Quantifying the reproducibility of lung ventilation images between 4-Dimensional Cone Beam CT and 4-Dimensional CT.

Henry C. Woodruff[1], Chun-Chien Shieh[1], Fiona Hegi-Johnson[1, 2, 3], Paul J. Keall[1], and John Kipritidis[1,4]

[1]Radiation Physics Laboratory, School of Medicine, University of Sydney, Sydney NSW, Australia, 2006

[2]Department of Medical Physics, School of Mathematical and Physical Sciences, University of Newcastle, Newcastle, NSW, Australia, 2300

[3]Radiation Oncology Centre, Seventh Day Adventist Hospital, Wahroonga, NSW, Australia, NSW 2076

[4]Northern Sydney Cancer Center, Royal North Shore Hospital, Sydney NSW 2065, Australia

Email of corresponding author: john.kipritidis@health.nsw.gov.au

ABSTRACT

Purpose: Deformable image registration (DIR) based computed tomography ventilation imaging derived from four-dimensional cone-beam CT ($CTVI^{4DCBCT}$) can complement existing 4DCT-based methods ($CTVI^{4DCT}$) to track lung function changes over a course of lung cancer radiation therapy. However the accuracy of $CTVI^{4DCBCT}$ needs to be assessed since anatomic 4DCBCT has demonstrably poor image quality and small field of view (FOV) compared to treatment planning 4DCT. We perform a direct comparison between short interval $CTVI^{4DCBCT}$ and $CTVI^{4DCT}$ pairs to understand the patient specific image quality factors affecting the intermodality CTVI reproducibility in the clinic.

Methods and Materials: We analysed 51 pairs of 4DCBCT and 4DCT scans acquired within one day of each other for 9 lung cancer patients. To assess the impact of image quality, CTVIs were derived from 4DCBCT scans reconstructed using both standard Feldkamp-Davis-Kress backprojection ($CTVI^{4DCBCT}_{FDK}$) and an iterative McKinnon-Bates Simultaneous Algebraic Reconstruction Technique ($CTVI^{4DCBCT}_{MKBSART}$). Also, the influence of FOV was assessed by deriving CTVIs from 4DCT scans that were cropped to a similar FOV as the 4DCBCT scans ($CTVI^{4DCT}_{crop}$), or uncropped ($CTVI^{4DCT}_{uncrop}$). All CTVIs were derived by evaluating the Jacobian determinant of deformation between the exhale and inhale phases. Reproducibility between corresponding $CTVI^{4DCBCT}$ and $CTVI^{4DCT}$ pairs was quantified using the voxel-wise Spearman rank correlation $\bar{r}$ and the Dice similarity coefficient (DSC) for ventilation defect regions (identified as the lower quartile of ventilation values). Mann-Whitney U-tests were applied to determine statistical significance of each reconstruction and cropping condition.

Results: The (mean ± SD) Spearman correlation between $CTVI^{4DCBCT}_{FDK}$ and $CTVI^{4DCT}_{uncrop}$ was $\bar{r} = 0.60 \pm 0.23$ (range -0.03 – 0.88) and the DSC was 0.64 ± 0.12 (0.34 – 0.83). By comparison, correlations between $CTVI^{4DCBCT}_{MKBSART}$ and $CTVI^{4DCT}_{uncrop}$ showed a small but statistically significant improvement with $\bar{r}$ = 0.64 ± 0.20 (range 0.06 – 0.90, p = 0.03) and DSC = 0.66 ± 0.13 (0.31– 0.87, p=0.02). Intermodal correlations were noted to decrease with an increasing fraction of lung truncation in 4DCBCT relative to 4DCT, albeit not significantly (Pearson correlation R = 0.58, p = 0.002).

Conclusions: This study demonstrates that DIR based CTVIs derived from 4DCBCT can exhibit reasonable to good voxel-level agreement with CTVIs derived from 4DCT. These correlations outperform previous cross-modality comparisons between 4DCT-based ventilation and nuclear medicine. The use of 4DCBCT scans with iterative reconstruction and minimal lung truncation is recommended to ensure better reproducibility between 4DCBCT- and 4DCT-based CTVIs.

Keywords: lung radiation therapy, functional imaging, ventilation, 4D cone beam CT, deformable image registration.

1 INTRODUCTION

In recent years a number of studies have investigated functionally adaptive lung cancer radiation therapy treatments guided by in-vivo lung ventilation imaging including single-photon-emission computed tomography (SPECT)[1] and positron emission tomography (PET)[2]. With the advent of respiratory correlated four-dimensional CT (4DCT) scans as part of the routine lung cancer radiation therapy workflow, new methods have been developed to quantify lung ventilation in terms of respiratory-related lung motion using deformable image registration (DIR)[3, 4]. CT ventilation imaging using 4DCT ($CTVI^{4DCT}$) has been shown to correlate with other ventilation imaging methods including ^{68}Ga PET[5, 6], ^{99m}Tc SPECT[7], and ^{3}He MRI[8]. $CTVI^{4DCT}$ demonstrates significant dosimetric correlations in terms of treatment-induced ventilation changes[9] and has recently been applied clinically to achieve functionally adaptive treatments[10].

Recently, our group reported on CTVI derived from daily 4D cone beam CT scans obtained in the treatment room ($CTVI^{4DCBCT}$)[11], which can complement $CTVI^{4DCT}$ by providing information on short timescale functional variations caused by patient breathing variability, changes in patient anatomy or radiation damage[12]. The need for adaptive ventilation guidance is highlighted by Meng *et al.*[1] who showed that 9 out of 15 patients receiving conventional lung cancer radiation therapy experienced >20% changes in ventilated lung volume between pre- and mid-treatment ventilation SPECT scans. Meanwhile Vinogradskiy *et al.*[12] observed weekly 4DCT-based CTVIs showing a pattern of increasing ventilation for patients with shrinking tumour volumes. It is clear from these two studies that the magnitude and direction of lung ventilation changes over a course of lung cancer radiation therapy is patient dependent and should ideally be monitored to ensure effective ventilation guidance. In order for $CTVI^{4DCBCT}$ to be applied clinically, it is necessary to demonstrate equivalence with $CTVI^{4DCT}$ as measured at a similar point in time. In an ideal imaging scenario where the image quality of 4DCBCT approaches that of 4DCT, and where a patient's breathing pattern is highly reproducible, then one might consider a comparison of DIR motion fields between 4DCBCT and 4DCT (and any derived ventilation quantities) to be trivial. But in the clinical imaging scenario, breathing can be far from regular and 4DCBCT image quality can be far from ideal. 4DCBCT has well-known image quality limitations, namely scatter, relatively small field of view (FOV) and projection aliasing ("reconstruction streaking") the latter of which is linked to projection undersampling and irregular breathing[13-16]. Given that the reproducibility of $CTVI^{4DCT}$ is known to be reduced by additive image noise in 4DCT[16, 17], we can anticipate that image artifacts in 4DCBCT will similarly reduce the accuracy of motion fields (and ventilation) derived from an intensity-based DIR. It is therefore of interest to perform a direct comparison between $CTVI^{4DCBCT}$ and $CTVI^{4DCT}$ generated under clinical treatment planning and delivery scenarios.

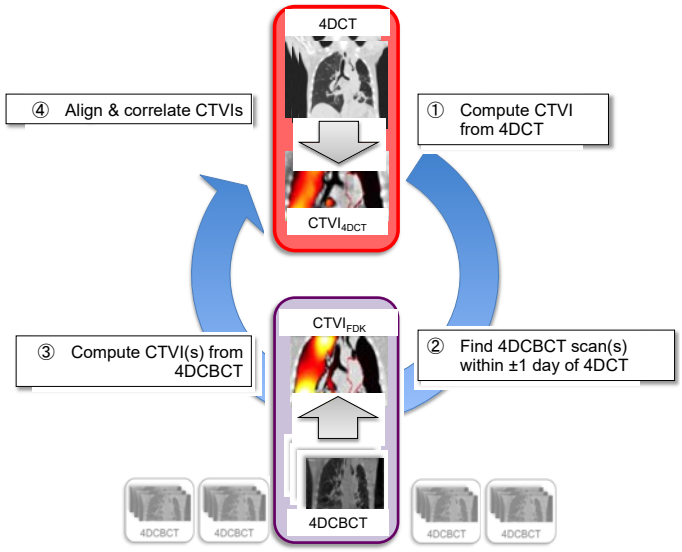
In this study we perform a direct comparison between $CTVI^{4DCBCT}$ and $CTVI^{4DCT}$ using 'short-interval' 4DCBCT and 4DCT scan pairs acquired within one day of each other for 11 lung cancer patients. We assess the intermodality reproducibility using two metrics which are dominant in the CT ventilation literature: (1) the Spearman rank correlation evaluated across all spatially matched lung voxels, and (2) the Dice similarity between functional defect structures corresponding to the lowest ventilation quartile[10, 17, 18]. Our focus on "ranked"

ventilation values is in line with both clinical[10] and in-silico[18] implementations of CTVI guided treatment planning, which prioritize the sparing of high-ventilation lung regions based on: (1) the functional rank of each specific voxel, or (2) a threshold between "defect" and "non-defect" regions set at a pre-defined rank (e.g. 20^{th} or 25^{th} ventilation percentile). In order to investigate the impact of 4DCBCT image quality, we consider two different 4DCBCT reconstruction algorithms with different values of the structure similarity index (SSI), which is an objective method for assessing perceptual image quality[19]. We additionally test the impact of limited FOV by analyzing 4DCT scans that have been cropped to the same FOV as 4DCBCT. A schematic workflow for this study is shown in Figure 1(a).

It is notable that the short-interval reproducibility between $CTVI^{4DCBCT}$ and $CTVI^{4DCT}$ can vary markedly from patient-to-patient and this is evidenced when comparing the single best and worst cases of intermodality reproducibility in this study; see Figure 1(b). This work represents an important step towards understanding the key patient-specific image quality parameters affecting reproducibility between 4DCBCT- and 4DCT-based CTVIs. Importantly this work also provides new insight into to the possible contra-indications for the use of $CTVI^{4DCBCT}$ guidance in lung cancer radiation therapy.

2 METHODS AND MATERIALS

a) Study schematic (for each 4DCT scan):



b) Best and worst cases in terms of Spearman correlation (*r*):

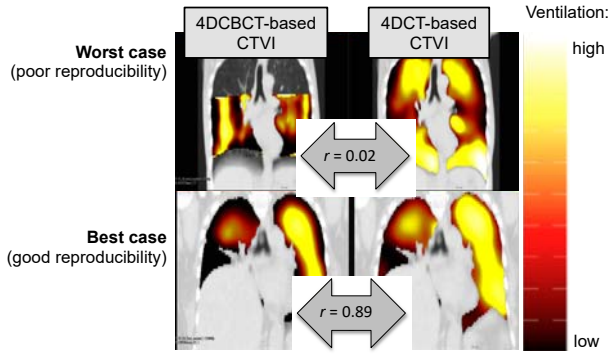


Figure 1 a) Schematic workflow of this current study. b) Examples of the best and the worst patient cases in terms of the Spearman correlation *r* between $CTVI^{4DCBCT}$ and $CTVI^{4DCT}$. In each panel the CTVI (amber colorwash) is superimposed on the time-averaged anatomic 4DCT. The *r*-values refer to all spatially matched lung voxels between the paired CTVIs.

2.1 Patient data and image acquisition

We analysed a total of 81 4DCBCT/4DCT scan pairs acquired between 2008 and 2012 for 13 locally advanced non-small cell lung cancer patients in an institutional review board approved study at Virginia Commonwealth University (VCU)[20, 21]. We excluded eight 4DCT scans (including all 4DCT scans for two patients) due to image quality considerations as described in Section 2.3. The details of the 4DCBCT and 4DCT acquisitions have been described in a previous publication[11].

Briefly, each patient received between one and five 4DCT scans over 4-6 weeks of lung cancer radiation therapy. For each 4DCT scan, at least one 4DBCT scan was acquired within ± 1 day, with the patient imaged in the treatment position either directly before or after each delivered fraction corresponding to a total of 30 4DCT scans and 51 4DCBCT scans. All 4DCT scans were acquired on a Brilliance Big Bore scanner (Philips Medical Systems) and reconstructed into 10 breathing phase bins using a 512x512 matrix with $0.98mm^2$ pixels and slice thickness 3 mm. 4DCBCT scans were performed in a research mode, with 2400 half-fan

projections acquired in 8 minutes using an On-Board Imager (Varian Medical Systems) and 360-degree gantry rotation. Audio-visual biofeedback was used for all scans to reduce breathing irregularities.

2.2 4DCBCT image reconstruction

In order to determine the impact of 4DCBCT image quality on the intermodality CTVI reproducibility, we employed two different 4DCBCT reconstruction methods: (i) the clinical standard Feldkamp-Davis-Kress (FDK) back projection method and (ii) an in-house reconstruction method that combines the McKinnon-Bates (MKB) algorithm[22] and the Simultaneous Algebraic Reconstruction Technique (SART)[23], which we refer to as MKB-SART. The MKB-SART method is an iterative algorithm, which was anticipated to produce 4DCBCT reconstructions with reduced image noise and less severe reconstruction streaking artefacts compared to FDK. Similar to the 4DCT scans, all 4DCBCT scans were reconstructed into 10 breathing phase bins with a 448x448x220 matrix with 0.88mm^2 square pixels and slice thickness 2mm.

2.3 Image quality assessment of 4DCBCT and 4DCT scans

To ensure that the 4DCT scans represented a "gold standard" comparator for 4DCBCT, we performed a visual assessment of the 4DCT scans to identify and exclude any cases where severe irregular breathing artifacts (e.g. anatomic truncation and duplication[15]) appeared in two or more phase images. We excluded 8 4DCT scans (corresponding to 30 4DCBCT/4DCT scan pairs including all scans for 2 patients), with the remaining 51 scan pairs used for CTVI generation. For these scans a quantitative image quality assessment was performed using the structure similarity index (SSI; see ref.[19]). The advantage of SSI with respect to techniques such as mean square error or signal to noise ratio is that these approaches estimate absolute errors while SSI is a perception-based model that considers image degradation as perceived change in structural information. The SSI is computed between each 4D phase image and the 4D time-average of each scan, where it can assume values between 1 (maximum image similarity) and 0 (no similarity).

2.4 CTVI Generation and DIR analysis

The CTVI^{4DCBCT} and CTVI^{4DCT} were all derived using our in-house ventilation image software VESPIR (VEntilation via Scripted Pulmonary Image Registration)[24]. For each 4DCBCT and 4DCT scan, B-spline DIR was used to compose a motion field between the exhale and inhale phase images by composing the individual DIR motion fields obtained between each neighbouring phase pair. The parameters for each DIR were similar to our previous 4DCBCT study[11] and were similar for both 4DCBCT and 4DCT.

Each DIR motion field, u, was chosen to minimize a cost function $C(u) = C_{MSE}(u) + \lambda \times C_{Reg}(u)$, which models the elastic nature of lung deformation. Here, C_{MSE} is the mean square error (MSE) intensity difference between the fixed and deformed moving image, whereas C_{Reg} enforces regularization (or "smoothness") of the motion based on the Laplacian of u. The parameter λ is a user-selected scalar value that controls the relative strength of the C_{MSE} and C_{Reg} terms. Too-small values of λ may cause the DIR to be driven by image noise/streaking, whereas too-large values may underestimate the tissue motion (see Appendix A in ref.[11]). To understand the impact of λ on the ventilation images, we tested $\lambda = 1$ and 5 for both 4DCBCT and 4DCT, reflecting the values used in our previous CTVI studies.

We note that the DIR cost function was not limited to lung voxels but rather was calculated across the whole of each fixed and moving image. This is because the poor image quality in

4DCBCT precluded accurate threshold-based lung segmentation. A visual check of all DIR results was performed by comparing the alignment of lung structures across all ten phase-images both before and after deformable registration. Where the apex of the diaphragm was visible, we checked that the DIR motion field vector direction matched the motion visible in the reconstructed 4D scans. We additionally performed a quantitative anatomic landmark-based DIR evaluation as described in Sec. 2.6.3. All 4DCBCT lung masks were derived from 4DCT lung masks using the intermodal alignment DIR as described in Section 2.5.

After composing the DIR motion field between exhale and inhale, the motion field Jacobian determinant was taken as a surrogate for regional ventilation as originally proposed by Reinhardt *et al.*[4]. The ventilation at each voxel was thus calculated as CTVI = (Jacobian - 1). We checked the CTVIs for any significant regions of negative Jacobian determinant values, which would indicate non-physical folding of tissue.

2.5 Alignment and segmentation of CTVIs

Each corresponding $CTVI^{4DCBCT}$ and $CTVI^{4DCT}$ pair was aligned and segmented using a procedure similar to our previous 4DCBCT study [11]. Briefly, each time-averaged 4DCT scan was first segmented using an active surface method driven by an intensity-based energy minimization functional implemented in the software package 'ITK-SNAP'[25]. The time-averaged 4DCT scans were then aligned to the corresponding time-averaged 4DCBCT using a combination of manual translations and (automated) rigid and deformable registrations using a mutual information image similarity metric. These registrations were also used to propagate the lung segmentation from the $CTVI^{4DCT}$ to the $CTVI^{4DCBCT}$. Based on this alignment procedure, we were additionally able to crop each 4DCT scan to a similar FOV as the corresponding 4DCBCT. In order to establish the impact of FOV on CTVI reproducibility, we then repeated the CTVI generation procedure based on the cropped 4DCTs.

2.6 Analysis of CTVI reproducibility

In this study we compare CTVIs generated from 4DCBCT scans (reconstructed using either the FDK or MKB-SART methods) versus CTVIs generated from treatment planning 4DCT (which are either cropped to the same FOV as the 4DCBCT, or uncropped). We therefore denote all CTVIs according to the 4DCBCT image reconstruction method or the 4DCT cropping condition: $CTVI^{4DCBCT}_{FDK}$, $CTVI^{4DCBCT}_{MKB}$, $CTVI^{4DCT}_{crop}$ or $CTVI^{4DCT}_{uncrop}$. The intermodality reproducibility between the 4DCBCT- and 4DCT-based CTVIs is quantified as follows:

2.6.1 Spearman correlation

For each 4DCBCT reconstruction method, FOV condition and regularization λ value we report the average Spearman correlation $\bar{r}$ between corresponding $CTVI^{4DCBCT}$ and $CTVI^{4DCT}$ pairs for all spatially matched lung voxels. The Spearman correlation takes a range of values between -1 and 1 for the case of oppositely ranked and identically ranked ventilation distributions respectively. Distributions of *r*-values were compared using Mann-Whitney U-tests [26] in order to compute the statistical significance of different 4DCBCT reconstruction methods and 4DCT cropping conditions. The null hypothesis was that these parameters do not influence the correlation between $CTVI^{4DCBCT}$ and $CTVI^{4DCT}$.

2.6.2 Dice coefficients

To further assess the relative distribution of lung function between $CTVI^{4DCBCT}$ and $CTVI^{4DCT}$, each ventilation image was divided into functional quartiles similar to the

approach by Castillo et al.[17] and binary maps produced for the lowest quartile of ventilation values representing the "defect" region. The Dice similarity coefficient (DSC) between defect regions was thus determined for corresponding $CTVI^{4DCBCT}$ and $CTVI^{4DCT}$ pairs. The DSC is a measure of the degree of overlap between two areas or volumes and is quantified as the ratio of twice the volume of intersection to the sum of the two volumes[27, 28].

2.6.3 Quantitative DIR evaluation

To explore any possible links between CTVI reproducibility and DIR accuracy, we performed quantitative evaluation of the DIR for both 4DCBCT and 4DCT scans using an intensity based MSE method, as well as an anatomic landmark-based method.

The MSE approach focused on the interphase DIR used to generate CTVIs. Specifically we evaluated the fraction of MSE intensity differences between lung voxels of the inhale/exhale phase images that were specifically resolved by the DIR (ΔMSE_{DIR}). The ΔMSE_{DIR} was calculated by comparing (i) the MSE in the lung between the exhale and inhale images before DIR (MSE_{before}) and (ii) the MSE in the lung between the exhale and deformed inhale images after DIR (MSE_{afer}). The ΔMSE_{DIR} was then expressed as a percentage difference

$$\Delta \mathrm{MSE}_{\mathrm{DIR}} = \frac{MSE_{before} - MSE_{after}}{MSE_{before}} * 100.$$

The landmark-based method computed the target registration error (TRE) for a set of computer-identified anatomic landmarks between the registered images. To evaluate the interphase DIR used to generate the CTVIs, landmarks were selected between each corresponding exhale and inhale phase image pair before DIR. The DIR motion field determinant was then used to warp the exhale landmarks to the inhale reference frame and determine the TRE after DIR. To evaluate the DIR used for the 4DCT/4DCBCT alignment, landmarks were selected from the time-averaged 4DCT and 4DCBCT images after DIR. For all of the landmarking studies, we applied the adaptive scale invariant feature transform (SIFT) algorithm which was implemented in a separate study by Paganelli *et al.*[29] and which has been validated both for 4DCT and for CT to CBCT registration[30]. All landmarks selected were within the lung parenchyma.

3 RESULTS

3.1 Impact of 4DBCT image quality on CTVI reproducibility

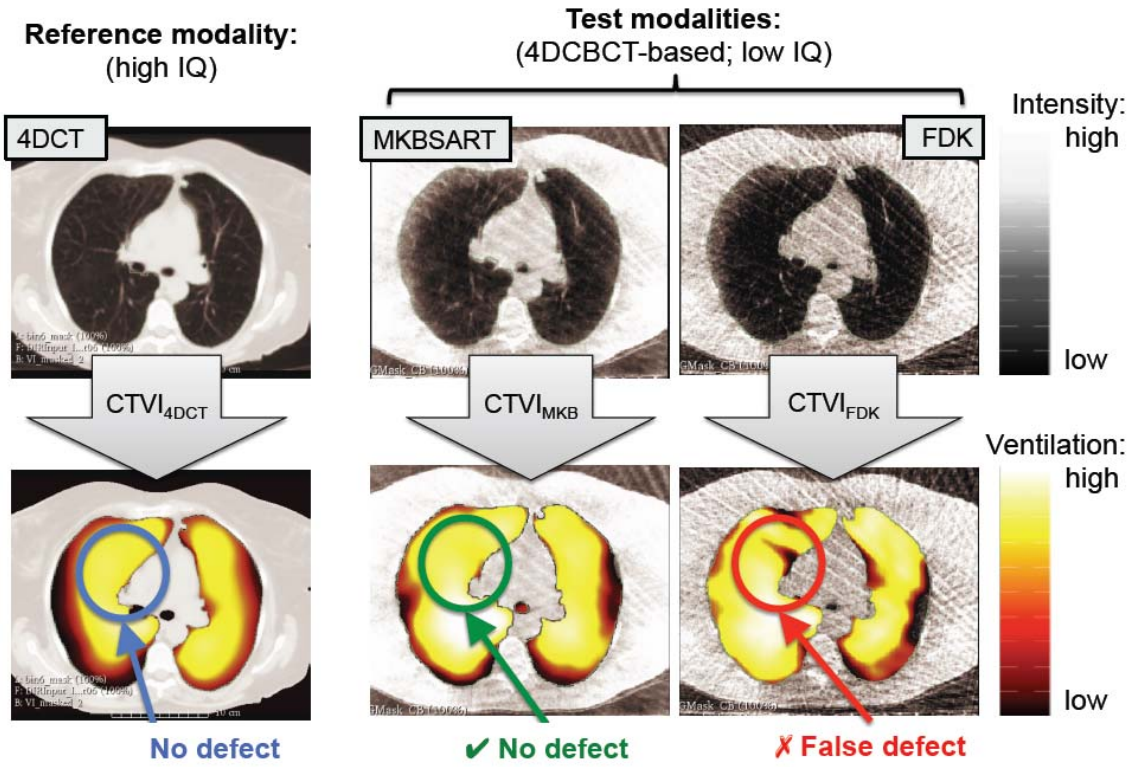


Figure 2 Effect of image quality on CTVI reproducibility. The circles indicate a region of lung that is ventilated in the reference CTVI (left column), but may appear as a false defect depending on the choice of 4DCBCT reconstruction method (middle and right columns). In this case the DIR regularization parameter λ was set at 1 and 5 for the 4DCT- and 4DCBCT-based CTVIs, respectively. Here IQ = image quality.

Figure 2 demonstrates the impact of 4DCBCT image quality on the reproducibility between 4DCBCT- and 4DCT-based CTVIs. The upper row of panels shows the central axial slice of 4DCT scan and the corresponding 4DCBCT scan reconstructed using both FDK and MKB-SART methods. Visual assessment of the anatomic MKB-SART image shows a reduction of reconstruction streaking at the cost of increased image blur and loss of contrast compared to FDK. The lower row demonstrates visually how imaging artefacts such as streaking in 4DCBCT reconstructed using the FDK method can lead to false defects in the CTVI, which are not apparent in the CTVI derived from 4DCBCT reconstructed with MKB-SART. As can be seen in Figure 3(a), we found that a linear model did not fit the data distribution for the CTVI correlation as measured by Spearman *r f*or either 4DCBCT reconstruction method, (Pearson correlation R = -0.24 for MKB-SART and -0.28 for FDK, both with $p > 0.4$). In Figure 3(b) we can see a similar result for the amount of overlap between defect regions as measured by the DSC (R = -0.20 for MKB-SART and -0.21 for FDK, both with $p > 0.1$). The (mean ± SD) SSI for all scans comparing the 3D time averaged volume to the all respiratory bins was 0.56 ± 0.06 for our in-house MKBSART reconstruction and 0.28 ± 0.02 for FDK, indicating better image quality for MKBSART reconstructions.

Table 1: Comparing 4DCBCT- and 4DCT-based CTVIs in terms of Spearman *r*-values and DSC for ventilation defect regions, across all combinations of 4DCBCT reconstruction method, 4DCT cropping condition and DIR regularization λ. Each mean value corresponds to 51 scan pairs (11 patients). The grey shaded cells show the only case were the $CTVI_{MKB}^{4DCBCT}$ did not show improved reproducibility than $CTVI_{FDK}^{4DCBCT}$.

Spearman correlations (mean ± SD)		**λ=1**		**λ=5**	
		$CTVI_{MKBSART}^{4DCBCT}$	$CTVI_{FDK}^{4DCBCT}$	$CTVI_{MKBSART}^{4DCBCT}$	$CTVI_{FDK}^{4DCBCT}$
λ=1	$CTVI_{uncrop}^{4DCT}$	0.51 ± 0.24	0.51 ± 0.24	0.64 ± 0.20	0.60 ± 0.23
	$CTVI_{crop}^{4DCT}$	0.56 ± 0.20	0.53 ± 0.24	0.63 ± 0.19	0.60 ± 0.21
λ=5	$CTVI_{uncrop}^{4DCT}$	0.50 ± 0.21	0.45 ± 0.22	0.64 ± 0.20	0.58 ± 0.23
	$CTVI_{crop}^{4DCT}$	0.56 ± 0.21	0.52 ± 0.23	0.62 ± 0.18	0.59 ± 0.20
Dice similarity coefficients (mean ± SD)					
λ=1	$CTVI_{uncrop}^{4DCT}$	0.51 ± 0.24	0.51 ± 0.24	0.64 ± 0.20	0.60 ± 0.23
	$CTVI_{crop}^{4DCT}$	0.56 ± 0.20	0.53 ± 0.24	0.63 ± 0.19	0.60 ± 0.21
λ=5	$CTVI_{uncrop}^{4DCT}$	0.50 ± 0.21	0.45 ± 0.22	0.64 ± 0.20	0.58 ± 0.23
	$CTVI_{crop}^{4DCT}$	0.56 ± 0.21	0.52 ± 0.23	0.62 ± 0.18	0.59 ± 0.20

Table 1 shows the (mean ± SD) Spearman correlations and DSC between common lung voxels of the 4DCBCT-based ventilation images ($CTVI_{FDK}^{4DCBCT}$ or $CTVI_{MKB}^{4DCBCT}$) and corresponding 4DCT-based ventilation images ($CTVI_{crop}^{4DCT}$ or $CTVI_{uncrop}^{4DCT}$) for different values of the DIR λ. We observed that $CTVI_{MKB}^{4DCBCT}$ delivered significantly better Spearman correlations with $CTVI^{4DCT}$, and smaller variance, than $CTVI_{FDK}^{4DCBCT}$; $p < 0.05$ for all combination of λ and cropping, with the exception for the combination shaded grey in Table 1; the DSC show a similar, yet weaker correlation. Table 1 also shows as an example the moderate yet significant (p=0.03) improvement of mean Spearman correlations between $CTVI_{uncrop}^{4DCT}$ (λ = 1) and $CTVI_{MKB}^{4DCBCT}$ (λ = 5) when compared to $CTVI_{FDK}^{4DCBCT}$ (λ = 5). Within each reconstruction method we found no clear correlation of intermodality reproducibility with the SSI; a linear model does not fit the data for either reconstruction method separately (Pearson's R = 0.24 and 0.28 for MKB-SART and FDK, respectively, $p > 0.4$). Increasing λ from 1 to 5 for the 4DCBCT DIR process showed a statistically significant improvement in correlation with 4DCT-based CTVI irrespective of the reconstruction method or λ. Decreasing λ for the 4DCT DIR from 5 to 1 either improves correlation with 4DCBCT-based CTVI for certain combinations of reconstruction method, or causes no change in mean values.

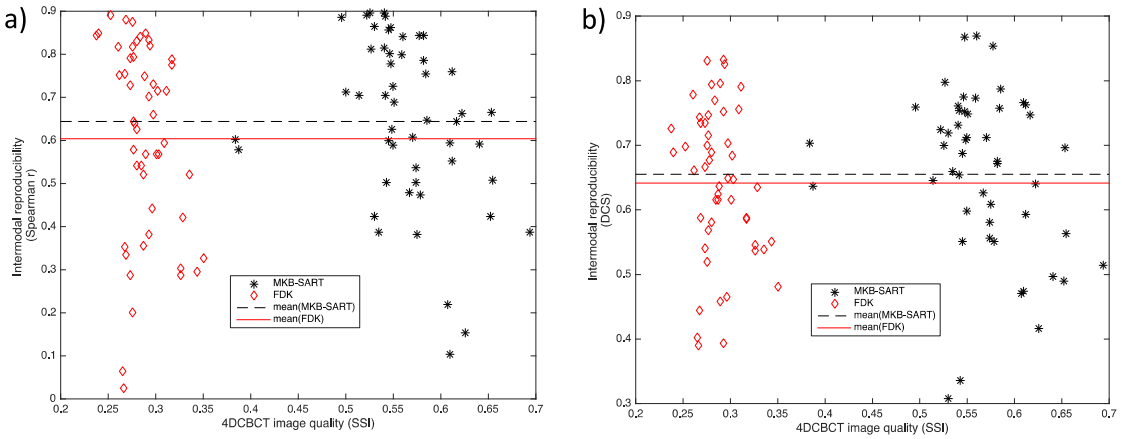


Figure 3 (a) Intermodality reproducibility between $CTVI^{4DCBCT}$ and $CTVI^{4DCT}$ as measured by the Spearman correlation *r*, plotted as a function of the image quality metric (SSI) derived from 4DCBCT. The black dashed and red lines show the mean Spearman *r* the case where 4DCBCT was reconstructed using MKB-SART ($r = 0.64$) and FDK ($r = 0.60$), respectively. (b) Same as for (a) but where the intermodality reproducibility is quantified by the DSC for lung defect regions. The black dashed and red lines show the mean DSC for the case where 4DCBCT was reconstructed using MKB-SART (DSC = 0.66) and FDK (DSC = 0.64) respectively.

Although this work focuses on relative ventilation distributions (rather than absolute ventilation values), we also evaluated the correlations outlined above using the linear Pearson correlation and found similar results.

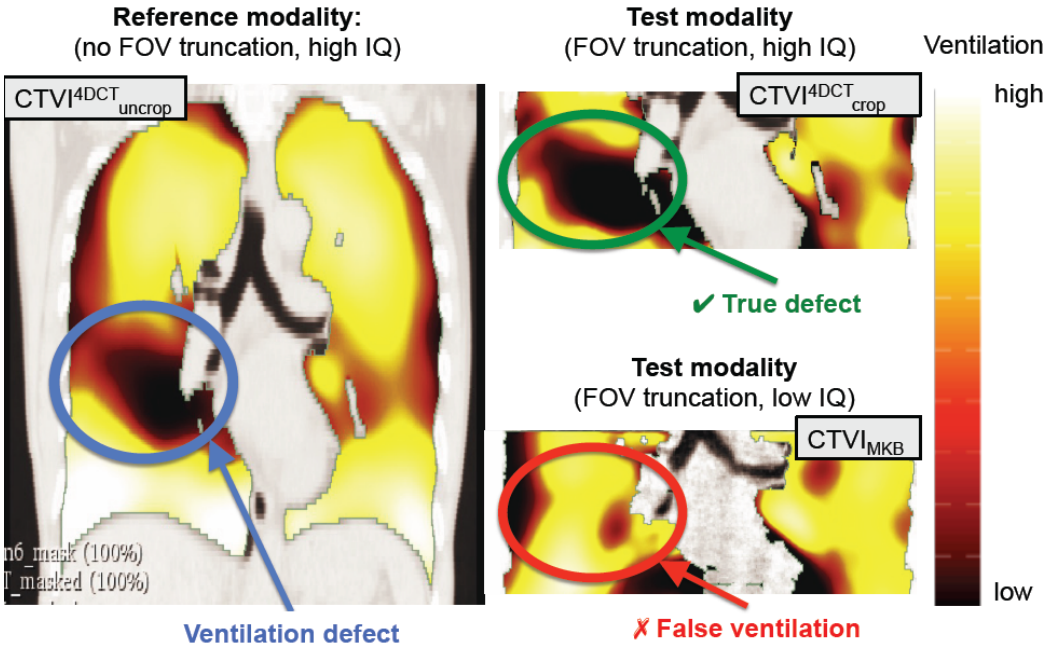


Figure 4: Comparing the impact of FOV and image quality (IQ) on CTVI reproducibility. The ellipses indicate a ventilation defect in the reference CTVI (left panel). The defect is preserved when generating the CTVI from a cropped 4DCT scan (upper right panel) but not in the 4DCBCT scan (lower right) which suffers both FOV truncation and lower image quality. Here IQ = image quality.

3.2 Impact of 4DCBCT FOV on CTVI reproducibility

We found that the limited FOV of the 4DCBCT reconstructions caused truncations of the imaged lung for all patients in this study. Lung volumes as calculated from the time-averaged 4DCT ranged from 1.6 L to 7.2 L, with a mean ± SD of (3.8 ± 1.6) L while lung volumes in the aligned 4DCBCT images ranged from 1.5 L to 5.2 L, with a mean ± SD of (2.8 ± 1.0) L. The 4DCBCT imaged lung volumes were on average 77% of the corresponding 4DCT, ranging from 58% to 96%. We found no correlation between 4DCT imaged lung volume and level of truncation (R = 0.3, p=0.03).

As is evident from Table 1, the FOV truncation in 4DCBCT (as modelled by cropping of the 4DCT) did not significantly affect the Spearman *r*-values or DSC values between 4DCBCT-based CTVIs and 4DCT-based CTVIs on average. In fact, the differences in CTVI correlation due to 4DCT cropping condition had p-values > 0.13 irrespective of the 4DCBCT reconstruction algorithm or regularization λ.

Figure 4 shows an extreme example of lung truncation in 4DCBCT, where 58% of the imaged lung volume is truncated relative to the uncropped 4DCT. This is the most extreme case of truncation in the study, yet the largest difference visually is not between $CTVI_{crop}^{4DCT}$ and $CTVI_{uncrop}^{4DCT}$, but rather between $CTVI_{crop}^{4DCT}$ and $CTVI^{4DCBCT}$. That is, the largest differences in CTVI reproducibility did not appear to arise strictly due to lung truncation, but rather the choice of image modality. We note that since all 4DCBCT were scans acquired as part of treatment setup, differences lesion position and hence patient positioning influence the level of truncation.

The intermodality CTVI reproducibility was seen to vary with other patient- and scan-specific factors, namely the lung volume as determined in the time-averaged 4DCT and the fraction of lung truncated in the time-averaged 4DCBCT. We found that the intermodality CTVI reproducibility showed a negative linear correlation (Pearson correlation R = -0.85, p = 0.02) with increasing lung volume as segmented from the time-averaged 4DCT; see Figure 5(a). As shown in Figure 5(b) we observed a similar, though less significant linear correlation (R = 0.55, p = 0.07) with the level of lung truncation in the 4DCBCT scans, quantified as $(Volume_{CT} - Volume_{CT\ crop})/Volume_{CT}$. There is no clear trend between time-averaged 4D-CT lung volume and amount of imaged lung truncation in 4DCBCT scans (R = 0.3, p=0.08). In Figure 5(c) we also report a positive trend (R = 0.79, p= 0.04) for the mean of the Jacobian map (a proxy for the respiratory effort, see Discussion) as a function of imaged lung volume. Figure 5(d) shows a decrease in intermodal correlation (mean Spearman *r*) with increasing mean Jacobian (R = 0.84, p = 0.01). In other words, increasing levels of lung truncation in 4DCBCT, and increasing levels of respiratory effort in 4DCT were both associated with reduced intermodality CTVI reproducibility; but to complicate matters, lung volume and respiratory effort were also correlated with each other.

3.3 Impact of DIR accuracy on CTVI reproducibility

We assed DIR accuracy using the MSE and TRE methods on a representative set of anatomical images used to generate $CTVI_{MKB}^{4DCBCT}$ (λ = 5) and $CTVI_{uncrop}^{4DCT}$ (λ = 1). As an initial self-consistency check of the DIR motion fields for 4DCT, we analysed the correlation between lung volume changes in the 4DCT exhale/inhale lung masks, and the corresponding sum of Jacobian determinant values in the CTVI. We found a good linear correlation of (Pearson's R = 0.88, p = 0.01) over all 30 4DCT scans. We also investigated the Jacobian determinant distributions derived from 4DCT (4DCBCT) for significant regions of negative

values. For any given DIR result, no more than 0.2% (0.5%) of lung voxels had negative Jacobian values.

In terms of the interphase (exhale-to-inhale) DIR, the MSE-based evaluation of DIR accuracy found $\Delta MSE_{DIR} = 21.1\%$ (78.6%) for 4DCBCT (4DCT). This indicates poorer DIR performance for 4DCBCT compared to 4DCT. However, as can be seen in Figure 5(e), we found no correlation between 4DCBCT ΔMSE_{DIR} and the intermodality CTVI reproducibility as measured by Spearman correlation (R = 0.14, p=0.7). A moderate linear correlation (R = 0.45, p=0.04) was found between 4DCT ΔMSE_{DIR} and Spearman *r*.

For the landmark-based TRE analysis for interphase DIR, the adaptive SIFT algorithm detected a mean ± SD number of landmarks 136 ± 44 for 4DCT inhale-exhale image pairs and 192 ± 76 for 4DCBCT. The mean ± SD TRE between landmarks after DIR was (1.7 ± 0.4) mm for 4DCBCT and (2.3 ± 0.5) mm for 4DCT. As shown in Figure 5(f), we found a moderately negative correlation (albeit not statistically significant) between TRE and accuracy of $CTVI_{MKB}^{4DCBCT}$ (R = -0.50, p = 0.22) but no correlation for the case of $CTVI_{uncrop}^{4DCT}$ (R = 0.27, p = 0.7).

For the landmark-based analysis of the intermodal DIR, we assessed the TRE between SIFT-identified matching features between the time-averaged 4DCBCT and the corresponding time-averaged 4DCT after registration. The mean ± SD number of landmarks was 140 ± 54 and the TRE was (0.9 ± 0.4) mm and showed no correlation with intermodal CTVI reproducibility (R = -0.20, p = 0.6).

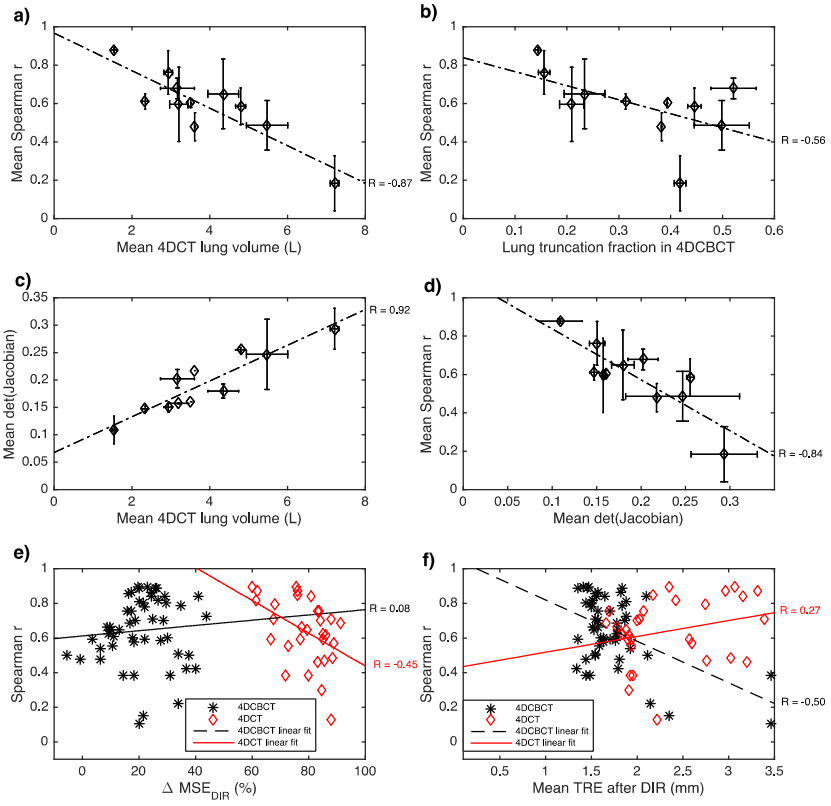


Figure 5 Dependence of the intermodality CTVI reproducibility (quantified by Spearman *r*-value) on: (a) lung volume in the time-average 4DCT, and (b) fraction of lung truncation in

the corresponding time-averaged 4DCBCT. Also plotted are: (c) the dependence of the mean Jacobian determinant on the lung volume in time-averaged 4DCT, and (d) the dependence of Spearman *r*-values on the mean Jacobian determinant values from (c). The Spearman-*r* is also plotted against DIR accuracy metrics including (e) the fraction of intensity MSE differences resolved by DIR and (d) anatomic landmark based TRE following DIR. All data points refer to the comparison between $CTVI_{MKB}^{4DCBCT}$ (with $\lambda = 5$) and $CTVI_{uncrop}^{4DCT}$ (with $\lambda = 1$). Pearson's R-values are reported for all linear fits.

4 DISCUSSION

The adaptation of lung cancer radiation therapy treatments to longitudinal changes in 4DCBCT-based CTVIs is a technically demanding task. In the absence of ventilation changes, it is implicitly required that the DIR should provide comparable ventilation distributions between daily 4DCBCT scans and the treatment planning 4DCT, despite the large image quality differences between those modalities. This is important to demonstrate the potential of 4DCBCT-based CTVI to track longitudinal ventilation changes over a course of radiotherapy. We found that the choice of 4DCBCT image reconstruction and DIR parameters have a significant impact on the strength of the intermodality reproducibility.

Using an in-house iterative 4DCBCT reconstruction method, we were able to demonstrate that improving perceived image quality moderately enhances 4DCBCT-based CTVI accuracy relative to 4DCT ($p=0.03$). Within each 4DCBCT reconstruction method we found no clear correlation of the intermodality reproducibility of 4DCBCT-VI with a perception-based image quality metric (SSI). This is possibly because the differences in image quality as measured by the SSI across a single 4DCBCT reconstruction method are minor compared to the differences between different reconstruction methods; see Figure 3. In addition, MKB-SART reconstruction uses the time average of all FDK respiratory phase bins as a prior, and is likely to deliver a higher SSI when computed using our method as described in Sec. 2.3; there also may be competing effects such as lung volume and truncation that are confounding the impact of image quality on intermodality reproducibility. Understanding these effects is important for clinical implementation, and image processing and reconstruction techniques need to be researched further. Image quality forms a contra-indication for this method when the reconstructed volume is of poor visual quality.

As pointed out in ref.[11] the choice of DIR parameters is important when deriving Jacobian based CTVI, which is explored in this work through variation of the motion field regularization. Higher regularization enforces smoother Jacobian maps, and a balance needs to be struck between intermodality reproducibility and spatial fidelity of regional ventilation information. Optimization of CTVI parameters becomes difficult without access to actual gold standards such as contrast-enhanced (Xe) CT, SPECT V/Q or hyperpolarized (3 He) magnetic resonance imaging (MRI). The intermodality correlation between 4DCBCT-based CTVIs and 4DCT-based CTVIs compares favourably to other cross modality CTVI validation studies. Kipritidis *et al.*[6] reported a mean Spearman value of $r = 0.42$ and a DSC of 0.52 for defect regions when comparing 4DCT-based CTVI with Galligas PET; meanwhile Castillo *et al.*[17] and Yamamoto *et al.*[7] reported mean DSC values in the range 0.3-0.5 when comparing defect regions between 4DCT-based CTVI and ^{99m}Tc- DTPA SPECT. Kida *et al.*[18] found that Spearman correlations of around $r = 0.4$ are consistent with equivalent functional dosimetry between CTVI- and SPECT-guided functional avoidance treatment plans.

It is important to note that the reproducibility between 4DCBCT- and 4DCT-based ventilation in our study appears less strong than the reproducibility of ventilation images derived purely from repeat 4DCT. For example, Du et al. [31] found Spearman r ~ 0.8 for Jacobian-based ventilation image pairs derived from short-interval 4DCT scans of lung cancer patients. Moreover, both Du et al.[31] and Latifi et al.[32] found that introducing a local scaling factor, for example based on the tidal volume in each repeat scan, can improve reproducibility of regional ventilation in absolute terms. We can anticipate that for 4DCBCT, the main difficulty with a tidal-volume normalization approach is that poor image quality may limit the accuracy of the 4D lung segmentation; but in any case the normalization of ventilation images is an important factor to be considered in further optimising the reproducibility between 4DCBCT- and 4DCT-based ventilation.

One of the major differences between the imaging modalities in this dataset was the much smaller FOV of the 4DCBCTs, as exemplified in Figure 4. By cropping the 4DCTs to the same FOV and comparing to the uncropped CTVIs, we were able to establish that the effect on the generated VIs was noticeable but not severe, and not significant to intramodal reproducibility. Furthermore, our results suggest that increasing the motion field regularization parameter may mitigate the effects of imaged lung truncation, but at the cost of loss of regional information.

In clinical practice, a limited FOV can be problematic when treating patients with large lungs. Even if the truncation had no impact on CTVI generation, the lack of regional information for the entire lung adds complexity to the task of accumulating functional dose for the organ as a whole. We observed a negative relationship between imaged CT lung volume and correlation between 4DCBCT- and 4DCT-based CTVIs. Increasing lung volume was also correlated with increases in the mean DIR motion field Jacobian determinant, which reflects the ratio of volumes before and after deformation in a specified region. Since the mean Jacobian is a measure of global volume change in the lung, it can be strongly correlated with respiratory effort and hence CTVI reproducibility[33]. It is not clear whether it is lung volume, or differences in respiratory effort that are the major factor in determining the intermodality reproducibility; a correction method based on respiratory effort may help to mitigate this effect. This is further implied by the clear positive linear correlation between the mean Jacobian and lung volume, and the decrease of correlation between CTVIs with increasing mean Jacobian.

For the purpose of this study we assumed that 4DCT, after a visual pre-selection process to exclude severe irregular breathing motion artefacts, always provides a “correct” reconstruction of the imaged patient, yet not all anatomic truncation or duplication, nor all respiratory phase binning errors are immediately detectable. Human studies of 4DCT-based CTVI generally report lower intermodal correlation than studies performed on ventilated animals (e.g. ref.[4]), an observation that could be attributed to changes in breathing pattern during or between imaging sessions. As adaptive ventilation guidance may be indicated for those patients experiencing ventilation changes during treatment, future studies should address this by, e.g. applying recorded breathing traces on real or virtual thoracic phantoms for the purpose of optimizing the extraction of CTVIs from 4DCT and 4DCBCT using a true gold standard.

This study relied on a single DIR implementation for both interphase DIR (for generating CTVIs), and intermodal DIR (for aligning intermodal CTVIs); in a sense this is a study limitation since DIR accuracy can vary from scan to scan and with different DIR algorithm

parameters. To understand the impact of DIR accuracy on CTVI reproducibility, we employed an analysis of intensity-based MSE and anatomic landmark-based TRE between different pairs of deformably registered images. For the interphase DIR in 4DCT, a moderate relationship was found between the CTVI reproducibility and the ΔMSE_{DIR}, but no significant relationship was found in terms of the landmark-based TRE. For interphase DIR with 4DCBCT, we observed the opposite (i.e. there was a significant link between CTVI reproducibility and TRE but not ΔMSE_{DIR}). For the intermodal DIR, CTVI reproducibility showed no significant link with the TRE. Based on the above we could not identify DIR accuracy as being the major factor affecting CTVI reproducibility in this study. We point out that there are two limitations of our DIR evaluation: (i) a lack of HU equivalence for 4DCBCT may limit a direct comparison between ΔMSE_{DIR} values between 4DCBCT and 4DCT, and (ii) the adaptive SIFT algorithm has not been validated for 4DCBCT.

We note that, while promising, these results are strictly only applicable to one particular DIR algorithm and one particular (non-clinical) 4DCBCT scan method. Other DIR methods (e.g. biomechanical DIR or diffeomorphisms) may produce different variable levels of intermodality CTVI reproducibility given the same input 4DCBCT and 4DCT images. Clinical 4DCBCT reconstructions may feature varying image quality depending on the acquisition protocol, number of projections or use of scatter correction. Since the 4DCBCT and 4DCT reconstructions are available in a publically available dataset[1], there is potential to further improve on these results using more robust DIR algorithms.

5 CONCLUSION

This is to our knowledge the first direct comparison of CT ventilation images for lung cancer radiation therapy patients using 4DCBCT and 4DCT imaged at similar time points. We found good intermodal correlation between the CTVIs on average, however the use of 4DCBCT scans with iterative reconstruction is recommended to achieve similar results to treatment planning 4DCT. Additional optimization of 4DCBCT image reconstruction and DIR algorithms may help to improve the accuracy of 4DCBCT-based CTVI. Breathing effort correction and patient pre-selection by means of lung volume might be necessary in clinical applications.

6 ACKNOWLEDGEMENTS

We thank Dr. Geoffrey D. Hugo (Virginia Commonwealth University, Richmond VA) who was instrumental in providing the lung imaging data for this data. We also thank Dr Paganelli for supplying the adaptive SIFT software. Dr Kipritidis and Dr Woodruff were supported by a Cancer Institute NSW Early Career Fellowship. This work was also supported by an NHMRC Australia Fellowship and US NCI P01CA116602. The authors have no relevant conflicts of interest to disclose.

7 REFERENCES

[1] X. Meng, K. Frey, M. Matuszak, S. Paul, R. Ten Haken, J. Yu, F.-M. Kong, "Changes in Functional Lung Regions During the Course of Radiation Therapy and Their Potential Impact on Lung Dosimetry for Non-Small Cell Lung Cancer," International Journal of Radiation Oncology*Biology*Physics **89**, 145-151 (2014).

[1] http://doi.org/10.7937/K9/TCIA.2016.ELN8YGLE

[2] S. Siva, R. Thomas, J. Callahan, N. Hardcastle, D. Pham, T. Kron, R.J. Hicks, M.P. MacManus, D.L. Ball, M.S. Hofman, "High-resolution pulmonary ventilation and perfusion PET/CT allows for functionally adapted intensity modulated radiotherapy in lung cancer," Radiotherapy and Oncology **115**, 157-162 (2015).

[3] T. Guerrero, K. Sanders, J. Noyola-Martinez, E. Castillo, Y. Zhang, R. Tapia, R. Guerra, Y. Borghero, R. Komaki, "Quantification of regional ventilation from treatment planning CT," Int J Radiat Oncol Biol Phys **62**, 630-634 (2005).

[4] J.M. Reinhardt, K. Ding, K. Cao, G.E. Christensen, E.A. Hoffman, S.V. Bodas, "Registration-based estimates of local lung tissue expansion compared to xenon CT measures of specific ventilation," Medical image analysis **12**, 752-763 (2008).

[5] M.S. Hofman, J.-M. Beauregard, T.W. Barber, O.C. Neels, P. Eu, R.J. Hicks, "68Ga PET/CT Ventilation–Perfusion Imaging for Pulmonary Embolism: A Pilot Study with Comparison to Conventional Scintigraphy," Journal of Nuclear Medicine **52**, 1513-1519 (2011).

[6] J. Kipritidis, S. Siva, M.S. Hofman, J. Callahan, R.J. Hicks, P.J. Keall, "Validating and improving CT ventilation imaging by correlating with ventilation 4D-PET/CT using 68Ga-labeled nanoparticles," Med Phys **41**, 011910 (2014).

[7] T. Yamamoto, S. Kabus, C. Lorenz, E. Mittra, J.C. Hong, M. Chung, N. Eclov, J. To, M. Diehn, B.W. Loo Jr, P.J. Keall, "Pulmonary Ventilation Imaging Based on 4-Dimensional Computed Tomography: Comparison With Pulmonary Function Tests and SPECT Ventilation Images," International Journal of Radiation Oncology*Biology*Physics **90**, 414-422 (2014).

[8] L. Mathew, A. Wheatley, R. Castillo, E. Castillo, G. Rodrigues, T. Guerrero, G. Parraga, "Hyperpolarized 3He Magnetic Resonance Imaging: Comparison with Four-dimensional X-ray Computed Tomography Imaging in Lung Cancer," Academic Radiology **19**, 1546-1553 (2012).

[9] Y. Vinogradskiy, R. Castillo, E. Castillo, S.L. Tucker, Z. Liao, T. Guerrero, M.K. Martel, "Use of 4-Dimensional Computed Tomography-Based Ventilation Imaging to Correlate Lung Dose and Function With Clinical Outcomes," International Journal of Radiation Oncology*Biology*Physics **86**, 366-371 (2013).

[10] T. Yamamoto, S. Kabus, M. Bal, P. Keall, S. Benedict, M. Daly, "The first patient treatment of computed tomography ventilation functional image-guided radiotherapy for lung cancer," Radiotherapy and Oncology **118**, 227-231 (2016).

[11] J. Kipritidis, G. Hugo, E. Weiss, J. Williamson, P.J. Keall, "Measuring interfraction and intrafraction lung function changes during radiation therapy using four-dimensional cone beam CT ventilation imaging," Medical Physics **42**, 1255-1267 (2015).

[12] Y.Y. Vinogradskiy, R. Castillo, E. Castillo, A. Chandler, M.K. Martel, T. Guerrero, "Use of weekly 4DCT-based ventilation maps to quantify changes in lung function for patients undergoing radiation therapy," Medical Physics **39**, 289-298 (2012).

[13] S. Leng, J. Zambelli, R. Tolakanahalli, B. Nett, P. Munro, J. Star-Lack, B. Paliwal, G.-H. Chen, "Streaking artifacts reduction in four-dimensional cone-beam computed tomography," Medical physics **35**, 4649-4659 (2008).

[14] C.C. Shieh, J. Kipritidis, R.T. O'Brien, Z. Kuncic, P.J. Keall, "Image quality in thoracic 4D cone-beam CT: a sensitivity analysis of respiratory signal, binning method, reconstruction algorithm, and projection angular spacing," Med Phys **41**, 041912 (2014).

[15] T. Yamamoto, U. Langner, B.W. Loo, Jr., J. Shen, P.J. Keall, "Retrospective Analysis of Artifacts in Four-Dimensional CT Images of 50 Abdominal and Thoracic

Radiotherapy Patients," International Journal of Radiation Oncology • Biology • Physics **72**, 1250-1258 (2008).

16 T. Yamamoto, S. Kabus, C. Lorenz, E. Johnston, P.G. Maxim, M. Diehn, N. Eclov, C. Barquero, B.W. Loo, P.J. Keall, "4D CT lung ventilation images are affected by the 4D CT sorting method," Medical Physics **40**, 101907-n/a (2013).

17 R. Castillo, E. Castillo, J. Martinez, T. Guerrero, "Ventilation from four-dimensional computed tomography: density versus Jacobian methods," Physics in Medicine and Biology **55**, 4661 (2010).

18 S. Kida, M. Bal, S. Kabus, M. Negahdar, X. Shan, B.W. Loo Jr, P.J. Keall, T. Yamamoto, "CT ventilation functional image-based IMRT treatment plans are comparable to SPECT ventilation functional image-based plans," Radiotherapy and Oncology **118**, 521-527 (2016).

19 W. Zhou, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing **13**, 600-612 (2004).

20 N.O. Roman, W. Shepherd, N. Mukhopadhyay, G.D. Hugo, E. Weiss, "Interfractional Positional Variability of Fiducial Markers and Primary Tumors in Locally Advanced Non-Small-Cell Lung Cancer During Audiovisual Biofeedback Radiotherapy," International Journal of Radiation Oncology*Biology*Physics **83**, 1566-1572 (2012).

21 S. Balik, E. Weiss, N. Jan, N. Roman, W.C. Sleeman, M. Fatyga, G.E. Christensen, C. Zhang, M.J. Murphy, J. Lu, P. Keall, J.F. Williamson, G.D. Hugo, "Evaluation of 4-dimensional Computed Tomography to 4-dimensional Cone-Beam Computed Tomography Deformable Image Registration for Lung Cancer Adaptive Radiation Therapy," International Journal of Radiation Oncology*Biology*Physics **86**, 372-379 (2013).

22 G.C. Mc Kinnon, R. Bates, "Towards imaging the beating heart usefully with a conventional CT scanner," Biomedical Engineering, IEEE Transactions on **28**, 123-127 (1981).

23 A.H. Andersen, A.C. Kak, "Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm," Ultrasonic Imaging **6**, 81-94 (1984).

24 J. Kipritidis, H.C. Woodruff, E.M. Eslick, F. Hegi-Johnson, P.J. Keall, "New pathways for end-to-end validation of CT ventilation imaging (CTVI) using deformable image registration," 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), 939-942 (2016).

25 P.A. Yushkevich, J. Piven, H.C. Hazlett, R.G. Smith, S. Ho, J.C. Gee, G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," NeuroImage **31**, 1116-1128 (2006).

26 H.B. Mann, D.R. Whitney, "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," The Annals of Mathematical Statistics **18**, 50-60 (1947).

27 L.R. Dice, "Measures of the amount of ecologic association between species," Ecology **26**, 297-302 (1945).

28 A.C. Riegel, J.Y. Chang, S.S. Vedam, V. Johnson, P.-C.M. Chi, T. Pan, "Cine Computed Tomography Without Respiratory Surrogate in Planning Stereotactic Radiotherapy for Non–Small-Cell Lung Cancer," International Journal of Radiation Oncology*Biology*Physics **73**, 433-441 (2009).

29 P. Chiara, P. Marta, R. Marco, C.S. Gregory, C. Delia, A. Daniela, O. Roberto, B. Guido, "Scale invariant feature transform in adaptive radiation therapy: a tool for

deformable image registration assessment and re-planning indication," Physics in Medicine and Biology **58**, 287 (2013).

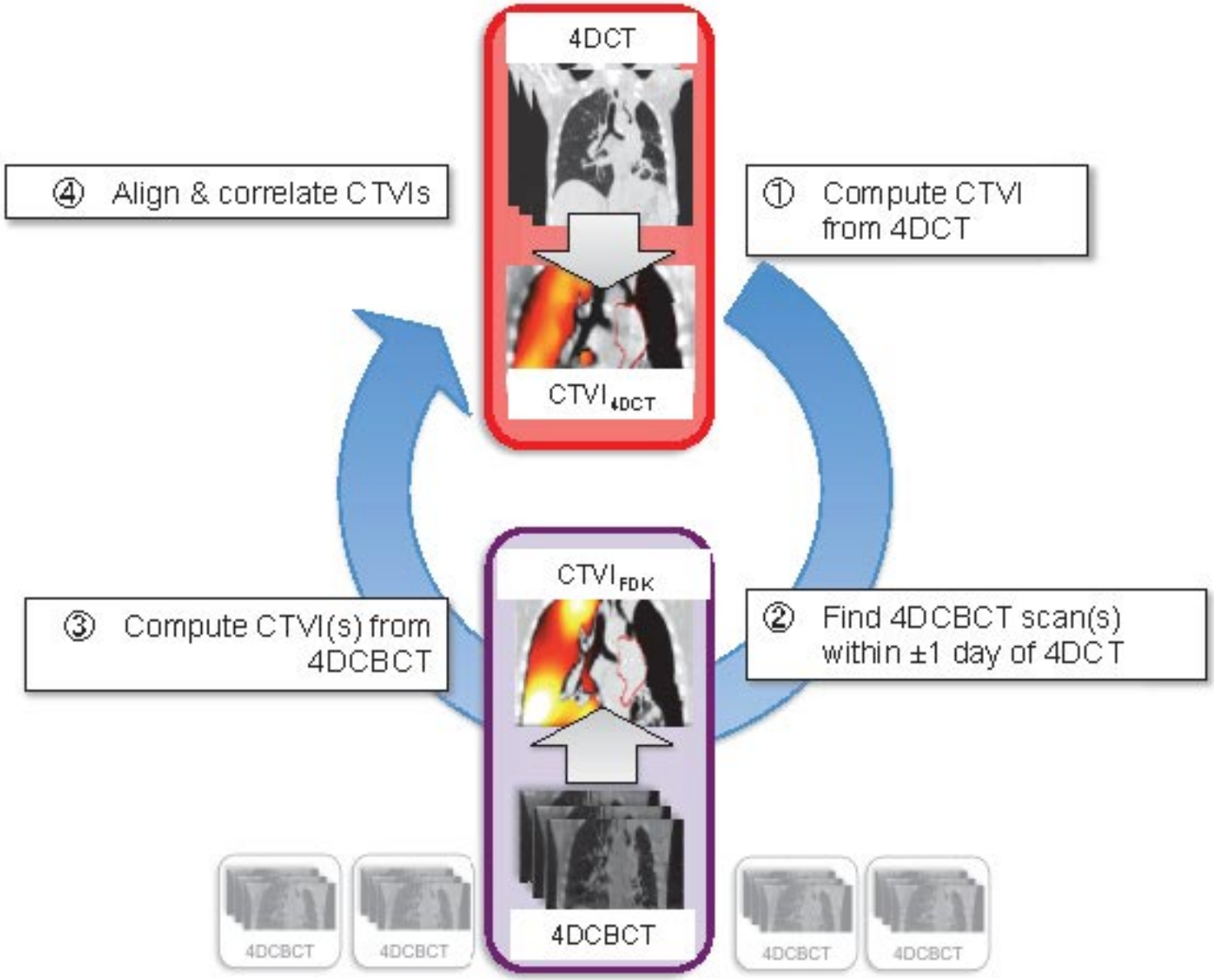
30 G. Landry, R. Nijhuis, G. Dedes, J. Handrack, C. Thieke, G. Janssens, J. Orban de Xivry, M. Reiner, F. Kamp, J.J. Wilkens, C. Paganelli, M. Riboldi, G. Baroni, U. Ganswindt, C. Belka, K. Parodi, "Investigating CT to CBCT image registration for head and neck proton therapy as a tool for daily dose recalculation," Medical Physics **42**, 1354-1366 (2015).

31 K. Du, J.E. Bayouth, K. Ding, G.E. Christensen, K. Cao, J.M. Reinhardt, "Reproducibility of intensity-based estimates of lung ventilation," Medical Physics **40**, 063504 (2013).

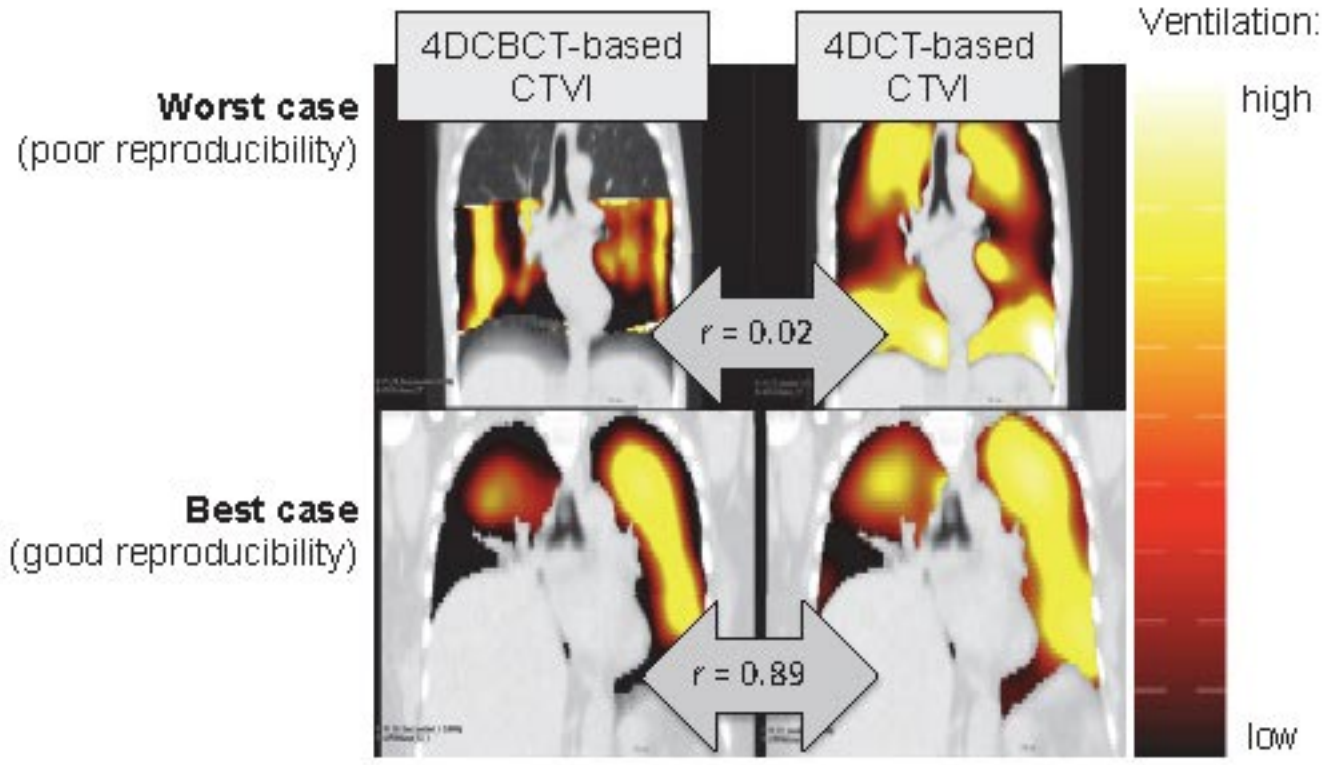
32 K. Latifi, V. Feygelman, E.G. Moros, T.J. Dilling, C.W. Stevens, G.G. Zhang, "Normalization of Ventilation Data from 4D-CT to Facilitate Comparison between Datasets Acquired at Different Times," PloS one **8**, e84083 (2013).

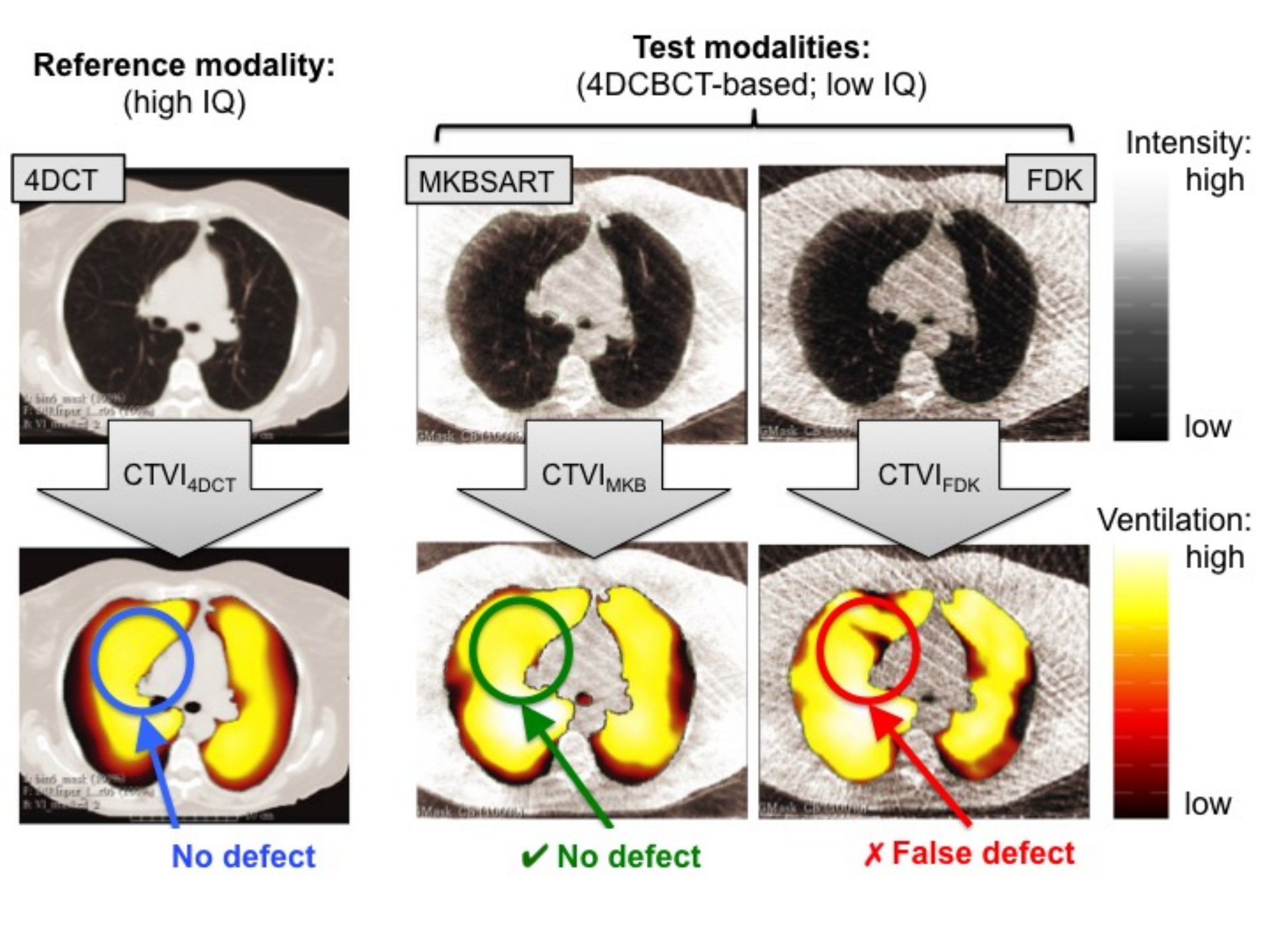
33 K. Du, J.M. Reinhardt, G.E. Christensen, K. Ding, J.E. Bayouth, "Respiratory effort correction strategies to improve the reproducibility of lung expansion measurements," Med Phys **40**, 123504 (2013).

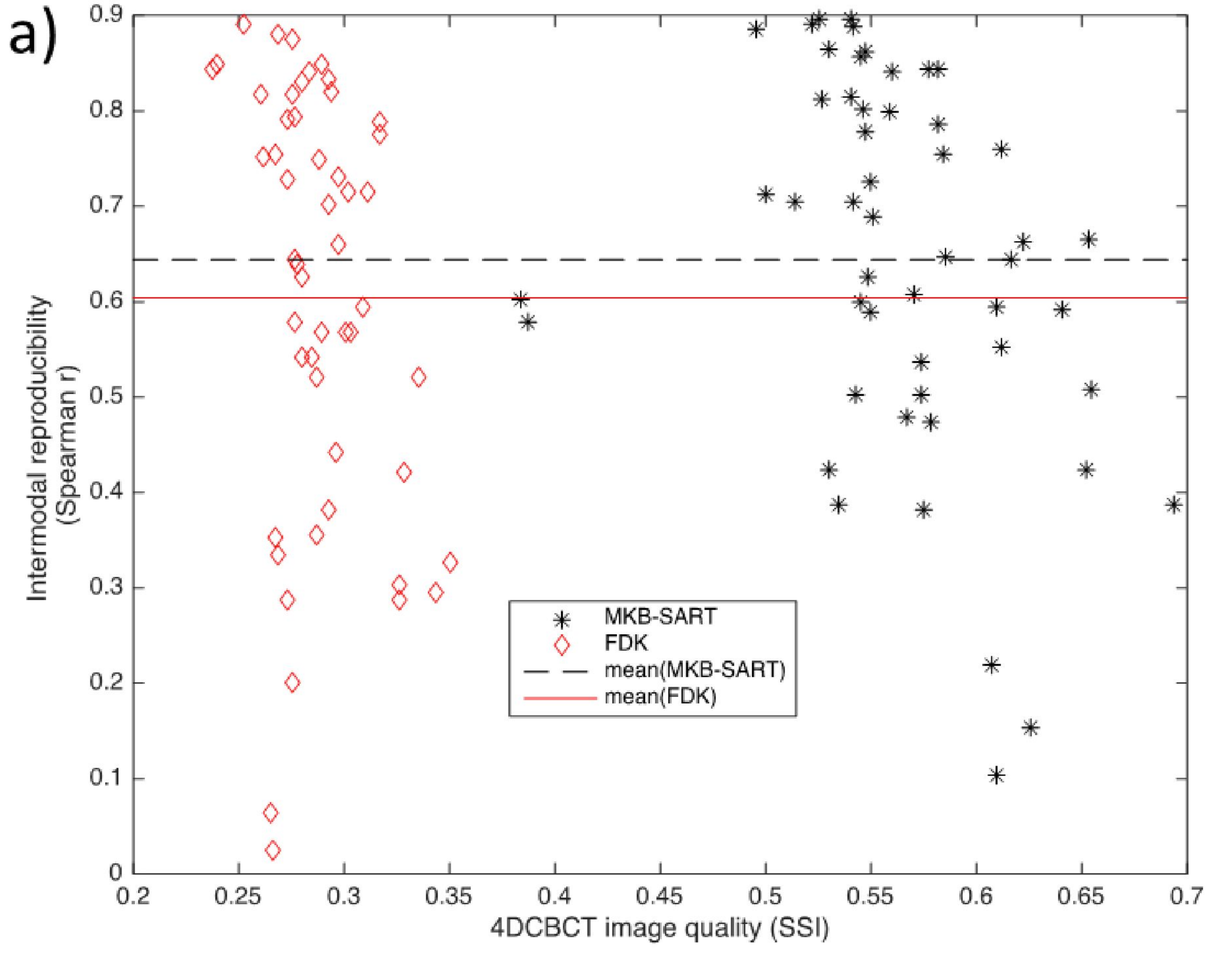
a) Study schematic (for each 4DCT scan):

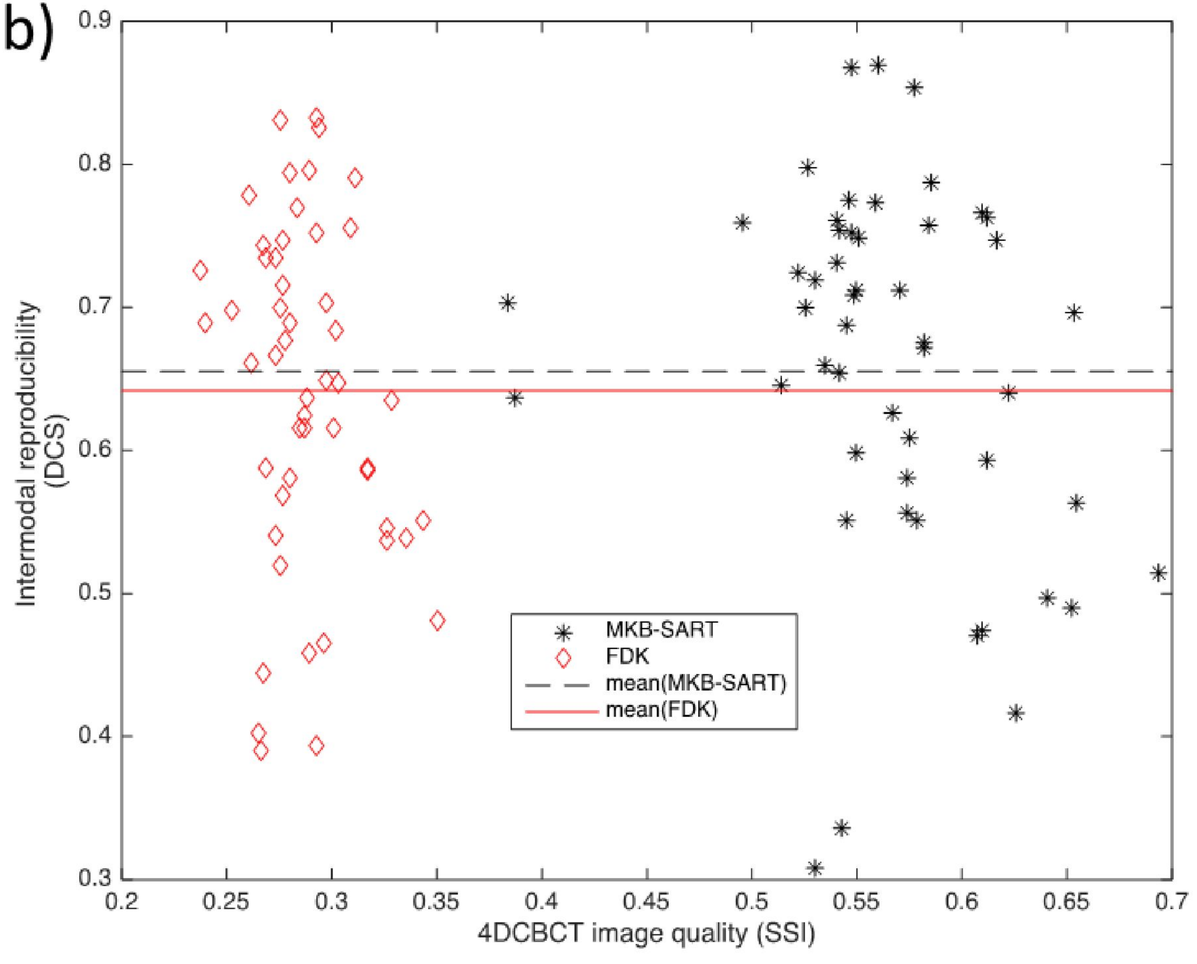


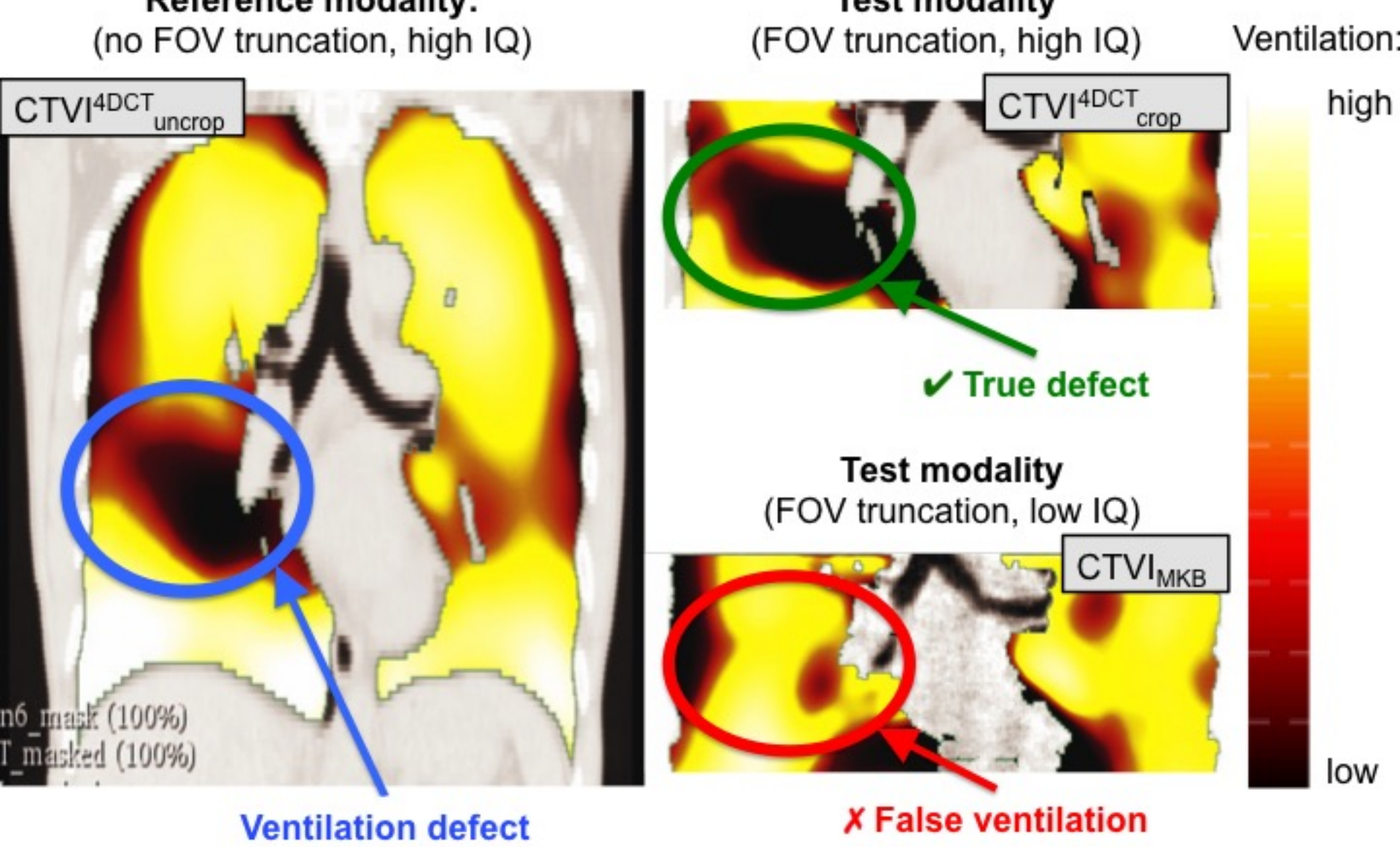
b) Best and worst cases in terms of Spearman correlation (*r*):

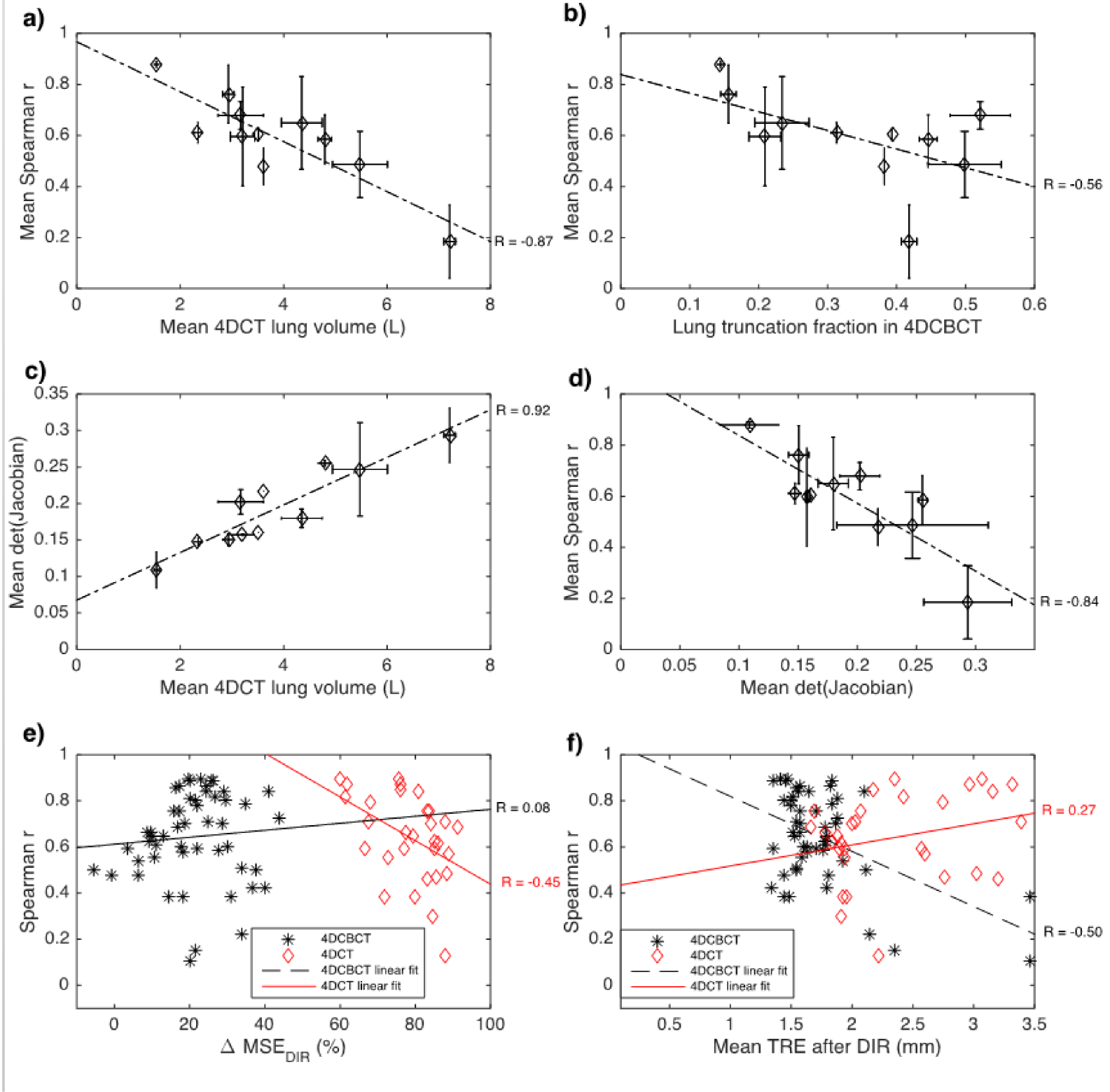



Reference modality:
(high IQ)
Test modalities:
(4DCBCT-based; low IQ)
4DCT
MKBSART
FDK
Intensity:
high
low
CTVI4DCT
CTVIMKB
CTVIFDK
Ventilation:
high
low
No defect
✔ No defect
✗ False defect


a)
Intermodal reproducibility (Spearman r)
4DCBCT image quality (SSI)
MKB-SART
FDK
mean(MKB-SART)
mean(FDK)


b)
Intermodal reproducibility (DCS)
4DCBCT image quality (SSI)
MKB-SART
FDK
mean(MKB-SART)
mean(FDK)


Reference modality:
(no FOV truncation, high IQ)
CTVI4DCT uncrop
n6_mask (100%)
T_masked (100%)
Ventilation defect
Test modality
(FOV truncation, high IQ)
CTVI4DCT crop
✔ True defect
Test modality
(FOV truncation, low IQ)
CTVIMKB
✗ False ventilation
Ventilation:
high
low


a)
Mean Spearman r
Mean 4DCT lung volume (L)
R = -0.87
b)
Mean Spearman r
Lung truncation fraction in 4DCBCT
R = -0.56
c)
Mean det(Jacobian)
Mean 4DCT lung volume (L)
R = 0.92
d)
Mean Spearman r
Mean det(Jacobian)
R = -0.84
e)
Spearman r
Δ MSE_DIR (%)
R = 0.08
R = -0.45
4DCBCT
4DCT
4DCBCT linear fit
4DCT linear fit
f)
Spearman r
Mean TRE after DIR (mm)
R = 0.27
R = -0.50
4DCBCT
4DCT
4DCBCT linear fit
4DCT linear fit