET scanning. This also influences the measurement of SUV. In the case of ^{18}F- FDG, the most common uptake time is 60 min. The SUV after FDG injection continuously increases as metabolically active cells take up the glucose analogue, which is subsequently trapped.[69,112] Therefore, the use of a fixed uptake time

Field Code Changed

Formatted: Superscript

Formatted: Normal, Indent: Left: 0.56", First line: 0.56", No bullets or numbering

Field Code Changed

is important for the consistency of SUV measurement. On the other hand total scan duration or scan mode (2D vs. 3D) does not have a significant effect on SUV accuracy.[70][113]

In the reconstruction protocols, the attenuation correction method, reconstruction method (analytical vs. statistical/iterative methods), and smoothing filter are major factors affecting SUV measurement. For example, increased smoothing results in decreased noise and increased bias. Increased bias will result in reduced SUV.[71][114]

~~c.~~ c. Patient factors

Even with the same PET/CT protocols and within the same patient repeatability is an important issue. SUV can vary due to the biological process such as different blood glucose and insulin levels, this leads to a high test-retest variability.[72][115] It is well known that plasma serum blood glucose level is inversely correlated with SUVs.[73][116]

~~d.~~ d. Types of quantitative PET parameters

Most quantitative PET parameters have important problems related to measurement variability, precision and repeatability. This includes maximum SUV, average SUV, peak SUV, metabolic tumor volume (MTV), and total lesion glycolysis (TLG).[117] Although maximum SUV is usually not affected by the determination of lesion region of interest (ROI) or VOI, in other PET parameters, ROI/VOI has significant influence. However, there is a persistent concern that maximum SUV represents a single pixel value that may not be representative of the total metabolic profile of the tumor. According to previous studies of NSCLC, MTV and TLG were better prognostic measures than maximum SUV and mean SUV,[118,119][74] suggesting that volume-based parameters of PET may have a role in providing further prognostic information.[74,75][120]

~~e.~~ e. Harmonization of PET parameters

Based on the literature, the measurement variability of maximum SUV, average SUV, and peak SUV expressed as a coefficient of variation is approximately 10%.[76][121]

Field Code Changed

Field Code Changed

Formatted: Font: (Default) Verdana, Font color: Text 1

Formatted: Normal, Indent: Left: 0.83", First line: 0.28", No bullets or numbering

Field Code Changed

Field Code Changed

Formatted: Font: (Default) Verdana, Font color: Text 1

Formatted: Normal, Indent: Left: 0.83", First line: 0.28", No bullets or numbering

Formatted: Superscript

Formatted: Superscript

Formatted: Indent: Left: 1.14", No bullets or numbering

Field Code Changed

Due to these measurement variabilities the harmonization of PET response criteria has been studied. For example, image reconstruction-related variability can be solved using a standardized filter such as EQ.PET.[77][122] To apply this kind of standardized filter for image reconstruction, it is necessary to obtain recovery coefficients according to the lesion size by National Electrical Manufacturers Association (NEMA) NU-2 phantom.[78][123] However, further efforts are necessary to standardize the quantitative measurement of PET parameters.

Field Code Changed

Field Code Changed

f. f. PET Radiotracers for lung cancer

Formatted: Indent: Left: 1.14", No bullets or numbering

Representative PET radiotracers for lung cancer and their clinical utilities are summarized in Table 3. FDG, a glucose analogue, is the most widely used PET radiotracer for lung cancer. It is clinically useful for the single pulmonary nodule evaluation, initial staging, detecting recurrence, and therapy response evaluation.[79][124] ^{18}F-Fluorothymidine PET is good for evaluating therapy response to radiotherapy or chemotherapy early in lung cancer.[80][125] ^{18}F-fluoromisonidazole PET shows the hypoxic portion within the tumor, which can be used for radiotherapy planning in lung cancer by boosting radiation dose to hypoxic tumor.[81][126] ^{18}F-alfatide, reflecting tumor angiogenesis, has a potential to evaluate therapy response in lung cancer, although published studies were very few.[82][127] PET tracers targeting tumor epidermal growth factor receptor (EGFR) such as ^{11}C-PD153035 and ^{11}C-erlotinib may be applicable for therapy response evaluation to EGFR–tyrosine kinase inhibitors.[128,129] [83]

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Formatted: Superscript

3) Special considerations on radiomics analysis

A. Bin number

Radiomics analysis computes hundreds or sometimes thousands of features from the underlying imaging modalities and ROIs. The features are different from semantic features and are agnostic computational features whose formulae are defined with

various parameters.[130] Thus, for a given radiomics feature, if the associated parameter changes, the ensuing radiomics feature might change as well. Many radiomics features, noted as histogram-based features, are computed from the intensity histogram using the underlying imaging data within the ROI. Histograms are affected by binning parameters of bin width and range (Figure 5). Range is application dependent, and we typically use 4096 for CT. Many people also use a number of bins, which is range divided by bin width for the binning parameter. Using many bins allows fine differentiation between intensity values, but using too many bins leads to very narrow bin width. A narrow bin width leads to unreliable histogram estimates, as we may not have enough samples for some bins. The Freedman-Diaconis rule can be used to set bin width.[84131,132]

In addition, ~~we are~~ computing hundreds and thousands of features from a given ROI, ~~which might~~can frequently lead to having too many parameters in an analytical model. This ~~could be~~results in ~~thought as~~ overfitting the of data ~~when we are~~and is an important limitation when training an artificial intelligence model with limited samples. ~~Thus, it is desirable to have as many as possible.~~As a general rule there should not be more features than patients in the training set. Many recent radiomics papers included hundreds or sometimes over one thousand samples.[133-13585,86] In practice, not all the features are included in the resultant radiomics model. There are often feature selection procedure (through least absolute shrinkage and selection operator or something equivalent), where the number of features is reduced to a few (i.e., typically tens of features).[85,87133,136] In this case, having 70-80 samples could still be adequate ~~regarding~~ to avoid ~~ing~~ overfitting.

B. Texture features

Texture features are widely recognized radiomics features.[133,137,13885] The most representative texture features are computed from gray-level co-occurrence matrix (GLCM) and intensity size zone matrix (ISZM). These matrices are built out of 2D

Formatted: Superscript

Formatted: Superscript

Field Code Changed

Formatted: Superscript

histograms, which measure the frequency of a pair of observations compared to a 1D histogram, where researchers consider the frequency of one observation (e.g., intensity). GLCM measures the frequency of intensity pairs in the neighborhood, while ISZM measures the frequency of blobs with certain size and intensity. For GLCM, the 2D histogram is built using intensity of the given voxel as the first axis and intensity of the neighboring voxel as the second axis. The GLCM quantifies how intensity pairs occur in a neighbor and hence can reflect textural information. Similar to the 1D histogram case, the number of bins is a major parameter in 2D histograms. In general, the samples in a given bin for a 2D histogram is less than those in an equivalent 1D histogram as the voxels need to fill the bins spanning the 2D histogram compared to filling the 1D histogram. ~~In general, the number of samples in a the 2D histogram areis a lot less as the voxels need to fill the bins spanning the 2D space compared to the 1D histogram.~~ Due to this sparsity in the 2D histogram, researchers typically use 128/256 bins for GLCM.[88][139] Figure 6 shows typical 1D intensity histogram and 2D GLCM and ISZM histograms.

Commented [mls2]: Why wouldn't this be more information needed to fill a 2D histogram than a 1D histogram?

Field Code Changed

The size of the ROI also affects the 1D/2D histogram measures. If the~~your~~ ROI is big enough to contain thousands of voxels, then the above approaches are suitable. If the~~a~~ ROI has a very small number of voxels (~~perhaps around~~e.g. 100), then researchers need to reduce the number of bins significantly to make sure there are enough voxels occupying the bins.

C. Shape features

Shape features are important parts of radiomics analysis.[~~133,140~~85] The shape of ROI is quantified with various formulae. ROI is composed of voxels that could be isotropic or non-isotropic. In ~~many cases, we have~~chest imaging there is often good in-plane resolution and ~~poor~~variable out-of-plane resolution (i.e., non-isotropic voxels). For non-isotropic voxels, the shape features are more sensitive to shape change

Formatted: Superscript

occurring in-plane while less sensitive to shape change occurring out-of-plane. For isotropic voxels, the shape features are equally sensitive into all directions. The shape of the target ROI may change in any direction; thus, isotropic voxels are preferable over non-isotropic ones. If the imaging data is non-isotropic, we can interpolate the imaging data to make it isotropic. This interpolation makes the data smoother but at least reducesalso reduces the shape variability. In other words, the fine edge detail is lost. This is similar to iterative reconstruction methods in CT used to decrease patient dose. The dose is decreased, but fine detail can beis lost in these smoother more "plastic" appearing images.

D. Filter and Wavelet

Some researchers have applied an edge enhancement filter such as Laplacian of Gaussian (LoG) to the reconstructed image data and only then compute radiomics features from the filtered image.[89141] The LoG filter has a scale parameter that controls the scale at which enhancement occurs (Figure 7). Researchers need to specify the scale parameter to suit their intended application. The scale should be set based on image quality and the size of ROI. If researchers have poor quality image with large ROIs, large scale operations are recommended.

Field Code Changed

Some studies also apply wavelet decomposition to imaging data.[85133] The imaging data are decomposed into many output data, and radiomics features are then computed from the decomposed data. There are many wavelet transforms to choose from, each with a plethora of parameters. Coiflets are widely used for their simplicity. Researchers can decompose one 3D scan into 8 3D decomposed scans in its simplest version. Different wavelet transform leads to different decomposed data and thus affects the radiomics features. Researchers should fully consider the various parameters of wavelets before applying them in their projects.

Field Code Changed

Future considerations

~~3. Future Suggestions and Conclusion~~

As radiomics features show promising benefits for quantification of lung cancer biology and response to treatment, many researchers are now paying close attention to the clinical usefulness of radiomics in oncologic studies. However, as the number of radiomics studies explodes, it should be clearly noted that the extracted radiomics features are subject to lack of precision and repeatability. Therefore, in order to homogenize evaluation criteria and reporting guidelines for radiomics, Lambin et al. proposed the radiomics quality score (RQS) (Table 4).[86~~134~~] The RQS evaluates the necessary steps in radiomics analysis including 16 key components of which each is given a number of points corresponding to the importance of the respective component.[86~~134~~] Major check points in the RQS are data selection, medical imaging, features extraction, exploratory analysis, and modeling. The highest possible total RQS for quantification of the overall methodology and analysis of radiomics practice is 36 points. Therefore, efforts should be made to consider RQS in future studies and to stablish collaborative foundations to control and fully realize the potential of radiomics.

~~One last consideration~~Another feature that may help in tumor response evaluation is delta radiomics.[86,90~~134,142~~] In contrast to most radiomics studies, which are based on features extracted at a single time point (usually at the time of diagnosis), delta radiomics evaluates changes in radiomics features between interval studies. Delta-radiomics features have shown potential in predicting response or survival in patients with colorectal cancer liver metastasis, metastatic renal cell, and lung cancer.[90-92~~142-144~~] According to a study of 107 NSCLC patients, pretreatment radiomics features were not prognostic, while texture-strength measured at the end of treatment significantly stratified high- and low-risk patients, thus suggesting the potential of delta-radiomics features.[90~~142~~] Nevertheless, if delta radiomics were to be employed in clinical practice,

Commented [mls3]: Good table

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

standardization of technical factors and high reproducibility of the features remain prerequisites.

In conclusion, compared to the current RECIST version 1.1, tumor volumetric measurement and radiomics are more quantitative measures and supplement the limitations of RECIST in the current era of precision cancer therapy. Nevertheless, substantial variability can be introduced in the process of measuring tumor burden due to various technical factors. Furthermore, the increasing role of software post-processing and radiomics support the need for increased awareness of technical factors of image acquisition among radiologists. We await the incorporation of these advanced image processing metrics of Radiomics and Artificial Intelligence into the new RECIST criteria for tumor response assessment. ~~In the future, the traditional role of radiological practice in oncological studies is likely to change, and the concepts and knowledge described in this review will support radiologists with a new perspective for tumor response evaluations in the cutting edge cancer patient care.~~

References

1. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45:228-247.
2. Kang H, Lee HY, Lee KS, et al. Imaging-based tumor treatment response evaluation: review of conventional, new, and emerging concepts. *Korean J Radiol*. 2012;13:371-390.
3. Erasmus JJ, Gladish GW, Broemeling L, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *J Clin Oncol*. 2003;21:2574-2582.
4. Nishino M, Tirumani SH, Ramaiya NH, et al. Cancer immunotherapy and immune-related response assessment: The role of radiologists in the new arena of cancer treatment. *Eur J Radiol*. 2015;84:1259-1268.
5. Nishino M, Ramaiya NH, Hatabu H, et al. Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol*. 2017;14:655-668.
6. Lee G, Lee HY, Park H, et al. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art. *Eur J Radiol*. 2017;86:297-307.
7. Scholten ET, de Hoop B, Jacobs C, et al. Semi-automatic quantification of subsolid pulmonary nodules: comparison with manual measurements. *PLoS One*. 2013;8:e80249.
8. Nishino M, Dahlberg SE, Cardarella S, et al. Tumor volume decrease at 8 weeks is associated with longer survival in EGFR-mutant advanced non-small-cell lung cancer patients treated with EGFR TKI. *J Thorac Oncol*. 2013;8:1059-1068.
9. Nishino M, Guo M, Jackman DM, et al. CT tumor volume measurement in advanced non-small-cell lung cancer: Performance characteristics of an emerging clinical tool. *Acad Radiol*. 2011;18:54-62.
10. Han D, Heuvelmans MA, Oudkerk M. Volume versus diameter assessment of small pulmonary nodules in CT lung cancer screening. *Transl Lung Cancer Res*. 2017;6:52-61.
11. Zhao B, Oxnard GR, Moskowitz CS, et al. A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development. *Clin Cancer Res*. 2010;16:4647-4653.
12. Plathow C, Schoebinger M, Fink C, et al. Quantification of lung tumor volume and rotation at 3D dynamic parallel MR imaging with view sharing: preliminary results. *Radiology*. 2006;240:537-545.
13. Rios Velazquez E, Aerts HJ, Gu Y, et al. A semiautomatic CT-based ensemble segmentation of lung tumors: comparison with oncologists' delineations and with the surgical specimen. *Radiother Oncol*. 2012;105:167-173.
14. Parmar C, Rios Velazquez E, Leijenaar R, et al. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One*. 2014;9:e102107.
15. Kalpathy-Cramer J, Zhao B, Goldgof D, et al. A Comparison of Lung Nodule Segmentation Algorithms: Methods and Results from a Multi-institutional Study. *J Digit Imaging*.

Commented [mls4]: Trim to 80 references

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

2016;29:476-487.

16. Ko JP, Rusinek H, Jacobs EL, et al. Small pulmonary nodules: volume measurement at chest CT—phantom study. *Radiology*. 2003;228:864-870.
17. Lassen B, Jacobs C, Kuhnigk J, et al. Robust semi-automatic segmentation of pulmonary subsolid nodules in chest computed tomography scans. *Physics in Medicine & Biology*. 2015;60:1307.
18. Nishino M, Dahlberg SE, Fulton LE, et al. Volumetric Tumor Response and Progression in EGFR-mutant NSCLC Patients Treated with Erlotinib or Gefitinib. *Acad Radiol*. 2016;23:329-336.
19. Havaei M, Davy A, Warde-Farley D, et al. Brain tumor segmentation with deep neural networks. *Medical image analysis*. 2017;35:18-31.
20. Trebeschi S, van Griethuysen JJ, Lambregts DM, et al. Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR. *Scientific reports*. 2017;7:5301.
21. Ashraf H, de Hoop B, Shaker SB, et al. Lung nodule volumetry: segmentation algorithms within the same software package cannot be used interchangeably. *Eur Radiol*. 2010;20:1878-1885.
22. Devaraj A, van Ginneken B, Nair A, et al. Use of Volumetry for Lung Nodule Management: Theory and Practice. *Radiology*. 2017;284:630-644.
23. Huber A, Landau J, Ebner L, et al. Performance of ultralow-dose CT with iterative reconstruction in lung cancer screening: limiting radiation exposure to the equivalent of conventional chest X-ray imaging. *Eur Radiol*. 2016;26:3643-3652.
24. Stoel BC, Vrooman HA, Stolk J, et al. Sources of error in lung densitometry with CT. *Invest Radiol*. 1999;34:303-309.
25. Kim H, Park CM, Chae HD, et al. Impact of radiation dose and iterative reconstruction on pulmonary nodule measurements at chest CT: a phantom study. *Diagn Interv Radiol*. 2015;21:459-465.
26. Ohno Y, Yaguchi A, Okazaki T, et al. Comparative evaluation of newly developed model-based and commercially available hybrid-type iterative reconstruction methods and filter back projection method in terms of accuracy of computer-aided volumetry (CADv) for low-dose CT protocols in phantom study. *Eur J Radiol*. 2016;85:1375-1382.
27. Sui X, Meinel FG, Song W, et al. Detection and size measurements of pulmonary nodules in ultra-low-dose CT with iterative reconstruction compared to low dose CT. *Eur J Radiol*. 2016;85:564-570.
28. Cohen JG, Kim H, Park SB, et al. Comparison of the effects of model-based iterative reconstruction and filtered back projection algorithms on software measurements in pulmonary subsolid nodules. *Eur Radiol*. 2017;27:3266-3274.
29. Petrou M, Quint LE, Nan B, et al. Pulmonary nodule volumetric measurement variability as a function of CT slice thickness and nodule morphology. *AJR Am J Roentgenol*. 2007;188:306-312.

Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar

30. Tan Y, Guo P, Mann H, et al. Assessing the effect of CT slice interval on unidimensional, bidimensional and volumetric measurements of solid tumours. *Cancer Imaging*. 2012;12:497-505.
31. Zhao B, Tan Y, Tsai WY, et al. Exploring Variability in CT Characterization of Tumors: A Preliminary Phantom Study. *Transl Oncol*. 2014;7:88-93.
32. He L, Huang Y, Ma Z, et al. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Sci Rep*. 2016;6:34921.
33. Wang Y, de Bock GH, van Klaveren RJ, et al. Volumetric measurement of pulmonary nodules at low-dose chest CT: effect of reconstruction setting on measurement variability. *Eur Radiol*. 2010;20:1180-1187.
34. Christe A, Bronnimann A, Vock P. Volumetric analysis of lung nodules in computed tomography (CT): comparison of two different segmentation algorithm softwares and two different reconstruction filters on automated volume calculation. *Acta Radiol*. 2014;55:54-61.
35. Oliver JA, Budzevich M, Zhang GG, et al. Variability of Image Features Computed from Conventional and Respiratory-Gated PET/CT Images of Lung Cancer. *Transl Oncol*. 2015;8:524-534.
36. Honda O, Johkoh T, Sumikawa H, et al. Pulmonary nodules: 3D volumetric measurement with multidetector CT--effect of intravenous contrast medium. *Radiology*. 2007;245:881-887.
37. Rampinelli C, Raimondi S, Padrenostro M, et al. Pulmonary nodules: Contrast-enhanced volumetric variation at different CT scan delays. *AJR Am J Roentgenol*. 2010;195:149-154.
38. Gourtsoyianni S, Doumou G, Prezzi D, et al. Primary Rectal Cancer: Repeatability of Global and Local-Regional MR Imaging Texture Features. *Radiology*. 2017;284:552-561.
39. Incoronato M, Aiello M, Infante T, et al. Radiogenomic Analysis of Oncological Data: A Technical Survey. *Int J Mol Sci*. 2017;18.
40. Usmani N, Sloboda R, Kamal W, et al. Can images obtained with high field strength magnetic resonance imaging reduce contouring variability of the prostate? *Int J Radiat Oncol Biol Phys*. 2011;80:728-734.
41. Saha A, Yu X, Sahoo D, et al. Effects of MRI scanner parameters on breast cancer radiomics. *Expert Systems with Applications*. 2017;87:384-391.
42. Soher BJ, Dale BM, Merkle EM. A review of MR physics: 3T versus 1.5T. *Magn Reson Imaging Clin N Am*. 2007;15:277-290, v.
43. Leach MO, Morgan B, Tofts PS, et al. Imaging vascular function for early stage clinical trials using dynamic contrast-enhanced magnetic resonance imaging. *Eur Radiol*. 2012;22:1451-1464.
44. Thoeny HC, Ross BD. Predicting and monitoring cancer treatment response with diffusion-weighted MRI. *J Magn Reson Imaging*. 2010;32:2-16.
45. Knobloch G, Colgan T, Wiens CN, et al. Relaxivity of Ferumoxytol at 1.5 T and 3.0 T. *Invest*

Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar
Formatted: Check spelling and grammar
Formatted: Font: Italic, Check spelling and grammar

Radiol. 2018;53:257-263.

46. Rohrer M, Bauer H, Mintorovitch J, et al. Comparison of magnetic properties of MRI contrast media solutions at different magnetic field strengths. *Invest Radiol*. 2005;40:715-724.

47. Ciliberto M, Kishida Y, Seki S, et al. Update of MR Imaging for Evaluation of Lung Cancer. *Radiol Clin North Am*. 2018;56:437-469.

48. Johnson KM, Fain SB, Schiebler ML, et al. Optimized 3D ultrashort echo time pulmonary MRI. *Magn Reson Med*. 2013;70:1241-1250.

49. Ohno Y, Koyama H, Yoshikawa T, et al. Standard-, Reduced-, and No-Dose Thin-Section Radiologic Examinations: Comparison of Capability for Nodule Detection and Nodule Type Assessment in Patients Suspected of Having Pulmonary Nodules. *Radiology*. 2017;284:562-573.

50. Cha MJ, Park HJ, Paek MY, et al. Free-breathing ultrashort echo time lung magnetic resonance imaging using stack-of-spirals acquisition: A feasibility study in oncology patients. *Magn Reson Imaging*. 2018;51:137-143.

51. Koyama H, Ohno Y, Kono A, et al. Quantitative and qualitative assessment of non-contrast-enhanced pulmonary MR imaging for management of pulmonary nodules in 161 subjects. *Eur Radiol*. 2008;18:2120-2131.

52. Ohno Y, Kauczor HU, Hatabu H, et al. MRI for solitary pulmonary nodule and mass assessment: Current state of the art. *J Magn Reson Imaging*. 2018;47:1437-1458.

53. Koyama H, Ohno Y, Seki S, et al. Value of diffusion-weighted MR imaging using various parameters for assessment and characterization of solitary pulmonary nodules. *Eur J Radiol*. 2015;84:509-515.

54. Bernardin L, Douglas NH, Collins DJ, et al. Diffusion-weighted magnetic resonance imaging for assessment of lung lesions: repeatability of the apparent diffusion coefficient measurement. *Eur Radiol*. 2014;24:502-511.

55. Sasaki M, Shibata E, Kanbara Y, et al. Enhancement effects and relaxivities of gadolinium-DTPA at 1.5 versus 3 Tesla: a phantom study. *Magn Reson Med Sci*. 2005;4:145-149.

56. Yoon SH, Park CM, Park SJ, et al. Tumor Heterogeneity in Lung Cancer: Assessment with Dynamic Contrast-enhanced MR Imaging. *Radiology*. 2016;280:940-948.

57. Ingrisch M, Maxien D, Schwab F, et al. Assessment of pulmonary perfusion with breath-hold and free-breathing dynamic contrast-enhanced magnetic resonance imaging: quantification and reproducibility. *Invest Radiol*. 2014;49:382-389.

58. Fink C, Ley S, Risse F, et al. Effect of inspiratory and expiratory breathhold on pulmonary perfusion: assessment by pulmonary perfusion magnetic resonance imaging. *Invest Radiol*. 2005;40:72-79.

59. Johns CS, Swift AJ, Hughes PJC, et al. Pulmonary MR angiography and perfusion imaging-A review of methods and applications. *Eur J Radiol*. 2017;86:361-370.

60. Ohno Y, Koyama H, Lee HY, et al. Contrast-enhanced CT- and MRI-based perfusion assessment for pulmonary diseases: basics and clinical applications. *Diagn Interv Radiol*.

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

2016;22:407-421.

61. Gaddikeri S, Gaddikeri RS, Tailor T, et al. Dynamic Contrast-Enhanced MR Imaging in Head and Neck Cancer: Techniques and Clinical Applications. *AJNR Am J Neuroradiol*. 2016;37:588-595.

62. Sourbron SP, Buckley DL. On the scope and interpretation of the Tofts models for DCE-MRI. *Magn Reson Med*. 2011;66:735-745.

63. Ashton E, Raunig D, Ng C, et al. Scan-rescan variability in perfusion assessment of tumors in MRI using both model and data-derived arterial input functions. *J Magn Reson Imaging*. 2008;28:791-796.

64. Morgan B, Utting JF, Higginson A, et al. A simple, reproducible method for monitoring the treatment of tumours using dynamic contrast-enhanced MR imaging. *Br J Cancer*. 2006;94:1420-1427.

65. Yuan X, Zhang J, Ao G, et al. Lung cancer perfusion: can we measure pulmonary and bronchial circulation simultaneously? *Eur Radiol*. 2012;22:1665-1671.

66. Li XS, Fan HX, Fang H, et al. Value of whole-tumor dual-input perfusion CT in predicting the effect of multiarterial infusion chemotherapy on advanced non-small cell lung cancer. *AJR Am J Roentgenol*. 2014;203:W497-505.

67. Adams MC, Turkington TG, Wilson JM, et al. A systematic review of the factors affecting accuracy of SUV measurements. *AJR Am J Roentgenol*. 2010;195:310-320.

68. Kim CK, Gupta NC, Chandramouli B, et al. Standardized uptake values of FDG: body surface area correction is preferable to body weight correction. *J Nucl Med*. 1994;35:164-167.

69. Lowe VJ, DeLong DM, Hoffman JM, et al. Optimum scanning protocol for FDG-PET evaluation of pulmonary malignancy. *J Nucl Med*. 1995;36:883-887.

70. Kinahan PE, Fletcher JW. Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. *Semin Ultrasound CT MR*. 2010;31:496-505.

71. Doot RK, Scheuermann JS, Christian PE, et al. Instrumentation factors affecting variance and bias of quantifying tracer uptake with PET/CT. *Med Phys*. 2010;37:6035-6046.

72. Hofheinz F, Apostolova I, Oehme L, et al. Test-Retest Variability in Lesion SUV and Lesion SUR in (18)F-FDG PET: An Analysis of Data from Two Prospective Multicenter Trials. *J Nucl Med*. 2017;58:1770-1775.

73. Huang SC. Anatomy of SUV. Standardized uptake value. *Nucl Med Biol*. 2000;27:643-646.

74. Liao S, Penney BC, Wroblewski K, et al. Prognostic value of metabolic tumor burden on 18F-FDG PET in nonsurgical patients with non-small cell lung cancer. *Eur J Nucl Med Mol Imaging*. 2012;39:27-38.

75. Kim K, Kim SJ, Kim IJ, et al. Prognostic value of volumetric parameters measured by F-18 FDG PET/CT in surgically resected non-small-cell lung cancer. *Nucl Med Commun*. 2012;33:613-620.

76. Lodge MA. Repeatability of SUV in Oncologic (18)F-FDG PET. *J Nucl Med*. 2017;58:523-532.

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

77. Quak E, Le Roux PY, Lasnon C, et al. Does PET SUV Harmonization Affect PERCIST Response Classification? *J Nucl Med*. 2016;57:1699-1706.

78. Lasnon C, Desmonts C, Quak E, et al. Harmonizing SUVs in multicentre trials when using different generation PET systems: prospective validation in non-small cell lung cancer patients. *Eur J Nucl Med Mol Imaging*. 2013;40:985-996.

79. Volpi S, Ali JM, Tasker A, et al. The role of positron emission tomography in the diagnosis, staging and response assessment of non-small cell lung cancer. *Ann Transl Med*. 2018;6:95.

80. Everitt SJ, Ball DL, Hicks RJ, et al. Differential (18)F-FDG and (18)F-FLT Uptake on Serial PET/CT Imaging Before and During Definitive Chemoradiation for Non-Small Cell Lung Cancer. *J Nucl Med*. 2014;55:1069-1074.

81. Vera P, Thureau S, Chaumet-Riffaud P, et al. Phase II Study of a Radiotherapy Total Dose Increase in Hypoxic Lesions Identified by (18)F-Misonidazole PET/CT in Patients with Non-Small Cell Lung Carcinoma (RTEP5 Study). *J Nucl Med*. 2017;58:1045-1053.

82. Luan X, Huang Y, Gao S, et al. (18)F-alfatide PET/CT may predict short-term outcome of concurrent chemoradiotherapy in patients with advanced non-small cell lung cancer. *Eur J Nucl Med Mol Imaging*. 2016;43:2336-2342.

83. Dai D, Li XF, Wang J, et al. Predictive efficacy of (11)C-PD153035 PET imaging for EGFR-tyrosine kinase inhibitor sensitivity in non-small cell lung cancer patients. *Int J Cancer*. 2016;138:1003-1012.

84. Parekh V, Jacobs MA. Radiomics: a new application from established techniques. *Expert Rev Precis Med Drug Dev*. 2016;1:207-226.

85. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.

86. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14:749-762.

87. Huang Y, Liu Z, He L, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. *Radiology*. 2016;281:947-957.

88. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys*. 2017;44:1050-1062.

89. Aerts HJ, Grossmann P, Tan Y, et al. Defining a Radiomic Response Phenotype: A Pilot Study using targeted therapy in NSCLC. *Sci Rep*. 2016;6:33860.

90. Fave X, Zhang L, Yang J, et al. Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. *Sci Rep*. 2017;7:588.

91. Goh V, Ganeshan B, Nathan P, et al. Assessment of response to tyrosine kinase inhibitors in metastatic renal cell cancer: CT texture as a predictive biomarker. *Radiology*. 2011;261:165-171.

92. Rao SX, Lambregts DM, Schnerr RS, et al. CT texture analysis in colorectal liver metastases: A better way than size and volume measurements to assess response to chemotherapy? *United European Gastroenterol J*. 2016;4:257-263.

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

Formatted: Font: Italic, Check spelling and grammar

Formatted: Check spelling and grammar

1. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer.* 2009;45:228-247.
2. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst.* 2000;92:205-216.
3. Kang H, Lee HY, Lee KS, et al. Imaging based tumor treatment response evaluation: review of conventional, new, and emerging concepts. *Korean J Radiol.* 2012;13:371-390.
4. Erasmus JJ, Gladish GW, Broemeling L, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *J Clin Oncol.* 2003;21:2574-2582.
5. Tran LN, Brown MS, Goldin JG, et al. Comparison of treatment response classifications between unidimensional, bidimensional, and volumetric measurements of metastatic lung lesions on chest computed tomography. *Acad Radiol.* 2004;11:1355-1360.
6. Nishino M, Hatabu H, Johnson BE, et al. State of the art: Response assessment in lung cancer in the era of genomic medicine. *Radiology.* 2014;271:6-27.
7. Nishino M, Ramaiya NH, Hatabu H, et al. Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol.* 2017;14:655-668.
8. Ott PA, Hodi FS, Robert C. CTLA-4 and PD-1/PD-L1 blockade: new immunotherapeutic modalities with durable clinical benefit in melanoma patients. *Clin Cancer Res.* 2013;19:5300-5309.
9. Nishino M, Giobbie-Hurder A, Gargano M, et al. Developing a common language for tumor response to immunotherapy: immune-related response criteria using unidimensional measurements. *Clin Cancer Res.* 2013;19:3936-3943.
10. Nishino M, Tirumani SH, Ramaiya NH, et al. Cancer immunotherapy and immune-related response assessment: The role of radiologists in the new arena of cancer treatment. *Eur J Radiol.* 2015;84:1259-1268.
11. Choi H, Charnsangavej C, Faria SC, et al. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. *J Clin Oncol.* 2007;25:1753-1759.
12. Nishino M, Cryer SK, Okajima Y, et al. Tumoral cavitation in patients with non-small-cell lung cancer treated with antiangiogenic therapy using bevacizumab. *Cancer Imaging.* 2012;12:225-235.
13. Nishino M, Jagannathan JP, Krajewski KM, et al. Personalized tumor response assessment in the era of molecular medicine: cancer-specific and therapy-specific response criteria to complement pitfalls of RECIST. *AJR Am J Roentgenol.* 2012;198:737-745.
14. Lee G, Lee HY, Park H, et al. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art. *Eur J Radiol.* 2017;86:297-

Formatted: Indent: Left: 0", Hanging: 0.5", Space After: 0 pt, Line spacing: single

307.

15. Padhani AR, Miles KA. Multiparametric imaging of tumor response to therapy. *Radiology.* 2010;256:348-364.
16. Scholten ET, de Hoop B, Jacobs C, et al. Semi-automatic quantification of subsolid pulmonary nodules: comparison with manual measurements. *PLoS One.* 2013;8:e80249.
17. Goldmacher GV, Conklin J. The use of tumour volumetrics to assess response to therapy in anticancer clinical trials. *Br J Clin Pharmacol.* 2012;73:846-854.
18. Jennings SG, Winer-Muram HT, Tarver RD, et al. Lung tumor growth: assessment with CT--comparison of diameter and cross-sectional area with volume measurements. *Radiology.* 2004;231:866-871.
19. Mozley PD, Bendtsen C, Zhao B, et al. Measurement of tumor volumes improves RECIST-based response assessments in advanced lung cancer. *Transl Oncol.* 2012;5:19-25.
20. Nishino M, Dahlberg SE, Cardarella S, et al. Tumor volume decrease at 8 weeks is associated with longer survival in EGFR-mutant advanced non-small-cell lung cancer patients treated with EGFR TKI. *J Thorac Oncol.* 2013;8:1059-1068.
21. Nishino M, Guo M, Jackman DM, et al. CT tumor volume measurement in advanced non-small-cell lung cancer: Performance characteristics of an emerging clinical tool. *Acad Radiol.* 2011;18:54-62.
22. Zhao B, Schwartz LH, Moskowitz CS, et al. Lung cancer: computerized quantification of tumor response--initial results. *Radiology.* 2006;241:892-898.
23. Han D, Heuvelmans MA, Oudkerk M. Volume versus diameter assessment of small pulmonary nodules in CT lung cancer screening. *Transl Lung Cancer Res.* 2017;6:52-61.
24. Zhao B, Oxnard GR, Moskowitz CS, et al. A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development. *Clin Cancer Res.* 2010;16:4647-4653.
25. Plathow C, Schoebinger M, Fink C, et al. Quantification of lung tumor volume and rotation at 3D dynamic parallel MR imaging with view sharing: preliminary results. *Radiology.* 2006;240:537-545.
26. Altorki N, Lane ME, Bauer T, et al. Phase II proof-of-concept study of pazopanib monotherapy in treatment-naive patients with stage I/II resectable non-small-cell lung cancer. *J Clin Oncol.* 2010;28:3131-3137.
27. Hayes SA, Pietanza MC, O'Driscoll D, et al. Comparison of CT volumetric measurement with RECIST response in patients with lung cancer. *Eur J Radiol.* 2016;85:524-533.
28. Rios Velazquez E, Aerts HJ, Gu Y, et al. A semiautomatic CT-based ensemble segmentation of lung tumors: comparison with oncologists' delineations and with the surgical specimen. *Radiother Oncol.* 2012;105:167-173.
29. van Dam IE, van Sornsen de Koste JR, Hanna GG, et al. Improving target delineation on 4-dimensional CT scans in stage I NSCLC using a deformable registration tool. *Radiother Oncol.* 2010;96:67-72.
30. Heye T, Merkle EM, Reiner CS, et al. Reproducibility of dynamic contrast-enhanced MR-

imaging. Part II. Comparison of intra- and interobserver variability with manual region of interest placement versus semiautomatic lesion segmentation and histogram analysis. *Radiology.* 2013;266:812-821.

31. Parmar C, Rios Velazquez E, Leijenaar R, et al. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One.* 2014;9:e102107.
32. Kalpathy-Cramer J, Zhao B, Goldgof D, et al. A Comparison of Lung Nodule Segmentation Algorithms: Methods and Results from a Multi-institutional Study. *J Digit Imaging.* 2016;29:476-487.
33. Ko JP, Rusinek H, Jacobs EL, et al. Small pulmonary nodules: volume measurement at chest CT—phantom study. *Radiology.* 2003;228:864-870.
34. Lassen B, Jacobs C, Kuhnigk J, et al. Robust semi-automatic segmentation of pulmonary subsolid nodules in chest computed tomography scans. *Physics in Medicine & Biology.* 2015;60:1307.
35. Oda S, Awai K, Murao K, et al. Computer-aided volumetry of pulmonary nodules exhibiting ground-glass opacity at MDCT. *American Journal of Roentgenology.* 2010;194:398-406.
36. Nishino M, Dahlberg SE, Fulton LE, et al. Volumetric Tumor Response and Progression in EGFR-mutant NSCLC Patients Treated with Erlotinib or Gefitinib. *Acad Radiol.* 2016;23:329-336.
37. Havaei M, Davy A, Warde-Farley D, et al. Brain tumor segmentation with deep neural networks. *Medical image analysis.* 2017;35:18-31.
38. Trebeschi S, van Griethuysen JJ, Lambregts DM, et al. Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR. *Scientific reports.* 2017;7:5301.
39. Wang S, Zhou M, Liu Z, et al. Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. *Med Image Anal.* 2017;40:172-183.
40. Zhao L, Jia K. Multiscale cnns for brain tumor segmentation and diagnosis. *Computational and mathematical methods in medicine.* 2016;2016.
41. Ashraf H, de Hoop B, Shaker SB, et al. Lung nodule volumetry: segmentation algorithms within the same software package cannot be used interchangeably. *Eur Radiol.* 2010;20:1878-1885.
42. de Hoop B, Gietema H, van Ginneken B, et al. A comparison of six software packages for evaluation of solid lung nodules using semi-automated volumetry: what is the minimum increase in size to detect growth in repeated CT examinations. *Eur Radiol.* 2009;19:800-808.
43. Devaraj A, van Ginneken B, Nair A, et al. Use of Volumetry for Lung Nodule Management: Theory and Practice. *Radiology.* 2017;284:630-644.
44. Zhao YR, van Ooijen PM, Dorrius MD, et al. Comparison of three software systems for semi-automatic volumetry of pulmonary nodules on baseline and follow-up CT examinations. *Acta Radiol.* 2014;55:691-698.
45. Huber A, Landau J, Ebner L, et al. Performance of ultralow-dose CT with iterative

reconstruction in lung cancer screening: limiting radiation exposure to the equivalent of conventional chest X-ray imaging. *Eur Radiol.* 2016;26:3643-3652.

46. Kemerink GJ, Lamers RJ, Thelissen GR, et al. Scanner conformity in CT densitometry of the lungs. *Radiology.* 1995;197:749-752.
47. Stoel BC, Bode F, Rames A, et al. Quality control in longitudinal studies with computed tomographic densitometry of the lungs. *Proc Am Thorac Soc.* 2008;5:929-933.
48. Stoel BC, Vrooman HA, Stolk J, et al. Sources of error in lung densitometry with CT. *Invest Radiol.* 1999;34:303-309.
49. Kim H, Park CM, Chae HD, et al. Impact of radiation dose and iterative reconstruction on pulmonary nodule measurements at chest CT: a phantom study. *Diagn Interv Radiol.* 2015;21:459-465.
50. Kim H, Park CM, Song YS, et al. Influence of radiation dose and iterative reconstruction algorithms for measurement accuracy and reproducibility of pulmonary nodule volumetry: A phantom study. *Eur J Radiol.* 2014;83:848-857.
51. Ohno Y, Yaguchi A, Okazaki T, et al. Comparative evaluation of newly developed model-based and commercially available hybrid-type iterative reconstruction methods and filter back projection method in terms of accuracy of computer-aided volumetry (CADv) for low-dose CT protocols in phantom study. *Eur J Radiol.* 2016;85:1375-1382.
52. Siegelman JW, Supanich MP, Gavrielides MA. Pulmonary nodules with ground-glass opacity can be reliably measured with low-dose techniques regardless of iterative reconstruction: results of a phantom study. *AJR Am J Roentgenol.* 2015;204:1242-1247.
53. Doo KW, Kang EY, Yong HS, et al. Accuracy of lung nodule volumetry in low-dose CT with iterative reconstruction: an anthropomorphic thoracic phantom study. *Br J Radiol.* 2014;87:20130644.
54. Sakai N, Yabuuchi H, Kondo M, et al. Volumetric measurement of artificial pure ground-glass nodules at low-dose CT: Comparisons between hybrid iterative reconstruction and filtered back projection. *Eur J Radiol.* 2015;84:2654-2662.
55. Sui X, Meinel FG, Song W, et al. Detection and size measurements of pulmonary nodules in ultra-low-dose CT with iterative reconstruction compared to low-dose CT. *Eur J Radiol.* 2016;85:564-570.
56. Cohen JG, Kim H, Park SB, et al. Comparison of the effects of model-based iterative reconstruction and filtered back projection algorithms on software measurements in pulmonary subsolid nodules. *Eur Radiol.* 2017;27:3266-3274.
57. Petrou M, Quint LE, Nan B, et al. Pulmonary nodule volumetric measurement variability as a function of CT slice thickness and nodule morphology. *AJR Am J Roentgenol.* 2007;188:306-312.
58. Tan Y, Guo P, Mann H, et al. Assessing the effect of CT slice interval on unidimensional, bidimensional and volumetric measurements of solid tumours. *Cancer Imaging.* 2012;12:497-505.
59. Winer-Muram HT, Jennings SG, Meyer CA, et al. Effect of varying CT section width on

volumetric measurement of lung tumors and application of compensatory equations. *Radiology.* 2003;229:184-194.

60. Zhao B, Schwartz LH, Moskowitz CS, et al. Pulmonary metastases: effect of CT section thickness on measurement--initial experience. *Radiology.* 2005;234:934-939.
61. Zhao B, Tan Y, Bell DJ, et al. Exploring intra- and inter-reader variability in uni-dimensional, bi-dimensional, and volumetric measurements of solid tumors on CT scans reconstructed at different slice intervals. *Eur J Radiol.* 2013;82:959-968.
62. Zhao B, Tan Y, Tsai WY, et al. Exploring Variability in CT Characterization of Tumors: A Preliminary Phantom Study. *Transl Oncol.* 2014;7:88-93.
63. He L, Huang Y, Ma Z, et al. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Sci Rep.* 2016;6:34921.
64. Wang Y, de Bock GH, van Klaveren RJ, et al. Volumetric measurement of pulmonary nodules at low-dose chest CT: effect of reconstruction setting on measurement variability. *Eur Radiol.* 2010;20:1180-1187.
65. Christe A, Bronnimann A, Vock P. Volumetric analysis of lung nodules in computed tomography (CT): comparison of two different segmentation algorithm softwares and two different reconstruction filters on automated volume calculation. *Acta Radiol.* 2014;55:54-61.
66. Mansoor A, Bagci U, Foster B, et al. Segmentation and Image Analysis of Abnormal Lungs at CT: Current Approaches, Challenges, and Future Trends. *Radiographics.* 2015;35:1056-1076.
67. Oliver JA, Budzevich M, Zhang GG, et al. Variability of Image Features Computed from Conventional and Respiratory-Gated PET/CT Images of Lung Cancer. *Transl Oncol.* 2015;8:524-534.
68. Honda O, Johkoh T, Sumikawa H, et al. Pulmonary nodules: 3D volumetric measurement with multidetector CT--effect of intravenous contrast medium. *Radiology.* 2007;245:881-887.
69. Rampinelli C, Raimondi S, Padrenostro M, et al. Pulmonary nodules: Contrast-enhanced volumetric variation at different CT scan delays. *AJR Am J Roentgenol.* 2010;195:149-154.
70. Gourtsoyianni S, Doumou G, Prezzi D, et al. Primary Rectal Cancer: Repeatability of Global and Local-Regional MR Imaging Texture Features. *Radiology.* 2017;284:552-561.
71. Incoronato M, Aiello M, Infante T, et al. Radiogenomic Analysis of Oncological Data: A Technical Survey. *Int J Mol Sci.* 2017;18.
72. Sommer G, Tremper J, Koenigkam-Santos M, et al. Lung nodule detection in a high-risk population: comparison of magnetic resonance imaging and low-dose computed tomography. *Eur J Radiol.* 2014;83:600-605.
73. Rasch C, Barillot I, Remeijer P, et al. Definition of the prostate in CT and MRI: a multi-observer study. *Int J Radiat Oncol Biol Phys.* 1999;43:57-66.
74. Usmani N, Sloboda R, Kamal W, et al. Can images obtained with high field strength

magnetic resonance imaging reduce contouring variability of the prostate? *Int J Radiat Oncol Biol Phys.* 2011;80:728-734.

75. Saha A, Yu X, Sahoo D, et al. Effects of MRI scanner parameters on breast cancer radiomics. *Expert Systems with Applications.* 2017;87:384-391.
76. Soher BJ, Dale BM, Merkle EM. A review of MR physics: 3T versus 1.5T. *Magn Reson Imaging Clin N Am.* 2007;15:277-290, v.
77. Donati OF, Chong D, Nanz D, et al. Diffusion-weighted MR imaging of upper abdominal organs: field strength and intervendor variability of apparent diffusion coefficients. *Radiology.* 2014;270:454-463.
78. Leach MO, Morgan B, Tofts PS, et al. Imaging vascular function for early stage clinical trials using dynamic contrast-enhanced magnetic resonance imaging. *Eur Radiol.* 2012;22:1451-1464.
79. Thoeny HC, Ross BD. Predicting and monitoring cancer treatment response with diffusion-weighted MRI. *J Magn Reson Imaging.* 2010;32:2-16.
80. Li SP, Padhani AR. Tumor response assessments with diffusion and perfusion MRI. *J Magn Reson Imaging.* 2012;35:745-763.
81. Dale BM, Braithwaite AC, Boll DT, et al. Field strength and diffusion encoding technique affect the apparent diffusion coefficient measurements in diffusion-weighted imaging of the abdomen. *Invest Radiol.* 2010;45:104-108.
82. Knobloch G, Colgan T, Wiens CN, et al. Relaxivity of Ferumoxytol at 1.5 T and 3.0 T. *Invest Radiol.* 2018;53:257-263.
83. Rohrer M, Bauer H, Mintorovitch J, et al. Comparison of magnetic properties of MRI contrast media solutions at different magnetic field strengths. *Invest Radiol.* 2005;40:715-724.
84. Ciliberto M, Kishida Y, Seki S, et al. Update of MR Imaging for Evaluation of Lung Cancer. *Radiol Clin North Am.* 2018;56:437-469.
85. Johnson KM, Fain SB, Schiebler ML, et al. Optimized 3D ultrashort echo time pulmonary MRI. *Magn Reson Med.* 2013;70:1241-1250.
86. Ohno Y, Koyama H, Yoshikawa T, et al. Standard-, Reduced-, and No-Dose Thin-Section Radiologic Examinations: Comparison of Capability for Nodule Detection and Nodule Type Assessment in Patients Suspected of Having Pulmonary Nodules. *Radiology.* 2017;284:562-573.
87. Cha MJ, Park HJ, Paek MY, et al. Free-breathing ultrashort echo time lung magnetic resonance imaging using stack-of-spirals acquisition: A feasibility study in oncology patients. *Magn Reson Imaging.* 2018;51:137-143.
88. Koyama H, Ohno Y, Aoyama N, et al. Comparison of STIR turbo SE imaging and diffusion-weighted imaging of the lung: capability for detection and subtype classification of pulmonary adenocarcinomas. *Eur Radiol.* 2010;20:790-800.
89. Koyama H, Ohno Y, Kono A, et al. Quantitative and qualitative assessment of non-contrast-enhanced pulmonary MR imaging for management of pulmonary nodules in 161 subjects.

Eur Radiol. 2008;18:2120-2131.

90. Ohno Y, Kauczor HU, Hatabu H, et al. MRI for solitary pulmonary nodule and mass assessment: Current state of the art. *J Magn Reson Imaging.* 2018;47:1437-1458.
91. Koyama H, Ohno Y, Seki S, et al. Value of diffusion-weighted MR imaging using various parameters for assessment and characterization of solitary pulmonary nodules. *Eur J Radiol.* 2015;84:509-515.
92. Bernardin L, Douglas NH, Collins DJ, et al. Diffusion-weighted magnetic resonance imaging for assessment of lung lesions: repeatability of the apparent diffusion coefficient measurement. *Eur Radiol.* 2014;24:502-511.
93. Sasaki M, Shibata E, Kanbara Y, et al. Enhancement effects and relaxivities of gadolinium-DTPA at 1.5 versus 3 Tesla: a phantom study. *Magn Reson Med Sci.* 2005;4:145-149.
94. Yoon SH, Park CM, Park SJ, et al. Tumor Heterogeneity in Lung Cancer: Assessment with Dynamic Contrast-enhanced MR Imaging. *Radiology.* 2016;280:940-948.
95. Ingrisch M, Maxien D, Schwab F, et al. Assessment of pulmonary perfusion with breath-hold and free-breathing dynamic contrast-enhanced magnetic resonance imaging: quantification and reproducibility. *Invest Radiol.* 2014;49:382-389.
96. Fink C, Ley S, Risse F, et al. Effect of inspiratory and expiratory breathhold on pulmonary perfusion: assessment by pulmonary perfusion magnetic resonance imaging. *Invest Radiol.* 2005;40:72-79.
97. Johns CS, Swift AJ, Hughes PJC, et al. Pulmonary MR angiography and perfusion imaging-A review of methods and applications. *Eur J Radiol.* 2017;86:361-370.
98. Ohno Y, Nishio M, Koyama H, et al. Dynamic contrast-enhanced CT and MRI for pulmonary nodule assessment. *AJR Am J Roentgenol.* 2014;202:515-529.
99. Ohno Y, Koyama H, Lee HY, et al. Contrast-enhanced CT- and MRI-based perfusion assessment for pulmonary diseases: basics and clinical applications. *Diagn Interv Radiol.* 2016;22:407-421.
100. Gaddikeri S, Gaddikeri RS, Tailor T, et al. Dynamic Contrast-Enhanced MR Imaging in Head and Neck Cancer: Techniques and Clinical Applications. *AJNR Am J Neuroradiol.* 2016;37:588-595.
101. Ley S, Ley-Zaporozhan J. Pulmonary perfusion imaging using MRI: clinical application. *Insights Imaging.* 2012;3:61-71.
102. Sourbron SP, Buckley DL. On the scope and interpretation of the Tofts models for DCE-MRI. *Magn Reson Med.* 2011;66:735-745.
103. Tofts PS. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. *J Magn Reson Imaging.* 1997;7:91-101.
104. Tofts PS, Brix G, Buckley DL, et al. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. *J Magn Reson Imaging.* 1999;10:223-232.
105. Ashton E, Raunig D, Ng C, et al. Scan-rescan variability in perfusion assessment of tumors in MRI using both model and data-derived arterial input functions. *J Magn Reson Imaging.*

2008;28:791-796.
106. Morgan B, Utting JF, Higginson A, et al. A simple, reproducible method for monitoring the treatment of tumours using dynamic contrast-enhanced MR imaging. *Br J Cancer.* 2006;94:1420-1427.
107. Yuan X, Zhang J, Ao G, et al. Lung cancer perfusion: can we measure pulmonary and bronchial circulation simultaneously? *Eur Radiol.* 2012;22:1665-1671.
108. Li XS, Fan HX, Fang H, et al. Value of whole-tumor dual-input perfusion CT in predicting the effect of multiarterial infusion chemotherapy on advanced non-small cell lung cancer. *AJR Am J Roentgenol.* 2014;203:W497-505.
109. Adams MC, Turkington TG, Wilson JM, et al. A systematic review of the factors affecting accuracy of SUV measurements. *AJR Am J Roentgenol.* 2010;195:310-320.
110. Kim CK, Gupta NC, Chandramouli B, et al. Standardized uptake values of FDG: body surface area correction is preferable to body weight correction. *J Nucl Med.* 1994;35:164-167.
111. Zasadny KR, Wahl RL. Standardized uptake values of normal tissues at PET with 2-[fluorine-18]-fluoro-2-deoxy-D-glucose: variations with body weight and a method for correction. *Radiology.* 1993;189:847-850.
112. Lowe VJ, DeLong DM, Hoffman JM, et al. Optimum scanning protocol for FDG-PET evaluation of pulmonary malignancy. *J Nucl Med.* 1995;36:883-887.
113. Kinahan PE, Fletcher JW. Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. *Semin Ultrasound CT MR.* 2010;31:496-505.
114. Doot RK, Scheuermann JS, Christian PE, et al. Instrumentation factors affecting variance and bias of quantifying tracer uptake with PET/CT. *Med Phys.* 2010;37:6035-6046.
115. Hofheinz F, Apostolova I, Oehme L, et al. Test-Retest Variability in Lesion SUV and Lesion SUR in (18)F-FDG PET: An Analysis of Data from Two Prospective Multicenter Trials. *J Nucl Med.* 2017;58:1770-1775.
116. Huang SC. Anatomy of SUV. Standardized uptake value. *Nucl Med Biol.* 2000;27:643-646.
117. Moon SH, Hyun SH, Choi JY. Prognostic significance of volume-based PET parameters in cancer patients. *Korean J Radiol.* 2013;14:1-12.
118. Liao S, Penney BC, Wroblewski K, et al. Prognostic value of metabolic tumor burden on 18F-FDG PET in nonsurgical patients with non-small cell lung cancer. *Eur J Nucl Med Mol Imaging.* 2012;39:27-38.
119. Satoh Y, Onishi H, Nambu A, et al. Volume-based parameters measured by using FDG-PET/CT in patients with stage I NSCLC treated with stereotactic body radiation therapy: prognostic value. *Radiology.* 2014;270:275-281.
120. Kim K, Kim SJ, Kim IJ, et al. Prognostic value of volumetric parameters measured by F-18 FDG PET/CT in surgically resected non-small cell lung cancer. *Nucl Med Commun.* 2012;33:613-620.
121. Lodge MA. Repeatability of SUV in Oncologic (18)F-FDG PET. *J Nucl Med.* 2017;58:523-532.

122. Quak E, Le Roux PY, Lasnon C, et al. Does PET SUV Harmonization Affect PERCIST Response Classification? *J Nucl Med.* 2016;57:1699-1706.
123. Lasnon C, Desmonts C, Quak E, et al. Harmonizing SUVs in multicentre trials when using different generation PET systems: prospective validation in non-small cell lung cancer patients. *Eur J Nucl Med Mol Imaging.* 2013;40:985-996.
124. Volpi S, Ali JM, Tasker A, et al. The role of positron emission tomography in the diagnosis, staging and response assessment of non-small cell lung cancer. *Ann Transl Med.* 2018;6:95.
125. Everitt SJ, Ball DL, Hicks RJ, et al. Differential (18)F-FDG and (18)F-FLT Uptake on Serial PET/CT Imaging Before and During Definitive Chemoradiation for Non-Small Cell Lung Cancer. *J Nucl Med.* 2014;55:1069-1074.
126. Vera P, Thureau S, Chaumet Riffaud P, et al. Phase II Study of a Radiotherapy Total Dose Increase in Hypoxic Lesions Identified by (18)F-Misonidazole PET/CT in Patients with Non-Small Cell Lung Carcinoma (RTEP5 Study). *J Nucl Med.* 2017;58:1045-1053.
127. Luan X, Huang Y, Gao S, et al. (18)F-alfatide PET/CT may predict short-term outcome of concurrent chemoradiotherapy in patients with advanced non-small cell lung cancer. *Eur J Nucl Med Mol Imaging.* 2016;43:2336-2342.
128. Dai D, Li XF, Wang J, et al. Predictive efficacy of (11)C-PD153035 PET imaging for EGFR-tyrosine kinase inhibitor sensitivity in non-small cell lung cancer patients. *Int J Cancer.* 2016;138:1003-1012.
129. Meng X, Loo BW, Jr., Ma L, et al. Molecular imaging with 11C-PD153035 PET/CT predicts survival in non-small cell lung cancer treated with EGFR-TKI: a pilot study. *J Nucl Med.* 2011;52:1573-1579.
130. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology.* 2016;278:563-577.
131. Parekh V, Jacobs MA. Radiomics: a new application from established techniques. *Expert Rev Precis Med Drug Dev.* 2016;1:207-226.
132. Szigeti K, Szabo T, Korom C, et al. Radiomics-based differentiation of lung disease models generated by polluted air based on X-ray computed tomography data. *BMC Med Imaging.* 2016;16:14.
133. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* 2014;5:4006.
134. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* 2017;14:749-762.
135. Wachinger C, Reuter M, Klein T. DeepNAT: Deep convolutional neural network for segmenting neuroanatomy. *Neuroimage.* 2018;170:434-445.
136. Huang Y, Liu Z, He L, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. *Radiology.* 2016;281:947-957.
137. Tixier F, Le Rest CC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant

radiochemotherapy in esophageal cancer. *J Nucl Med.* 2011;52:369-378.

138. Ganeshan B, Goh V, Mandeville HC, et al. Non-small cell lung cancer: histopathologic correlates for texture parameters at CT. *Radiology.* 2013;266:326-336.

139. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys.* 2017;44:1050-1062.

140. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. *Magn Reson Imaging.* 2012;30:1234-1248.

141. Aerts HJ, Grossmann P, Tan Y, et al. Defining a Radiomic Response Phenotype: A Pilot Study using targeted therapy in NSCLC. *Sci Rep.* 2016;6:33860.

142. Fave X, Zhang L, Yang J, et al. Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. *Sci Rep.* 2017;7:588.

143. Goh V, Ganeshan B, Nathan P, et al. Assessment of response to tyrosine kinase inhibitors in metastatic renal cell cancer: CT texture as a predictive biomarker. *Radiology.* 2011;261:165-171.

144. Rao SX, Lambregts DM, Schnerr RS, et al. CT texture analysis in colorectal liver metastases: A better way than size and volume measurements to assess response to chemotherapy? *United European Gastroenterol J.* 2016;4:257-263.

Figure Legends

Figure 1. Various methods of tumor segmentation. A) Part-solid adenocarcinoma with internal air-bronchogram at the right upper lobe. B) Automatic segmentation of the solid portion. C) Automatic segmentation of the ground glass opacity (GGO) portion. D) Semi-automatic segmentation with subjective tumor margin editing demonstrates the final solid portion (blue) and GGO portion (red).

Figure 2. Graph demonstrates the attenuation profile along a vertical line through the tumor. Tumor margins are assumed by the rapid slope of pixel values. The area of negative pixel values within the tumor suggests the presence of air-bronchogram.

Figure 3. As part-solid adenocarcinoma, 5 mm slice images (top row) show less solid portion compared to 1.25 mm slice images (bottom row). Thicker-slice image contains larger partial volume artifacts than a thinner-slice image, thus influencing the true details of lung adenocarcinoma.

Figure 4. The extended Tofts model. Assumption of this model is the equilibrium of the contrast agent between the plasma and extravascular extracellular space (EES)

Figure 5. Difference in grey-level co-occurrence matrix (GLCM) according to number of bins. The two GLCMs are from the same patient with different numbers of bins. The GLCMs are displayed using the same scale. The left figure has more bins thus can have fewer counts per bin, while the right figure has fewer bins and thus more counts per bin. These differences in counts per bin led to different features values of GLCM.

Figure 6. Histogram samples for radiomics. A) 1D Histogram within ROI. In most cases, the analysis using CT images considers a full range of HU values (i.e., 4096 bins). B) 2D histogram of grey-level co-occurrence matrix (GLCM). The 2D histogram of GLCM is built using intensity of given voxel as the first axis and intensity of the neighboring voxel as the second axis. C) 2D histogram of intensity size zone matrix (ISZM). The horizontal axis denotes intensity, and the vertical axis denotes size of a given blob.

Figure 7. Original image of lung cancer is enhanced by two different methods of Laplace of Gaussian.

Measurement Variability in Treatment Response Determination for Non-Small Cell Lung Cancer: Improvements using Radiomics

Abstract

Multimodality imaging measurements of treatment response are critical for clinical practice, oncology trials and the evaluation of new treatment modalities. The current standard for determining treatment response in non-small cell lung cancer (NSCLC) is based on tumor size using the RECIST criteria. Molecular targeted agents and immunotherapies often cause morphological change without reduction of tumor size. Therefore, it is difficult to evaluate therapeutic response by conventional methods. Radiomics is the study of cancer imaging features that are extracted using machine learning and other semantic features. This method can provide comprehensive information on tumor phenotypes and can be used to assess therapeutic response in this new age of immunotherapy. Delta radiomics, which evaluates the longitudinal changes in radiomics features, shows potential in gauging treatment response in NSCLC. It is well known that quantitative measurement methods may be subject to substantial variability due to differences in technical factors and require standardization. In this review, we describe measurement variability in the evaluation of non-small cell lung cancer and the emerging role of radiomics.

Key words: Molecular Targeted Therapy; Immunotherapy; Medical oncology; Phenotype; Lung cancer response; Image; Radiomics

Abbreviations:

ADC : apparent diffusion coefficient

AIF : arterial input function

CT : computed tomography

DCE : dynamic contrast enhanced

DWI : diffusion-weighted

EES : extravascular extracellular space

EGFR : epidermal growth factor receptor

FDG : fluoro-deoxyglucose

GGO : ground-glass opacity

GLCM : gray-level co-occurrence matrix

GRE : gradient-recalled echo

ISZM : intensity size zone matrix

LoG : Laplacian of Gaussian

MRI : magnetic resonance imaging

MTT : mean transit time

MTV : metabolic tumor volume

NEMA : National Electrical Manufacturers Association

NSCLC : non-small cell lung cancer

PBF : pulmonary blood flow

PBV :pulmonary blood volume

PET : positron emission tomography

RECIST : Response Evaluation Criteria In Solid Tumors

ROI : region of interest

RQS : radiomics quality score

SNR : signal-to-noise ratio

STIR : Short T1 inversion recovery

SUV : standardized uptake value

SUV_{bw} : normalization of standardized uptake value for patient body weight

SUV_{bsa} : normalization of standardized uptake value for body surface area

SUV_{lbm} : normalization of standardized uptake value for lean body mass

TLG : total lesion glycolysis

TSE : turbo spin-echo

UTE : ultrashort echo time

VOI : volume of interest

1 Introduction

Assessment of anti-tumor activity of cancer therapies is generally determined by an anatomical measurement of tumor burden. Since its first introduction in 2000 and subsequent revision in 2009, Response Evaluation Criteria In Solid Tumors (RECIST) has served as the reference standard for measuring tumor burden and confirming tumor response.[1] According to RECIST criteria, measurement is the maximal axial (in-plane) unidimensional measurement of a tumor's diameter. However, this conventional tumor size analysis is imperfect due to inter and intra reader measurement variability, heterogeneous tumor morphology, and different technical parameters at the time of scanning.[2] All of these factors contribute to measurement variability, which can lead to an erroneous determination of treatment response/progression and thereby misinform treatment decisons.[2,3]

During the past decade, due to an improved understanding of cancer biology, a vast collection of targeted molecular therapies have been developed. This has led to a paradigm shift in the local and systemic treatment of non-small cell lung cancer (NSCLC). While conventional chemotherapy is focused on destroying rapidly dividing tumor cells, molecular targeted therapy aims at transmembraneous receptors and intracellular molecules that are responsible for the survival and proliferation of tumor cells. Molecular targeted therapy has been shown to be effective in tumors with specific genomic driver mutations. This has opened a new era of tumor response evaluation where the limitations of RECIST-based approaches are increasingly being found.[4] Cancer immunotherapy with immune-checkpoint blockade helps to activate the cancer pateints own immune system to kill the cancer cells.[5] Cancer immunotherapy is associated with unconventional response patterns that are not accurately characterized by the RECIST criteria.[4,5] In addition, morphological changes such as tumor necrosis and cavitation without concurrent tumor size reduction are frequently observed in the setting of anti-angiogenic therapy.[2,4]

Radiomics is the process of extracting large amounts of advanced quantitative information embedded within radiological images. This approach to image analysis aids the field of oncology by providing a more quantitative approach for tumor response assessment.[6] Although state-of-the-art methods have been shown to work well for measuring tumor volume, a great deal of variability exists, and radiologists should be familiar with the technical variations, benefits, and drawbacks of radiomics regarding measurement variability. Furthermore, although computed tomography (CT) continues to play an important role, additional imaging modalities such as positron emission tomography (PET), magnetic resonance imaging (MRI) dynamic contrast enhanced (DCE) perfusion images and MRI diffusion-weighted (DWI) MRI allow for multiparametric assessment of tumor biology (e.g. glucose metabolism, tumor perfusion, and tumor hypoxia) of tumor biology.[6] Thus, by combining detailed functional and metabolic information, these protocols provide a more comprehensive depiction of the tumor microenvironment and may allow for an earlier determination of tumor response. *The purpose of the review is to focus on the technical issues regarding NSCLC tumor measurement variability and how radiomics is emerging for the early assessment of tumor response.*

2. Technical aspects of measurement

1) Measuring tumor volume

    A. Segmentation

Precise NSCLC tumor measurement between interval studies is the current basis for tumor response assessment. Currently the long axis diameter, a unidimensional measurement, remains the standard RECIST criteria for assessment of whether a tumor is growing or shrinking, discordant tumor response between primary reviewers and secondary reviewers has been reported.[3] In a previous study, using RECIST for NSCLC,

there was a significant difference between readers for unidimensional measurements of tumor size. The misclassification rates for progressive disease were 30% and 10% for interobserver and intraobserver measurements, respectively.[3] The primary reason for this variability is related to inter-reader differences in the manual measurements of the primary tumor. [3,7] For instance, in the case of a single unidimensional largest diameter measurement for RECIST, each reviewer may measure the tumor at different image slices. As a solution to this problem, most radiologists now agree that measuring the entire tumor volume is more accurate than a single unidimensional RECIST measurement.[8,9] Many recent publications have shown that volumetric measurements demonstrate better reproducibility and repeatability. According to a recent review article, although limit of agreement for both manual diameter measurements and semi-automated volume measurements lies in the same range in terms of absolute percentages, the percentage of lung nodules in which an actual inter-reader difference found was with 11% far lower for semi-automated nodule volume measurements compared to manual diameter measurements, where inter-reader variability occurs commonly. [10] Second, volumetric measurement is more sensitive in detecting even small changes than is unidimensional measurement.[11] For example, in a 10 mm spherical nodule, a 1 mm increase of unidimensional diameter corresponds to a 10% increase in cross-sectional diameter and a 33% increase in volume.[12] Finally, as lung CT post-processing computer software is becoming widespread, tumor volumetric measurements are gaining in popularity. This has now become the standard for oncological trials are beginning as their clinical response endpoint[11]. We will now discuss the factors affecting the variability of tumor volume measurement.[11]

Segmentation is the process by which humans (manual segmentation) and machines delineate tumor boundaries from the surrounding lung. Generally, the whole tumor is selected as the volume of interest (VOI), which is usually feasible, but in certain cases may be hampered due to indistinct tumor margins.[13] For example, when lung cancer is surrounded by a pathological abnormality such as post-obstructive pneumonia

or radiation-induced lung injury, the tumor boundary is frequently obscured. Tissue reorganization and post radiation therapy, scar formation disrupts accurate tumor segmentation, leading to variability in tumor measurement.

Among various methods of segmentation, automatic and semi-automatic methods using volumetric software have been shown to be more reproducible than manual segmentation.[13] Although the current "gold standard" is considered to be manual segmentation drawn by experts, this method has major drawbacks: (1) it is a time-consuming, (2) labor-intensive task and (3) has inter and intra-reader variability. In a study comparing manual and semi-automatic segmentation, the radiomics features derived from the latter demonstrated significantly higher reproducibility ($p=0.0009$; intra-class correlation coefficient values of 0.85 and 0.77 for semi-automatic segmentation and manual segmentation, respectively) and were more robust compared to those derived from manual contouring.[14] When comparing repeatability (intra-algorithm comparisons) and reproducibility (inter-algorithm comparisons) of segmentation algorithms, repeatability was significantly higher than the reproducibility ($p<0.007$; average Dice score of 0.95 and 0.81 for repeatability and reproducibility, respectively), recommending that the same software be used at all time points in longitudinal studies.[15] However, in cases of part-solid adenocarcinomas, which have a ground-glass opacity (GGO) component, fully automatic segmentation is also be problematic due to the reduced contrast between the GGO component and surrounding lung parenchyma.[16] Thus, as of today, for part-solid adenocarcinomas, semi-automatic segmentation with tumor margin editing based on subjective decision by an experienced expert remains the optimal choice for accurate volumetric assessments of NSCLC (Figure 1).[17] Likewise, advanced NSCLC lung cancer patients with large tumors having irregular margins, heterogeneous intra-tumoral texture and surrounding atelectasis or effusions often require semi-automated approach with expert radiologist manual editing of the segmentation.[9] In the setting of molecular targeted therapy for NSCLC, tumor volumes

obtained by a semi-automated approach have been shown to be a prognostic marker for improved survival, solidifying its value in this era of precision medicine.[8,18]

In terms of rapid and accurate tumor segmentation, fully automatic segmentation methods based on deep learning may be the solution. Several investigators have trained convolutional neural networks and demonstrated that deep learning is capable of performing accurate localization and segmentation of tumors in multiple organs.[19,20] Although most of these articles were based on MRI scans, such as brain, prostate, and rectum, deep learning technologies have shown potential to improve accuracy and robustness of tumor segmentation.

Another point that needs to be highlighted is the usage of different vendor volumetric software platforms. Studies comparing multiple volumetric software packages found considerable variation in nodule volume. This shows that that the results of software packages should not be used interchangeably.[21,22] Next we discuss the impact of technical factors at the time of CT acquisition and CT reconstruction such as radiation dose, iterative reconstruction, inspiration, and slice thickness on the variability in volumetric measurement.

2) Technical issues according to particular imaging modality

A. CT

Chest CT is the modality of choice in routine lung cancer imaging, and iterative reconstruction techniques have allowed for a significant reduction of radiation dose with overall similar image quality (Figure 2).[23] Quantitative analysis of CT data can provide accurate anatomical information about NSCLC and the surrounding lung. In principle, lung nodule volumetric measurement and comparison across interval CT scans is relatively easy and reproducible on the same scanner hardware. However, imagers and oncologists should keep in mind that there is variability in these “objective” CT metrics

that are introduced by individual reviewer RECIST measurements, manual segmentation of tumor volume, and technical factors (e.g. choice of reconstruction kernel, slice-thickness, and inter-scanner differences).[24] Any combination of these, factors may cause considerable measurement variability of the tumor burden, making the task more challenging for radiologists. We next carefully discuss the various technical factors that may impact tumor measurement accuracy.

a. CT Reconstruction algorithms and radiation dose

Previous studies have investigated the influence of the reconstruction kernel and radiation dose on lung nodule volume using chest phantoms.[25,26] The vast majority of those studies demonstrated that various iterative reconstructions (e.g. adaptive statistical iterative reconstruction, iDose, and model based iterative reconstruction) showed no significant variability in nodule diameter or volume measurement when compared to filtered back projection.[25,26] In fact, some studies reported that iterative reconstructions demonstrated better measurement accuracy at a reduced radiation dose. They suggested that reduced noise or increased image quality from iterative reconstruction helped reduce measurement errors.[25,26] In a study comparing lung cancer screening individuals who underwent low dose CT and ultra-low dose CT with iterative reconstruction, there was no significant difference in nodule size and volume measurement between the two protocols.[27] In a recent study comparing subsolid nodules between model-based iterative reconstruction and filtered back projection, Cohen et al. demonstrated that semi-automatic measurements of diameter, volume, and solid components of the subsolid nodules were within the range of measurement variability.[28] Thus, lung nodule volumetric measurements acquired from scans with different reconstruction techniques can be reliably compared.

b. Slice thickness and reconstruction kernel

Prior studies have investigated the impact of slice thickness on tumor measurement for cancer screening or tumor response evaluation.[29,30] Significant differences in volume according to CT slice thickness variation were noted for smaller lung nodules, where thicker slices introduced greater measurement variability.[29,30] The reason for this is related to partial volume effects. Given that a thicker CT slice contains larger partial volume artifacts than a thinner image, the margin of the tumor is blurred on thicker images. The lack of isotropic voxels for Lung CT influences NSCLC nodule segmentation and any extracted radiomics features.[31] In cases of subcentimeter nodules, which have very small VOI, partial volume artifacts substantially influences the volume measurement.[12]

Similarly, when employing radiomics, recent studies have shown that thin-slice images were better than thick-slice images for radiomics features.[31,32] In patients with lung cancer, He et al. reported that a radiomics signature based on thin slices (1.25 mm) demonstrated better diagnostic performance than when applied to thick slices (5 mm).[32] In a chest phantom study, thinner (1.25 mm and 2.5 mm) slice thickness was found to be better for radiomics features (e.g. quantifying tumor size, shape, and density).[31] To minimize measurement variability thinner slice images are recommend, slice thickness should be the same and slices of different thickness should not be mixed together for analysis (Figure 3).

Studies comparing tumor volumes at different reconstruction kernels are scarce and have conflicting results. One study reported that, compared to sharp kernels, soft tissue reconstructions demonstrated more repeatable volumetric measurements.[33] Another study reported that, compared to high frequency bone algorithms, low frequency soft algorithms demonstrated larger volumes.[34]

c. Effects of respiration and intravenous contrast

Differences in lung inflation should not be underestimated when measuring lung

nodules. For example, collapse of the alveoli at expiration may bring over and underestimation of tumor size, whereas stretching of the tumor parenchyma and blurring of the tumor margins could be responsible for an apparent larger tumor size at inspiration. Interestingly, results of significant changes in apparent tumor volume during the respiratory cycle have been previously reported.[12] Furthermore, motion artefacts during respiration can significantly affect the ability to segment lung nodules, rendering their outline and volume assessment unreliable. In addition, the presence of a pleural effusion or pneumothorax may also have a large influence on the apparent tumor volume. For radiomics, Oliver et al. suggested that approximately 75% of the current dictionary of CT radiomics features are susceptible to respiration.[35]

Another interesting point is the impact of intravenous contrast material on lung nodule volume. Due to increased attenuation of the peripheral portion of a nodule at post-contrast scans (more vascular and viable region of the tumor nodule), the contrast difference between the parenchyma and the nodule increases; thus, volume segmentation may include a greater area of the peripheral lung nodule.[22] Results from two studies showed that, although the precise increase in nodule volume was small, radiologists should be aware of this artifact on the contrast enhanced exams.[36,37]

B. MRI

Owing to the ability of MRI to gather multiparametric data from NSCLC, MRI may play an increasing role in categorizing the therapeutic response in lung cancer.[38,39] MRI is more reproducible in the identification of NSCLC and has superior soft tissue contrast in comparison to CT. MRI lung nodule volumes are smaller than CT lung nodule volumes due to the higher resolution of CT and the magnetic susceptibility of air surrounding.[40] Ideally, for quantitative analysis, MRI images should all have the same field of view and acquisition matrix, field strength, and slice thickness. Each one of these acquisition parameters which have a strong effect on signal-to-noise ratio.[39] However, the many choices for acquisition in Lung MRI complicates comparison between studies of the many

features extracted from the images.[41] Standardization of Lung MRI protocols in the setting of gathering radiomic features from NSCLC will be very helpful.

a. Magnetic field strengths

As MR field strength increases, the signal-to-noise (SNR) increases. This increase in the SNR can be utilized to for an increase in the number of phase encoded steps for better spatial resolution and improved anatomical identification.[40,42] The use of higher MR field strength improves the ability to contour tumor masses and reduces the measurement variability.[40] However, as MR field strength increases, it is accompanied by Bo and B1 inhomogeneity, an increased number of image artifacts due to changes in tissue magnetic susceptibility and increase in chemical shift,[42] B1 inhomogeneity results in systematic error for T1 measurement.[43] In oncology practice, tumor necrosis from response to anticancer therapies leads to increased water diffusion. This results in higher signal intensity on the higher-b-value images, and the apparent diffusion coefficient (ADC) value of the corresponding region will typically increase as there is no restriction to the diffusion of water with a destruction of the closely packed cell membranes. Changes in ADC on DWI have been shown to be effective for monitoring therapeutic response in solid tumore.[44] Therefore, the field strength should be considered to evaluate the therapeutic response with ADC value.

There are field strength-related changes on the relaxivity of MR contrast media. The relaxivity of gadolinium based MR contrast media increases 5% to 10% when changing from 1.5T to 3T.[42] The individual dependencies of relaxivities on field strength for the types of MR contrast media were significantly different (Table 1).[45,46] Successful treatment leads to decreased magnitude of enhancement. For detecting change of enhancement of tumor, dosing for contrast media would need to be modified according to field strength.

b. MRI Acquisition parameters for staging of lung cancer

In patients with lung cancer, MRIs have been widely used to evaluate invasion of mediastinum and organs because of superior soft tissue contrast. With advancement of MR techniques, size threshold for nodules have increased, and MR has potential role for assessing indeterminate lung nodules.[47] With short echo times, fast spin echo sequences have enabled to assess NSCLC, however, T2 blurring affected the reproducibility of the evaluation of NSCLC.[48] Ultrashort echo time (UTE) with the exceedingly short T2 and T2* relaxation times of the lungs has been used in nodule detection and nodule type classification[47,49]. However, UTE sequences have the disadvantages of long scan duration because of inefficient k-space coverage and are sensitive to motion artifacts.[50] 3D UTE provide isotropic spatial resolution with full chest coverage and is less sensitive to motion artifacts.[48] The use of limited field-of-view excitation, variable readout gradient, and radial oversampling improves image quality on 3D UTE.[48] Most 3D UTE sequences have acquired images using radial-based trajectories.[50] A recent study by Ohno et al. showed that UTE images with radial acquisition was useful in detection and classification of pulmonary nodules larger than 4 mm, and interobserver agreement for nodule classification was excellent ($\kappa = 0.95$).[49] In addition, UTEs with spiral trajectories over a radial readout are reported to have the advantage of high k-space coverage speed while preserving image quality.[50] A study reported that there was a 100% detection rate for nodules 5 mm or larger and 76.7% for 2-5 mm nodules on 3D UTE with stack-of-spirals trajectory.[50]

Short T1 inversion recovery (STIR) turbo spin-echo imaging sequence, which is very sensitive to change in T1 and T2, has been known as an important sequence in pulmonary MR imaging.[51] The specificity (60.6%) and accuracy (74.5%) of STIR were higher than T1 (37.9% and 67.9%) and T2 (48.5% and 67.9%) in distinguishing malignant from benign nodules.[51] Therefore, STIR sequence could be used to characterize lung nodule, and assess clinical stage of NSCLC. DWI could be useful in assessment of lung nodules, the staging, and early detection and prediction of treatment response of NSCLC. With high lesion-to-background ratio on high b-value images, DWI is

useful in the detection of lung nodules. In addition, DWI allows for the characterization of lung nodules using a quantitative assessment of diffusion of water molecules by calculating the ADC.[47] Thus, ADC is widely used a quantitative imaging biomarker in evaluating NSCLC. A study reported that ADC value increased by 25% after one cycle of chemotherapy due to tumor necrosis and apoptosis, and this suggest that early response of treatment can be predicted by means of ADC change.[52] However, the DWI-based evaluation of lung nodule can show the difference in value depending on the quantitative evaluation method and b value selection.[52] Due to the impact of b value selection on DWI, quantitative parameter values should be changed depend on b value selection. Susceptibility artifacts is one of the reasons for the lower ADC differentiation of lung nodule. As b value increased, the change of distortion and susceptibility artifacts increased, and results in poor SNR.[53]. The interobserver coefficient of variation of ADC in nodules less than 2 cm was relatively poor.[54]. Contrast-enhanced T1 sequence can be used to characterize lung nodules according to contrast enhancement patterns as well as difference in signal intensity before and after injection of contrast agent.[52] Contrast enhancement with gadolinium contrast agent on T1-weighted gradient-recalled echo (GRE) or turbo GRE sequences is superior to those on spin-echo and turbo spin-echo (TSE) sequences.[55].

Quantitative features than can be derived from medical images helps to evaluate of lung cancer.[6] Entropy was known as the most reproducible MR parameter reflecting tumor heterogeneity.[56] A recent study showed that histogram and texture parameters varied after contrast agent injection on DCE MRI, and the 120-150 second after contrast agent infection was optimal for analysis of MR texture parameter.[56] The effects of acquisition parameter variations on pixel signal intensities are masked because of blurring and partial volume effects, thus reducing the effect on the radiomics features. Repeatability of MR quantitative parameters is better for global features such as first-order statistical histogram and model-based fractal features than for local-regional texture parameters.[38]

c. Compensating for respiratory motion

There is an artificial increase in the volume of a solitary NSCLC during inspiration because of stretching of the tumor and the surrounding peritumoral lung parenchyma [12] Moreover, breathing-related motion can decrease the signal intensity on MRI particularly in areas of dynamic air trapping. During inspiration, lung volume is larger; thus, tissue density and MR signal are lower.[57] Perfusion could be evaluated qualitatively and quantitatively, and perfusion MRI is performed during breath hold to minimize artifacts from respiration motion because of the fast transit time of contrast agent. However, measurement of perfusion depends strongly on the level of inspiration. During inspiration, pulmonary vascular resistance is increased, while right atrial filling is increased due to the drop in intrapleural pressure. One consequence of this change in physiology is that perfusion at during a breath hold MR angiography exam performed at full inspiration (total lung capacity) is lower than perfusion performed at full expiration (residual volume). It is difficult to control the degree of inspiration during breath hold.[58] Therefore, measurements of perfusion performed with have relatively poor reproducibility.[57] Some authors have suggested that measurement of perfusion during quiet free breathing can be assessed more reproducibly because free breathing offers better patient compliance.[57]

d. Functional MR analysis

Angiogenesis is one of the important factors in the evaluation of lung cancer related to tumor survival and growth. DCE-MRI provides information of tumor angiogenesis such as blood flow, vascular volume, and permeability. Once multiple images can be acquired during the first transit or recirculation and washout of contrast medium, quantitative evaluation of contrast passage kinetics can be made.[59] Pulmonary blood flow (PBF), pulmonary blood volume (PBV) and mean transit time (MTT) can be generated by means of pixel-by-pixel analysis.[60] In addition, an arterial input function (AIF), rate of change in the concentration of contrast medium in the plasma with time, is quantified in the larger

arteries including main pulmonary.[59,61] An accurate AIF is necessary for quantitative analysis.[61] AIF allows conversion of signal time curves to concentration time curves from assumption of a linear relation between the signal intensity and the concentration of contrast agent.[59]

Quantitative evaluation of DCE-MRI is based on many pharmacokinetic models, and the Toft's and Kermode model (Toft's model, Table 2 and figure 4) are the most frequently used in DCE –MRI analysis.[61] The Toft's model was originally constructed with ignoring the effect of intravascular tracer[62]:

$$C(t) = K^{trans}\, e^{-tkep} * Ca(t) \quad [1]$$

Where "*" is convolution, and $C(t)$ and $Ca(t)$ are concentration-time curves in the tissues in interest and in the plasma of a feeding artery, respectively. The standardized terms are presented in Table 2. The parameters of K^{trans} and k_{ep} are defined as follows[62]:

$$K^{trans} = EF_p \qquad K_{ep} = EF_p / v_e \quad [2]$$

Where v_e is the fractional volume of the extravascular extracellular space (EES), F_p is the flow of plasma in the capillary bed, and E is related to F_p and the permeability-surface are permeability surface area product of the endothelial wall. The assumption of negligible plasma volume is invalid, particularly tumor tissues. To overcome the limitation, the Tofts model has been formulated to allow for an intravascular contribution, which is referred as the extended Tofts model.[62]

$$C(t) = v_p Ca(t) + K^{trans}\, e^{-tkep} * Ca(t) \quad [3]$$

v_p is the fractional volume of the plasma space.

Various factors could affect the reliability of results in DCE-MRI. The accuracy and precision of pharmacokinetic parameter estimates are strongly influenced by SNR, and temporal resolution.[61] Using a theoretical AIF, differences in injection rate and cardiac

output are ignored, which may differ between subjects and for a single subject over time.[63] Measurement of suboptimal AIF results in worse reproducibility than if a standardized AIF is used, although the AIF might be not so important for evaluating treatment response.[64] In addition, blood supply of lung takes places through both the dual pulmonary and bronchial arterial systems. Primary lung cancers are supplied by dual blood supply and the bronchial circulation plays an important role in lung cancer, especially when the size of tumor is larger.[65] The single input perfusion analysis according to the maximum slope method calculated the dominant circulation, and ignored the secondary circulation, therefore the result of perfusion of lung cancer are likely to be underestimated.[66] To overcome underestimation of perfusion, the dual-input perfusion analysis technique is employed in lung cancer perfusion analysis, and a study reported that dual-input perfusion analysis is helpful for predicting the treatment effect of multi-arterial infusion chemotherapy.[66]

C. PET

Due to its quantitative ability and ability to target cellular biology the use of PET has continuously increased for the assessment of therapeutic response in lung cancer. The most commonly used variable is the standardized uptake value (SUV) of 18F-deoxyglucose (FDG)-based quantitative PET parameters are used as radiomics features and therapeutic response criteria. Many biological and technical factors affect the measurement of SUV, which are described below.[67]

a. Normalization method for SUV calculation

SUV is calculated by activity concentration in tissue adjusted by the administered dose of radiopharmaceutical, background SUV of the blood pool and body size. Body size usually corresponds to the body weight of the patient (SUV_{bw}). However, other indexes such as lean body mass (SUV_{lbm}) or body surface area (SUV_{bsa}) can also be used. The choice of how to normalize SUV affects the measurement of SUV and how this value can

be used in comparing other studies for therapeutic response to specific agent. One disadvantage of SUV_{bw} is its known overestimation in obese patients. Both SUV_{bw} and SUV_{bsa} are less sensitive to patient weight.[68]

b. PET/CT scanner models and image acquisition/reconstruction protocol

PET/CT hardware models, image acquisition and reconstruction protocols also affect the quantitative measurement of SUV. For the performance of PET/CT scanners, the most important factors are the intrinsic resolution and detector sensitivity. These key parameters directly affect in-plane resolution and voxel size which determines the amount of partial volume artifact and, SUV variability. This is magnified in the lung bases where nodules move with respiration further adding to volume averaging artifacts and image misregistration with respect to the CT used for attenuation correction.

In the image acquisition protocol, one of most important factors is uptake time. Uptake time is defined as the time interval between the injection of the PET radiopharmaceutical and start of PET scanning. This also influences the measurement of SUV. In the case of ^{18}F- FDG, the most common uptake time is 60 min. The SUV after FDG injection continuously increases as metabolically active cells take up the glucose analogue, which is subsequently trapped.[69] Therefore, the use of a fixed uptake time is important for the consistency of SUV measurement. On the other hand total scan duration or scan mode (2D vs. 3D) does not have a significant effect on SUV accuracy.[70]

In the reconstruction protocols, the attenuation correction method, reconstruction method (analytical vs. statistical/iterative methods), and smoothing filter are major factors affecting SUV measurement. For example, increased smoothing results in decreased noise and increased bias. Increased bias will result in reduced SUV.[71]

c. Patient factors

Even with the same PET/CT protocols and within the same patient repeatability is an important issue. SUV can vary due to the biological process such as different blood

glucose and insulin levels, this leads to a high test-retest variability.[72] It is well known that plasma serum blood glucose level is inversely correlated with SUVs.[73]

d. Types of quantitative PET parameters

Most quantitative PET parameters have important problems related to measurement variability, precision and repeatability. This includes maximum SUV, average SUV, peak SUV, metabolic tumor volume (MTV), and total lesion glycolysis (TLG). Although maximum SUV is usually not affected by the determination of lesion region of interest (ROI) or VOI, in other PET parameters, ROI/VOI has significant influence. However, there is a persistent concern that maximum SUV represents a single pixel value that may not be representative of the total metabolic profile of the tumor. According to previous studies of NSCLC, MTV and TLG were better prognostic measures than maximum SUV and mean SUV,[74] suggesting that volume-based parameters of PET may have a role in providing further prognostic information.[74,75]

e. Harmonization of PET parameters

Based on the literature, the measurement variability of maximum SUV, average SUV, and peak SUV expressed as a coefficient of variation is approximately 10%.[76] Due to these measurement variabilities the harmonization of PET response criteria has been studied. For example, image reconstruction-related variability can be solved using a standardized filter such as EQ.PET.[77] To apply this kind of standardized filter for image reconstruction, it is necessary to obtain recovery coefficients according to the lesion size by National Electrical Manufacturers Association (NEMA) NU-2 phantom.[78] However, further efforts are necessary to standardize the quantitative measurement of PET parameters.

f. PET Radiotracers for lung cancer

Representative PET radiotracers for lung cancer and their clinical utilities are summarized in Table 3. FDG, a glucose analogue, is the most widely used PET radiotracer

for lung cancer. It is clinically useful for the single pulmonary nodule evaluation, initial staging, detecting recurrence, and therapy response evaluation.[79] ^{18}F-Fluorothymidine PET is good for evaluating therapy response to radiotherapy or chemotherapy early in lung cancer.[80] ^{18}F-fluoromisonidazole PET shows the hypoxic portion within the tumor, which can be used for radiotherapy planning in lung cancer by boosting radiation dose to hypoxic tumor.[81] ^{18}F-alfatide, reflecting tumor angiogenesis, has a potential to evaluate therapy response in lung cancer, although published studies were very few.[82] PET tracers targeting tumor epidermal growth factor receptor (EGFR) such as ^{11}C-PD153035 and ^{11}C-erlotinib may be applicable for therapy response evaluation to EGFR–tyrosine kinase inhibitors.[83]

3) Special considerations on radiomics analysis

A. Bin number

Radiomics analysis computes hundreds or sometimes thousands of features from the underlying imaging modalities and ROIs. The features are different from semantic features and are agnostic computational features whose formulae are defined with various parameters. Thus, for a given radiomics feature, if the associated parameter changes, the ensuing radiomics feature might change as well. Many radiomics features, noted as histogram-based features, are computed from the intensity histogram using the underlying imaging data within the ROI. Histograms are affected by binning parameters of bin width and range (Figure 5). Range is application dependent, and we typically use 4096 for CT. Many people also use a number of bins, which is range divided by bin width for the binning parameter. Using many bins allows fine differentiation between intensity values, but using too many bins leads to very narrow bin width. A narrow bin width leads to unreliable histogram estimates, as we may not have enough samples for some bins. The Freedman-Diaconis rule can be used to set bin width.[84]

In addition, computing hundreds and thousands of features from a given ROI,

can frequently lead to having too many parameters in an analytical model. This results in overfitting the of data and is an important limitation when training an artificial intelligence model with limited samples. As a general rule there should not be more features than patients in the training set. Many recent radiomics papers included hundreds or sometimes over one thousand samples.[85,86] In practice, not all the features are included in the resultant radiomics model. There are often feature selection procedure (through least absolute shrinkage and selection operator or something equivalent), where the number of features is reduced to a few (i.e., typically tens of features).[85,87] In this case, having 70-80 samples could still be adequate to avoid overfitting.

B. Texture features

Texture features are widely recognized radiomics features.[85] The most representative texture features are computed from gray-level co-occurrence matrix (GLCM) and intensity size zone matrix (ISZM). These matrices are built out of 2D histograms, which measure the frequency of a pair of observations compared to a 1D histogram, where researchers consider the frequency of one observation (e.g., intensity). GLCM measures the frequency of intensity pairs in the neighborhood, while ISZM measures the frequency of blobs with certain size and intensity. For GLCM, the 2D histogram is built using intensity of the given voxel as the first axis and intensity of the neighboring voxel as the second axis. The GLCM quantifies how intensity pairs occur in a neighbor and hence can reflect textural information. Similar to the 1D histogram case, the number of bins is a major parameter in 2D histograms. In general, the samples in a given bin for a 2D histogram is less than those in an equivalent 1D histogram as the voxels need to fill the bins spanning the 2D histogram compared to filling the 1D histogram. Due to this sparsity in the 2D histogram, researchers typically use 128/256 bins for GLCM.[88] Figure 6 shows typical 1D intensity histogram and 2D GLCM and ISZM

histograms.

The size of the ROI also affects the 1D/2D histogram measures. If the ROI is big enough to contain thousands of voxels, then the above approaches are suitable. If the ROI has a very small number of voxels (e.g. 100), then researchers need to reduce the number of bins significantly to make sure there are enough voxels occupying the bins.

C. Shape features

Shape features are important parts of radiomics analysis.[85] The shape of ROI is quantified with various formulae. ROI is composed of voxels that could be isotropic or non-isotropic. In chest imaging there is often good in-plane resolution and variable out-of-plane resolution (i.e., non-isotropic voxels). For non-isotropic voxels, the shape features are more sensitive to shape change occurring in-plane while less sensitive to shape change occurring out-of-plane. For isotropic voxels, the shape features are equally sensitive in all directions. The shape of the target ROI may change in any direction; thus, isotropic voxels are preferable over non-isotropic ones. If the imaging data is non-isotropic, we can interpolate the imaging data to make it isotropic. This interpolation makes the data smoother but also reduces the shape variability. In other words, the fine edge detail is lost. This is similar to iterative reconstruction methods in CT used to decrease patient dose. The dose is decreased, but fine detail is lost in these smoother more "plastic" appearing images.

D. Filter and Wavelet

Some researchers have applied an edge enhancement filter such as Laplacian of Gaussian (LoG) to the reconstructed image data and only then compute radiomics features from the filtered image.[89] The LoG filter has a scale parameter that controls the scale at which enhancement occurs (Figure 7). Researchers need to specify the scale

parameter to suit their intended application. The scale should be set based on image quality and the size of ROI. If researchers have poor quality image with large ROIs, large scale operations are recommended.

Some studies also apply wavelet decomposition to imaging data.[85] The imaging data are decomposed into many output data, and radiomics features are then computed from the decomposed data. There are many wavelet transforms to choose from, each with a plethora of parameters. Coiflets are widely used for their simplicity. Researchers can decompose one 3D scan into 8 3D decomposed scans in its simplest version. Different wavelet transform leads to different decomposed data and thus affects the radiomics features. Researchers should fully consider the various parameters of wavelets before applying them in their projects.

Future considerations

As radiomics features show promising benefits for quantification of lung cancer biology and response to treatment, many researchers are now paying close attention to the clinical usefulness of radiomics in oncologic studies. However, as the number of radiomics studies explodes, it should be clearly noted that the extracted radiomics features are subject to lack of precision and repeatability. Therefore, in order to homogenize evaluation criteria and reporting guidelines for radiomics, Lambin et al. proposed the radiomics quality score (RQS) (Table 4).[86] The RQS evaluates the necessary steps in radiomics analysis including 16 key components of which each is given a number of points corresponding to the importance of the respective component.[86] Major check points in the RQS are data selection, medical imaging, features extraction, exploratory analysis, and modeling. The highest possible total RQS for quantification of the overall methodology and analysis of radiomics practice is 36 points. Therefore, efforts should be made to consider RQS in future studies and to stablish collaborative foundations to control and fully realize the potential of radiomics.

Another feature that may help in tumor response evaluation is delta

radiomics.[86,90] In contrast to most radiomics studies, which are based on features extracted at a single time point (usually at the time of diagnosis), delta radiomics evaluates changes in radiomics features between interval studies. Delta-radiomics features have shown potential in predicting response or survival in patients with colorectal cancer liver metastasis, metastatic renal cell, and lung cancer.[90-92] According to a study of 107 NSCLC patients, pretreatment radiomics features were not prognostic, while texture-strength measured at the end of treatment significantly stratified high- and low-risk patients, thus suggesting the potential of delta-radiomics features.[90] Nevertheless, if delta radiomics were to be employed in clinical practice, standardization of technical factors and high reproducibility of the features remain prerequisites.

In conclusion, compared to the current RECIST version 1.1, tumor volumetric measurement and radiomics are more quantitative measures and supplement the limitations of RECIST in the current era of precision cancer therapy. Nevertheless, substantial variability can be introduced in the process of measuring tumor burden due to various technical factors. Furthermore, the increasing role of software post-processing and radiomics support the need for increased awareness of technical factors of image acquisition among radiologists. We await the incorporation of these advanced image processing metrics of Radiomics and Artificial Intelligence into the new RECIST criteria for tumor response assessment.

References

1. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45:228-247.
2. Kang H, Lee HY, Lee KS, et al. Imaging-based tumor treatment response evaluation: review of conventional, new, and emerging concepts. *Korean J Radiol*. 2012;13:371-390.
3. Erasmus JJ, Gladish GW, Broemeling L, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *J Clin Oncol*. 2003;21:2574-2582.
4. Nishino M, Tirumani SH, Ramaiya NH, et al. Cancer immunotherapy and immune-related response assessment: The role of radiologists in the new arena of cancer treatment. *Eur J Radiol*. 2015;84:1259-1268.
5. Nishino M, Ramaiya NH, Hatabu H, et al. Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol*. 2017;14:655-668.
6. Lee G, Lee HY, Park H, et al. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art. *Eur J Radiol*. 2017;86:297-307.
7. Scholten ET, de Hoop B, Jacobs C, et al. Semi-automatic quantification of subsolid pulmonary nodules: comparison with manual measurements. *PLoS One*. 2013;8:e80249.
8. Nishino M, Dahlberg SE, Cardarella S, et al. Tumor volume decrease at 8 weeks is associated with longer survival in EGFR-mutant advanced non-small-cell lung cancer patients treated with EGFR TKI. *J Thorac Oncol*. 2013;8:1059-1068.
9. Nishino M, Guo M, Jackman DM, et al. CT tumor volume measurement in advanced non-small-cell lung cancer: Performance characteristics of an emerging clinical tool. *Acad Radiol*. 2011;18:54-62.
10. Han D, Heuvelmans MA, Oudkerk M. Volume versus diameter assessment of small pulmonary nodules in CT lung cancer screening. *Transl Lung Cancer Res*. 2017;6:52-61.
11. Zhao B, Oxnard GR, Moskowitz CS, et al. A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development. *Clin Cancer Res*. 2010;16:4647-4653.
12. Plathow C, Schoebinger M, Fink C, et al. Quantification of lung tumor volume and rotation at 3D dynamic parallel MR imaging with view sharing: preliminary results. *Radiology*. 2006;240:537-545.
13. Rios Velazquez E, Aerts HJ, Gu Y, et al. A semiautomatic CT-based ensemble segmentation of lung tumors: comparison with oncologists' delineations and with the surgical specimen. *Radiother Oncol*. 2012;105:167-173.
14. Parmar C, Rios Velazquez E, Leijenaar R, et al. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One*. 2014;9:e102107.
15. Kalpathy-Cramer J, Zhao B, Goldgof D, et al. A Comparison of Lung Nodule Segmentation Algorithms: Methods and Results from a Multi-institutional Study. *J Digit Imaging*.

2016;29:476-487.
16. Ko JP, Rusinek H, Jacobs EL, et al. Small pulmonary nodules: volume measurement at chest CT—phantom study. *Radiology*. 2003;228:864-870.
17. Lassen B, Jacobs C, Kuhnigk J, et al. Robust semi-automatic segmentation of pulmonary subsolid nodules in chest computed tomography scans. *Physics in Medicine & Biology*. 2015;60:1307.
18. Nishino M, Dahlberg SE, Fulton LE, et al. Volumetric Tumor Response and Progression in EGFR-mutant NSCLC Patients Treated with Erlotinib or Gefitinib. *Acad Radiol*. 2016;23:329-336.
19. Havaei M, Davy A, Warde-Farley D, et al. Brain tumor segmentation with deep neural networks. *Medical image analysis*. 2017;35:18-31.
20. Trebeschi S, van Griethuysen JJ, Lambregts DM, et al. Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR. *Scientific reports*. 2017;7:5301.
21. Ashraf H, de Hoop B, Shaker SB, et al. Lung nodule volumetry: segmentation algorithms within the same software package cannot be used interchangeably. *Eur Radiol*. 2010;20:1878-1885.
22. Devaraj A, van Ginneken B, Nair A, et al. Use of Volumetry for Lung Nodule Management: Theory and Practice. *Radiology*. 2017;284:630-644.
23. Huber A, Landau J, Ebner L, et al. Performance of ultralow-dose CT with iterative reconstruction in lung cancer screening: limiting radiation exposure to the equivalent of conventional chest X-ray imaging. *Eur Radiol*. 2016;26:3643-3652.
24. Stoel BC, Vrooman HA, Stolk J, et al. Sources of error in lung densitometry with CT. *Invest Radiol*. 1999;34:303-309.
25. Kim H, Park CM, Chae HD, et al. Impact of radiation dose and iterative reconstruction on pulmonary nodule measurements at chest CT: a phantom study. *Diagn Interv Radiol*. 2015;21:459-465.
26. Ohno Y, Yaguchi A, Okazaki T, et al. Comparative evaluation of newly developed model-based and commercially available hybrid-type iterative reconstruction methods and filter back projection method in terms of accuracy of computer-aided volumetry (CADv) for low-dose CT protocols in phantom study. *Eur J Radiol*. 2016;85:1375-1382.
27. Sui X, Meinel FG, Song W, et al. Detection and size measurements of pulmonary nodules in ultra-low-dose CT with iterative reconstruction compared to low dose CT. *Eur J Radiol*. 2016;85:564-570.
28. Cohen JG, Kim H, Park SB, et al. Comparison of the effects of model-based iterative reconstruction and filtered back projection algorithms on software measurements in pulmonary subsolid nodules. *Eur Radiol*. 2017;27:3266-3274.
29. Petrou M, Quint LE, Nan B, et al. Pulmonary nodule volumetric measurement variability as a function of CT slice thickness and nodule morphology. *AJR Am J Roentgenol*. 2007;188:306-312.

30. Tan Y, Guo P, Mann H, et al. Assessing the effect of CT slice interval on unidimensional, bidimensional and volumetric measurements of solid tumours. *Cancer Imaging*. 2012;12:497-505.
31. Zhao B, Tan Y, Tsai WY, et al. Exploring Variability in CT Characterization of Tumors: A Preliminary Phantom Study. *Transl Oncol*. 2014;7:88-93.
32. He L, Huang Y, Ma Z, et al. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Sci Rep*. 2016;6:34921.
33. Wang Y, de Bock GH, van Klaveren RJ, et al. Volumetric measurement of pulmonary nodules at low-dose chest CT: effect of reconstruction setting on measurement variability. *Eur Radiol*. 2010;20:1180-1187.
34. Christe A, Bronnimann A, Vock P. Volumetric analysis of lung nodules in computed tomography (CT): comparison of two different segmentation algorithm softwares and two different reconstruction filters on automated volume calculation. *Acta Radiol*. 2014;55:54-61.
35. Oliver JA, Budzevich M, Zhang GG, et al. Variability of Image Features Computed from Conventional and Respiratory-Gated PET/CT Images of Lung Cancer. *Transl Oncol*. 2015;8:524-534.
36. Honda O, Johkoh T, Sumikawa H, et al. Pulmonary nodules: 3D volumetric measurement with multidetector CT--effect of intravenous contrast medium. *Radiology*. 2007;245:881-887.
37. Rampinelli C, Raimondi S, Padrenostro M, et al. Pulmonary nodules: Contrast-enhanced volumetric variation at different CT scan delays. *AJR Am J Roentgenol*. 2010;195:149-154.
38. Gourtsoyianni S, Doumou G, Prezzi D, et al. Primary Rectal Cancer: Repeatability of Global and Local-Regional MR Imaging Texture Features. *Radiology*. 2017;284:552-561.
39. Incoronato M, Aiello M, Infante T, et al. Radiogenomic Analysis of Oncological Data: A Technical Survey. *Int J Mol Sci*. 2017;18.
40. Usmani N, Sloboda R, Kamal W, et al. Can images obtained with high field strength magnetic resonance imaging reduce contouring variability of the prostate? *Int J Radiat Oncol Biol Phys*. 2011;80:728-734.
41. Saha A, Yu X, Sahoo D, et al. Effects of MRI scanner parameters on breast cancer radiomics. *Expert Systems with Applications*. 2017;87:384-391.
42. Soher BJ, Dale BM, Merkle EM. A review of MR physics: 3T versus 1.5T. *Magn Reson Imaging Clin N Am*. 2007;15:277-290, v.
43. Leach MO, Morgan B, Tofts PS, et al. Imaging vascular function for early stage clinical trials using dynamic contrast-enhanced magnetic resonance imaging. *Eur Radiol*. 2012;22:1451-1464.
44. Thoeny HC, Ross BD. Predicting and monitoring cancer treatment response with diffusion-weighted MRI. *J Magn Reson Imaging*. 2010;32:2-16.
45. Knobloch G, Colgan T, Wiens CN, et al. Relaxivity of Ferumoxytol at 1.5 T and 3.0 T. *Invest*

Radiol. 2018;53:257-263.

46. Rohrer M, Bauer H, Mintorovitch J, et al. Comparison of magnetic properties of MRI contrast media solutions at different magnetic field strengths. *Invest Radiol*. 2005;40:715-724.
47. Ciliberto M, Kishida Y, Seki S, et al. Update of MR Imaging for Evaluation of Lung Cancer. *Radiol Clin North Am*. 2018;56:437-469.
48. Johnson KM, Fain SB, Schiebler ML, et al. Optimized 3D ultrashort echo time pulmonary MRI. *Magn Reson Med*. 2013;70:1241-1250.
49. Ohno Y, Koyama H, Yoshikawa T, et al. Standard-, Reduced-, and No-Dose Thin-Section Radiologic Examinations: Comparison of Capability for Nodule Detection and Nodule Type Assessment in Patients Suspected of Having Pulmonary Nodules. *Radiology*. 2017;284:562-573.
50. Cha MJ, Park HJ, Paek MY, et al. Free-breathing ultrashort echo time lung magnetic resonance imaging using stack-of-spirals acquisition: A feasibility study in oncology patients. *Magn Reson Imaging*. 2018;51:137-143.
51. Koyama H, Ohno Y, Kono A, et al. Quantitative and qualitative assessment of non-contrast-enhanced pulmonary MR imaging for management of pulmonary nodules in 161 subjects. *Eur Radiol*. 2008;18:2120-2131.
52. Ohno Y, Kauczor HU, Hatabu H, et al. MRI for solitary pulmonary nodule and mass assessment: Current state of the art. *J Magn Reson Imaging*. 2018;47:1437-1458.
53. Koyama H, Ohno Y, Seki S, et al. Value of diffusion-weighted MR imaging using various parameters for assessment and characterization of solitary pulmonary nodules. *Eur J Radiol*. 2015;84:509-515.
54. Bernardin L, Douglas NH, Collins DJ, et al. Diffusion-weighted magnetic resonance imaging for assessment of lung lesions: repeatability of the apparent diffusion coefficient measurement. *Eur Radiol*. 2014;24:502-511.
55. Sasaki M, Shibata E, Kanbara Y, et al. Enhancement effects and relaxivities of gadolinium-DTPA at 1.5 versus 3 Tesla: a phantom study. *Magn Reson Med Sci*. 2005;4:145-149.
56. Yoon SH, Park CM, Park SJ, et al. Tumor Heterogeneity in Lung Cancer: Assessment with Dynamic Contrast-enhanced MR Imaging. *Radiology*. 2016;280:940-948.
57. Ingrisch M, Maxien D, Schwab F, et al. Assessment of pulmonary perfusion with breath-hold and free-breathing dynamic contrast-enhanced magnetic resonance imaging: quantification and reproducibility. *Invest Radiol*. 2014;49:382-389.
58. Fink C, Ley S, Risse F, et al. Effect of inspiratory and expiratory breathhold on pulmonary perfusion: assessment by pulmonary perfusion magnetic resonance imaging. *Invest Radiol*. 2005;40:72-79.
59. Johns CS, Swift AJ, Hughes PJC, et al. Pulmonary MR angiography and perfusion imaging-A review of methods and applications. *Eur J Radiol*. 2017;86:361-370.
60. Ohno Y, Koyama H, Lee HY, et al. Contrast-enhanced CT- and MRI-based perfusion assessment for pulmonary diseases: basics and clinical applications. *Diagn Interv Radiol*.

2016;22:407-421.
61. Gaddikeri S, Gaddikeri RS, Tailor T, et al. Dynamic Contrast-Enhanced MR Imaging in Head and Neck Cancer: Techniques and Clinical Applications. *AJNR Am J Neuroradiol*. 2016;37:588-595.
62. Sourbron SP, Buckley DL. On the scope and interpretation of the Tofts models for DCE-MRI. *Magn Reson Med*. 2011;66:735-745.
63. Ashton E, Raunig D, Ng C, et al. Scan-rescan variability in perfusion assessment of tumors in MRI using both model and data-derived arterial input functions. *J Magn Reson Imaging*. 2008;28:791-796.
64. Morgan B, Utting JF, Higginson A, et al. A simple, reproducible method for monitoring the treatment of tumours using dynamic contrast-enhanced MR imaging. *Br J Cancer*. 2006;94:1420-1427.
65. Yuan X, Zhang J, Ao G, et al. Lung cancer perfusion: can we measure pulmonary and bronchial circulation simultaneously? *Eur Radiol*. 2012;22:1665-1671.
66. Li XS, Fan HX, Fang H, et al. Value of whole-tumor dual-input perfusion CT in predicting the effect of multiarterial infusion chemotherapy on advanced non-small cell lung cancer. *AJR Am J Roentgenol*. 2014;203:W497-505.
67. Adams MC, Turkington TG, Wilson JM, et al. A systematic review of the factors affecting accuracy of SUV measurements. *AJR Am J Roentgenol*. 2010;195:310-320.
68. Kim CK, Gupta NC, Chandramouli B, et al. Standardized uptake values of FDG: body surface area correction is preferable to body weight correction. *J Nucl Med*. 1994;35:164-167.
69. Lowe VJ, DeLong DM, Hoffman JM, et al. Optimum scanning protocol for FDG-PET evaluation of pulmonary malignancy. *J Nucl Med*. 1995;36:883-887.
70. Kinahan PE, Fletcher JW. Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. *Semin Ultrasound CT MR*. 2010;31:496-505.
71. Doot RK, Scheuermann JS, Christian PE, et al. Instrumentation factors affecting variance and bias of quantifying tracer uptake with PET/CT. *Med Phys*. 2010;37:6035-6046.
72. Hofheinz F, Apostolova I, Oehme L, et al. Test-Retest Variability in Lesion SUV and Lesion SUR in (18)F-FDG PET: An Analysis of Data from Two Prospective Multicenter Trials. *J Nucl Med*. 2017;58:1770-1775.
73. Huang SC. Anatomy of SUV. Standardized uptake value. *Nucl Med Biol*. 2000;27:643-646.
74. Liao S, Penney BC, Wroblewski K, et al. Prognostic value of metabolic tumor burden on 18F-FDG PET in nonsurgical patients with non-small cell lung cancer. *Eur J Nucl Med Mol Imaging*. 2012;39:27-38.
75. Kim K, Kim SJ, Kim IJ, et al. Prognostic value of volumetric parameters measured by F-18 FDG PET/CT in surgically resected non-small-cell lung cancer. *Nucl Med Commun*. 2012;33:613-620.
76. Lodge MA. Repeatability of SUV in Oncologic (18)F-FDG PET. *J Nucl Med*. 2017;58:523-532.

77. Quak E, Le Roux PY, Lasnon C, et al. Does PET SUV Harmonization Affect PERCIST Response Classification? *J Nucl Med*. 2016;57:1699-1706.
78. Lasnon C, Desmonts C, Quak E, et al. Harmonizing SUVs in multicentre trials when using different generation PET systems: prospective validation in non-small cell lung cancer patients. *Eur J Nucl Med Mol Imaging*. 2013;40:985-996.
79. Volpi S, Ali JM, Tasker A, et al. The role of positron emission tomography in the diagnosis, staging and response assessment of non-small cell lung cancer. *Ann Transl Med*. 2018;6:95.
80. Everitt SJ, Ball DL, Hicks RJ, et al. Differential (18)F-FDG and (18)F-FLT Uptake on Serial PET/CT Imaging Before and During Definitive Chemoradiation for Non-Small Cell Lung Cancer. *J Nucl Med*. 2014;55:1069-1074.
81. Vera P, Thureau S, Chaumet-Riffaud P, et al. Phase II Study of a Radiotherapy Total Dose Increase in Hypoxic Lesions Identified by (18)F-Misonidazole PET/CT in Patients with Non-Small Cell Lung Carcinoma (RTEP5 Study). *J Nucl Med*. 2017;58:1045-1053.
82. Luan X, Huang Y, Gao S, et al. (18)F-alfatide PET/CT may predict short-term outcome of concurrent chemoradiotherapy in patients with advanced non-small cell lung cancer. *Eur J Nucl Med Mol Imaging*. 2016;43:2336-2342.
83. Dai D, Li XF, Wang J, et al. Predictive efficacy of (11)C-PD153035 PET imaging for EGFR-tyrosine kinase inhibitor sensitivity in non-small cell lung cancer patients. *Int J Cancer*. 2016;138:1003-1012.
84. Parekh V, Jacobs MA. Radiomics: a new application from established techniques. *Expert Rev Precis Med Drug Dev*. 2016;1:207-226.
85. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.
86. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14:749-762.
87. Huang Y, Liu Z, He L, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. *Radiology*. 2016;281:947-957.
88. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys*. 2017;44:1050-1062.
89. Aerts HJ, Grossmann P, Tan Y, et al. Defining a Radiomic Response Phenotype: A Pilot Study using targeted therapy in NSCLC. *Sci Rep*. 2016;6:33860.
90. Fave X, Zhang L, Yang J, et al. Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. *Sci Rep*. 2017;7:588.
91. Goh V, Ganeshan B, Nathan P, et al. Assessment of response to tyrosine kinase inhibitors in metastatic renal cell cancer: CT texture as a predictive biomarker. *Radiology*. 2011;261:165-171.
92. Rao SX, Lambregts DM, Schnerr RS, et al. CT texture analysis in colorectal liver metastases: A better way than size and volume measurements to assess response to chemotherapy? *United European Gastroenterol J*. 2016;4:257-263.

Figure Legends

Figure 1. Various methods of tumor segmentation. A) Part-solid adenocarcinoma with internal air-bronchogram at the right upper lobe. B) Automatic segmentation of the solid portion. C) Automatic segmentation of the ground glass opacity (GGO) portion. D) Semi-automatic segmentation with subjective tumor margin editing demonstrates the final solid portion (blue) and GGO portion (red).

Figure 2. Graph demonstrates the attenuation profile along a vertical line through the tumor. Tumor margins are assumed by the rapid slope of pixel values. The area of negative pixel values within the tumor suggests the presence of air-bronchogram.

Figure 3. As part-solid adenocarcinoma, 5 mm slice images (top row) show less solid portion compared to 1.25 mm slice images (bottom row). Thicker-slice image contains larger partial volume artifacts than a thinner-slice image, thus influencing the true details of lung adenocarcinoma.

Figure 4. The extended Tofts model. Assumption of this model is the equilibrium of the contrast agent between the plasma and extravascular extracellular space (EES)

Figure 5. Difference in grey-level co-occurrence matrix (GLCM) according to number of bins. The two GLCMs are from the same patient with different numbers of bins. The GLCMs are displayed using the same scale. The left figure has more bins thus can have fewer counts per bin, while the right figure has fewer bins and thus more counts per bin. These differences in counts per bin led to different features values of GLCM.

Figure 6. Histogram samples for radiomics. A) 1D Histogram within ROI. In most cases, the analysis using CT images considers a full range of HU values (i.e., 4096 bins). B) 2D histogram of grey-level co-occurrence matrix (GLCM). The 2D histogram of GLCM is built using intensity of given voxel as the first axis and intensity of the neighboring voxel as the second axis. C) 2D histogram of intensity size zone matrix (ISZM). The horizontal axis denotes intensity, and the vertical axis denotes size of a given blob.

Figure 7. Original image of lung cancer is enhanced by two different methods of Laplace of Gaussian.

Figure 1

Click here to download Figure Figure 1.tif

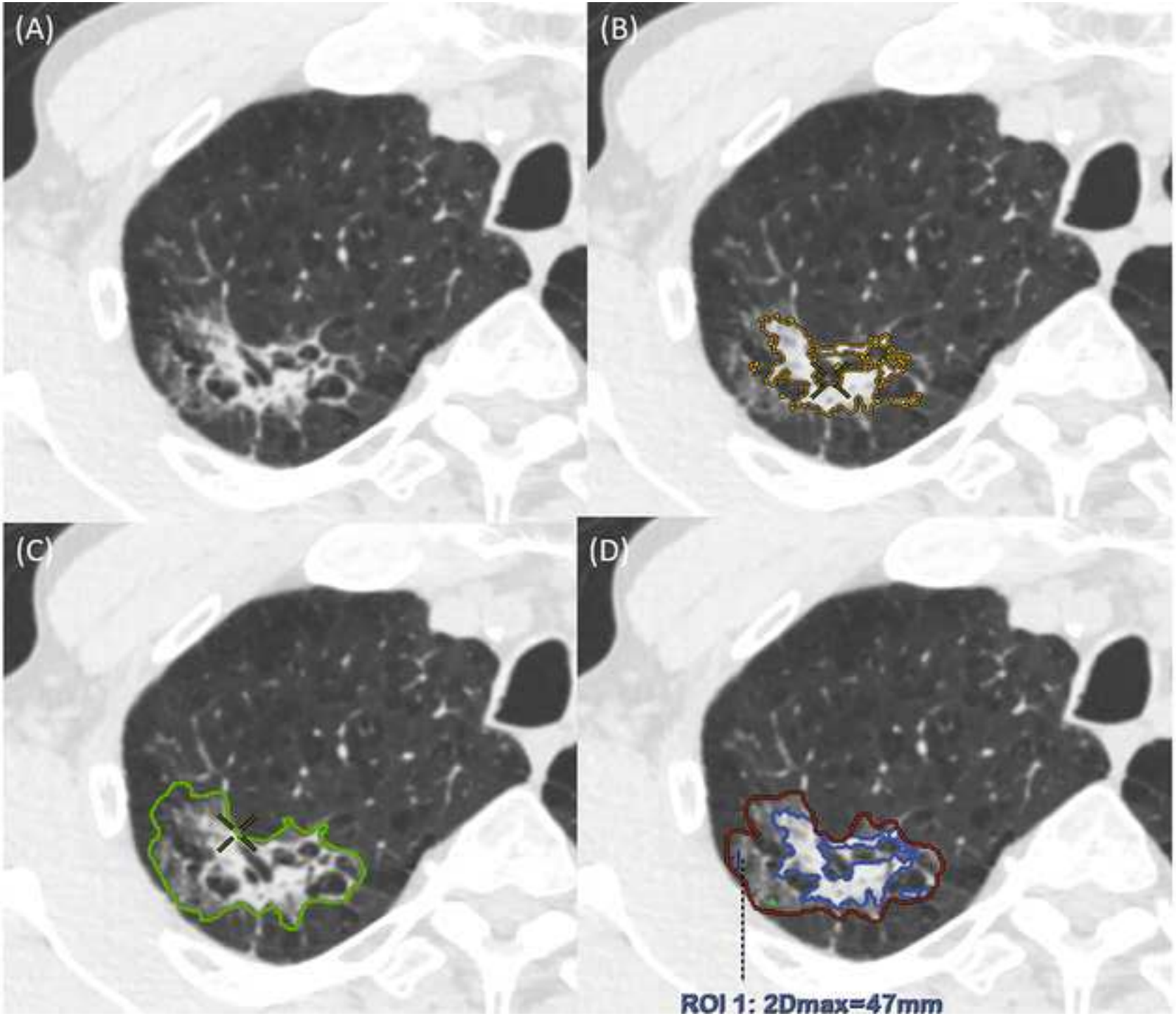

Figure 2

Click here to download Figure Figure 2.tif

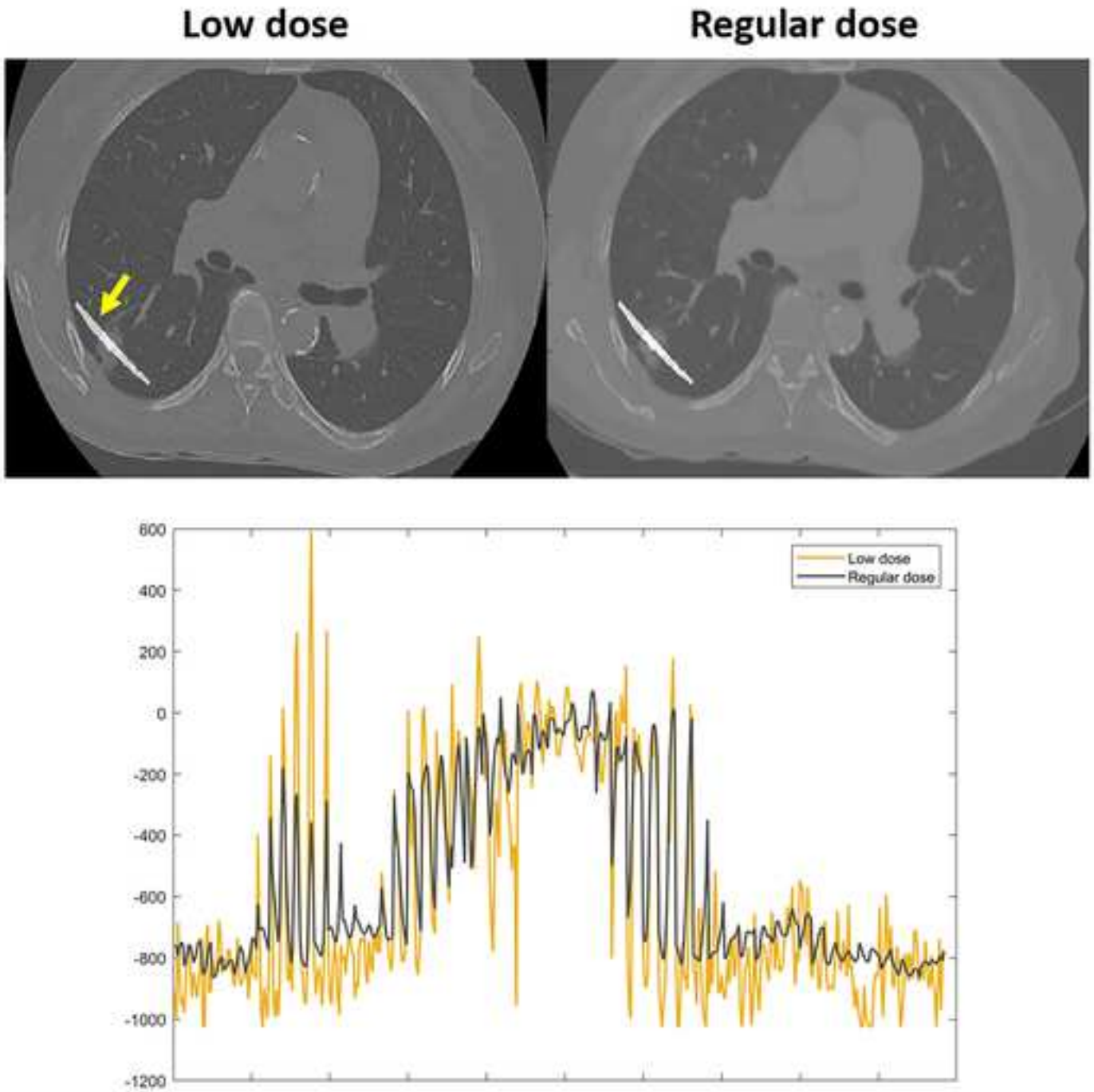

Figure 3

Click here to download Figure Figure 3.tif

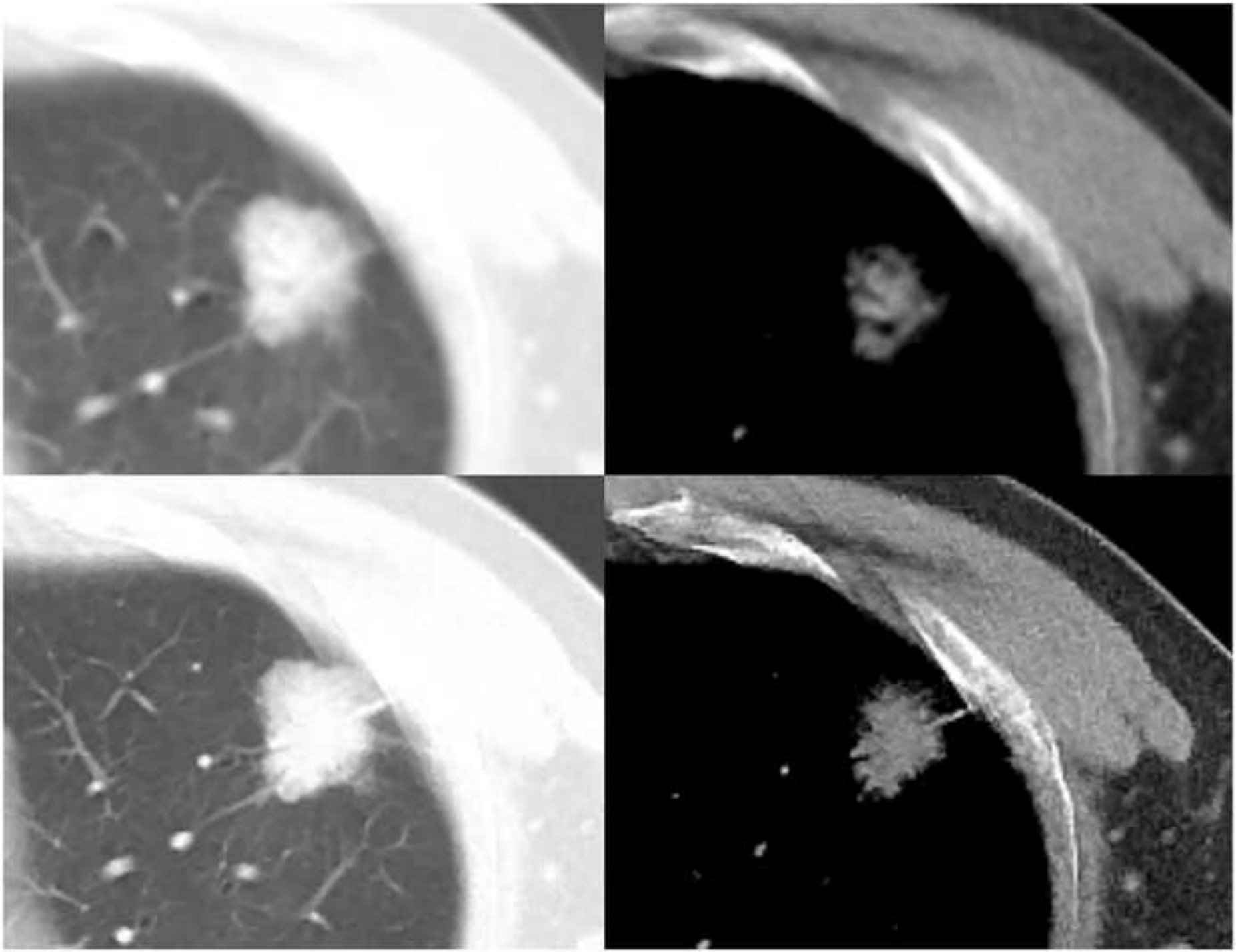

Figure 4

Click here to download Figure Figure 4-final.tif

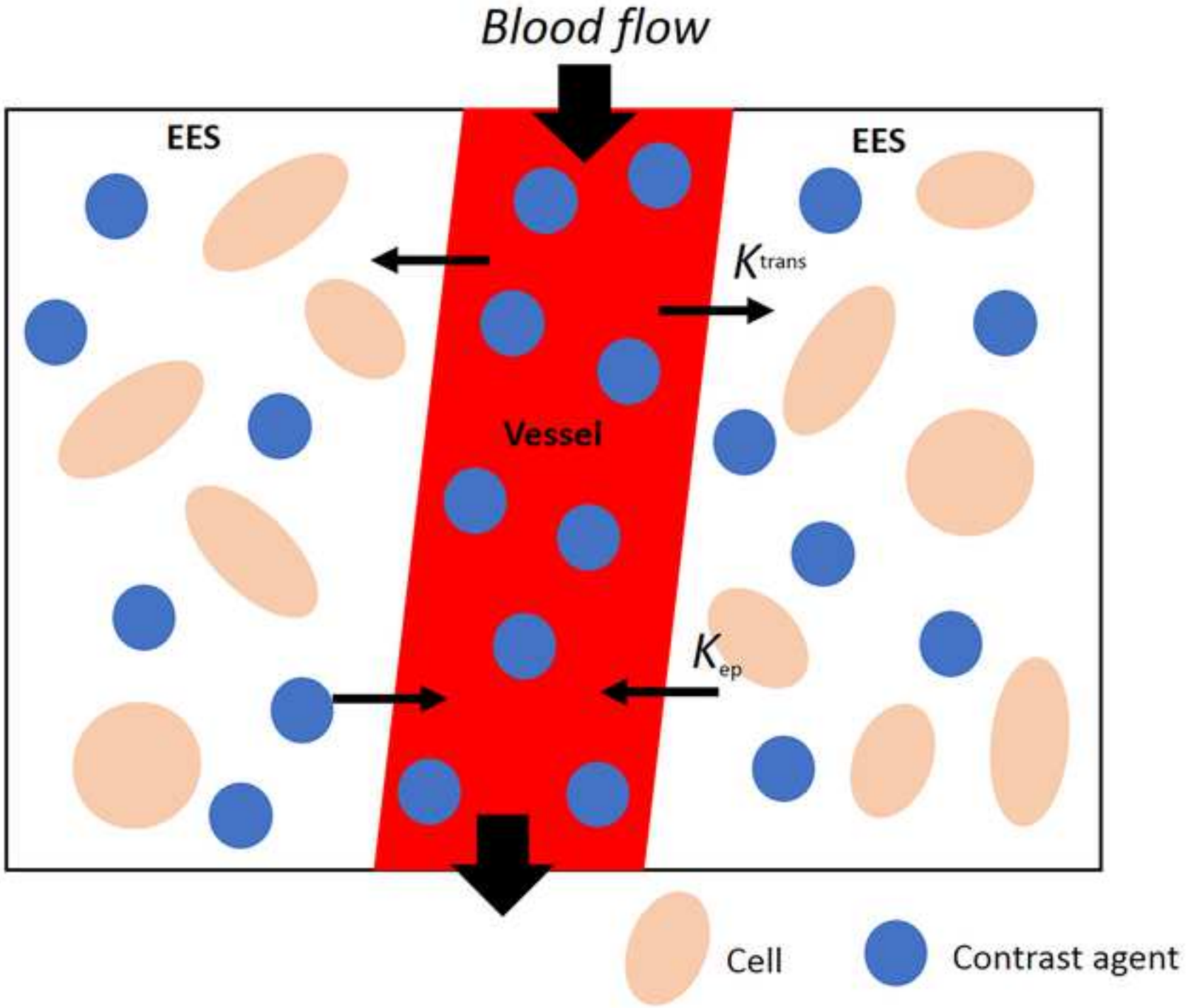

Figure 5

Click here to download Figure Figure 5.tif

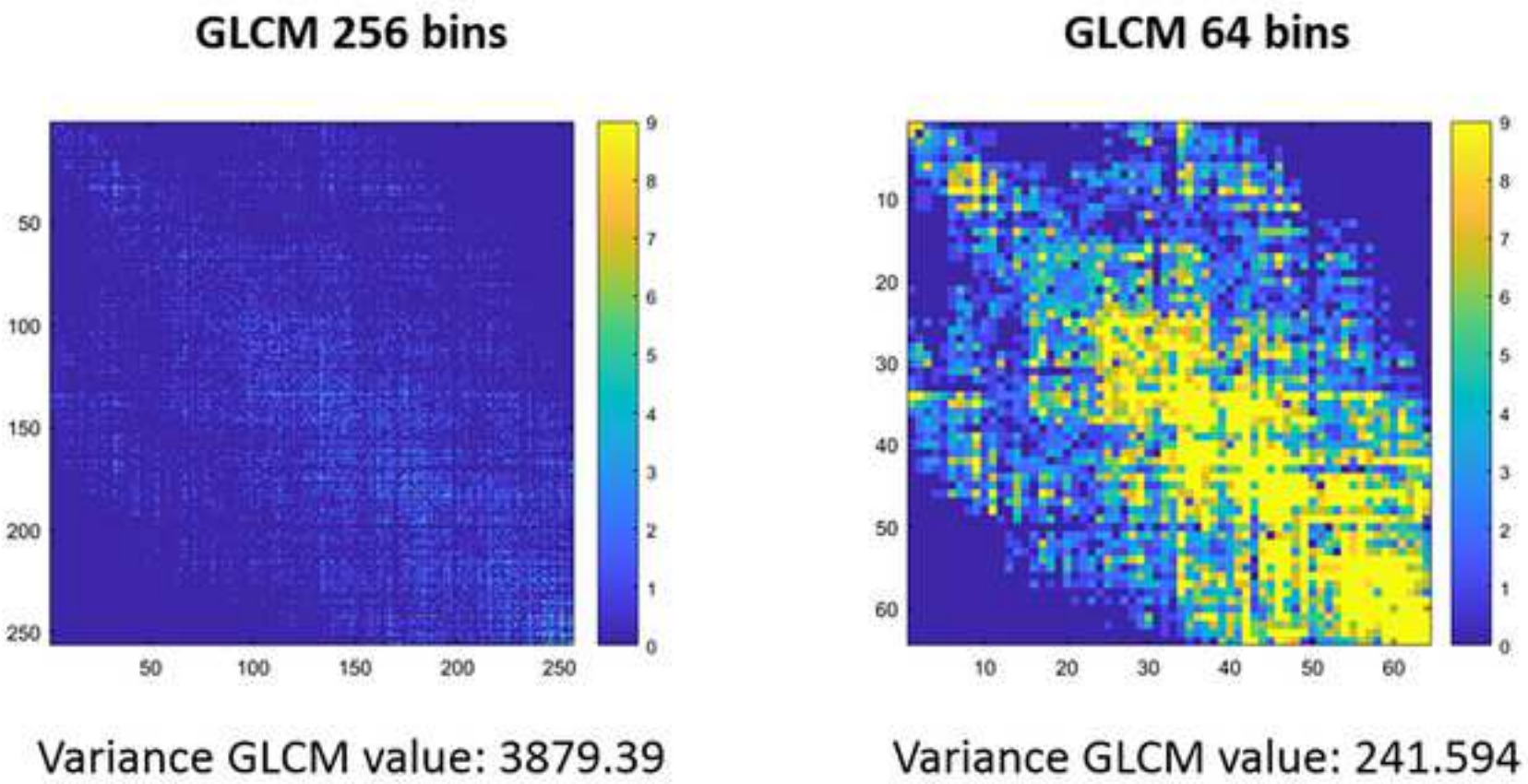

Figure 6

Click here to download Figure Figure 6.tif

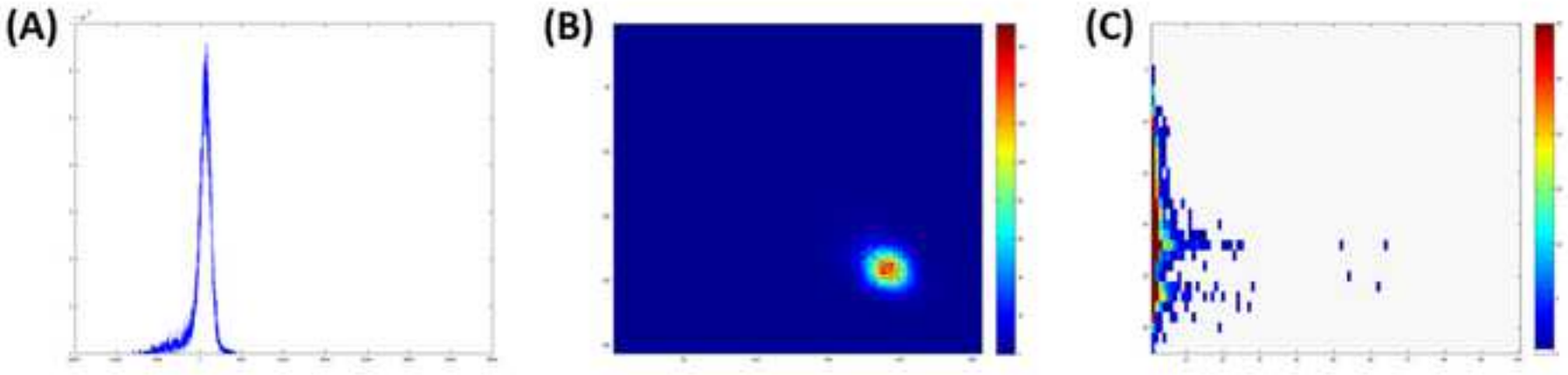

Figure 7

Click here to download Figure Figure 7.tif

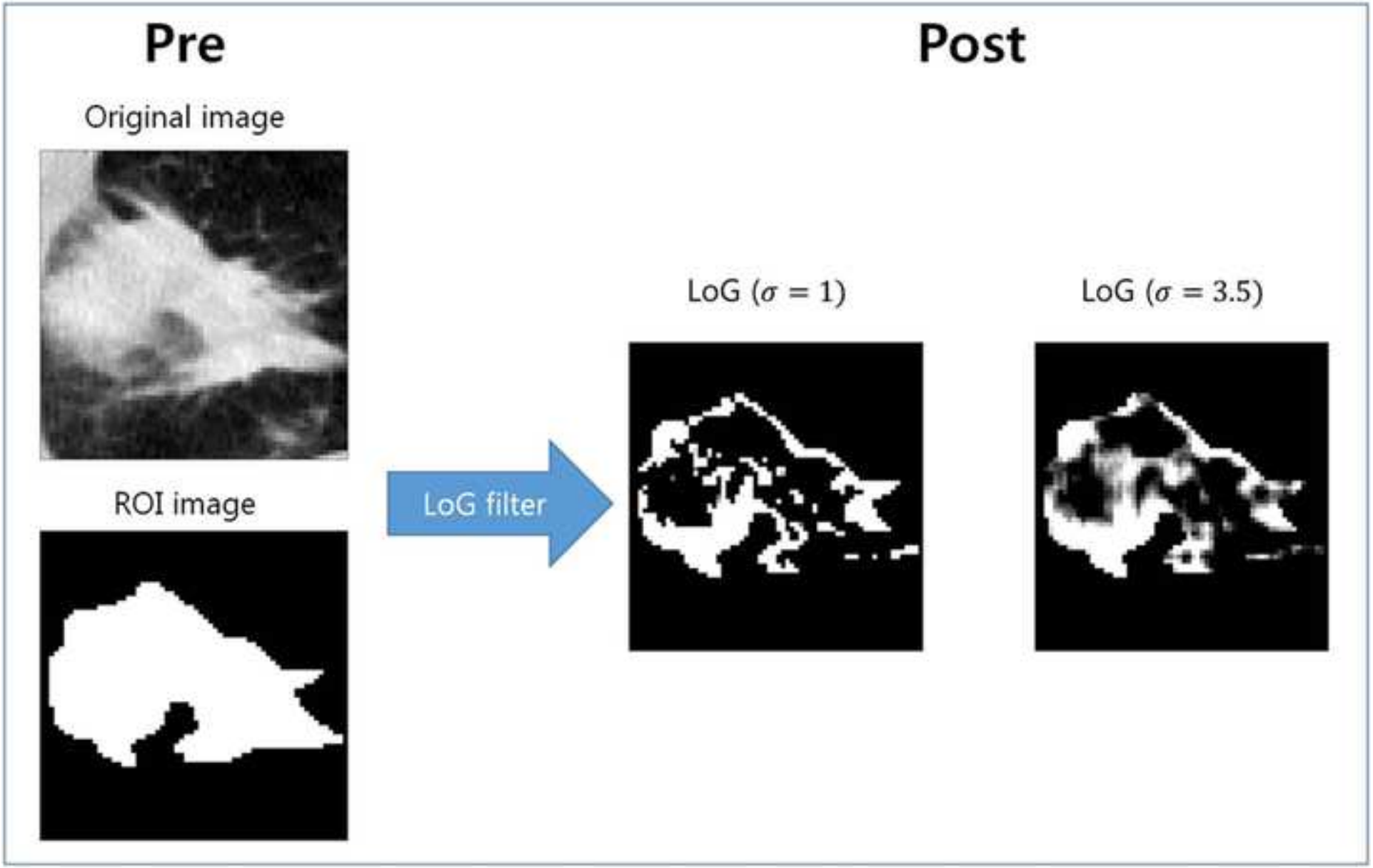

Table Legends.

Table 1. Dependence of gadolinium based contrast agent and Ferumoxytol relaxivity on field strength.[82,83]

Short name or Internal Code	Generic Name	1.5T		3T	
		R1	R2	R1	R2
Gd-BOPTA	Gadobenate dimeglumine	6.0-6.6	7.8-9.6	5.2-5.8	10.0-12.0
Gd-HP-DO3A	Gadoteridol	3.9-4.3	4.2-5.8	3.5-3.9	4.8-6.6
Gd-DO3A-butrol	Gadobutrol	4.9-5.5	5.2-7.0	4.7-5.3	6.2-8.0
Gd-DTPA	Gadopentetate dimeglumine	3.9-4.3	3.8-5.4	(3.5-3.9	4.3-6.1
Gd-DOTA	Gadoterate meglumine	3.4-3.8	3.4-5.2	3.3-3.7	4.0-5.8
Gd-DTPA-BMA	Gadodiamide	4.0-4.6	4.2-6.2	3.8-4.2	4.7-6.5
Code 7228*	Ferumoxytol	17.3-20.7	63.1-66.7	9.3-9.7	63.4-67.0

*Superparamagnetic iron oxide MR contrast agent

Table 2. Variables of Tofts equation

Quantity	Definition	Unit
C(t)	Tissue concentration as a function of time	mM
Ca(t)	Concentration in plasma as a function of time	mM
K^{trans}	Transfer constant from the blood plasma into the EES	mL/g/min
K_{ep}	Transfer constant from the EES back to the blood plasma	1/min
v_p	Volume fraction of the plasma space	none
v_e	Volume fraction of the EES	none

EES = extravascular extracellular space

Table 3. Representative PET radiotracers for lung cancer

PET radiotracers	Uptake mechanism	Clinical utility
FDG	Tumor glucose metabolism	Diagnosis, initial staging, detecting recurrence, and therapy response evaluation
^{18}F-Fluorothymidine	Tumor cell proliferation	Early therapy response evaluation
^{18}F-fluoromisonidazole	Tumor hypoxia	Radiotherapy planning
^{18}F-alfatide	Tumor angiogenesis	Therapy response evaluation
^{11}C-PD153035, ^{11}C-erlotinib	Tumor epidermal growth factor receptor (EGFR)	Therapy response evaluation to EGFR–tyrosine kinase inhibitors

Table 4. Criteria of radiomics quality score (RQS) and corresponding points.

Criteria	Score
Image protocol quality	+1 or +2
Multiple segmentation	+1
Phantom study	+1
Imaging at multiple time points	+1
Feature reduction or adjustment for multiple testing	-3 or +3
Multivariable analysis	+1
Biological correlates	+1
Cut-off analysis	+1
Discrimination statistics	+1 or +2
Calibration statistics	+1 or +2
Prospective study	+7
Validation	-5 to +5
Comparison to 'gold standard'	+2
Potential clinical applications	+2
Cost-effectiveness analysis	+1
Open science and data	+1 to +4
RQS total	36