Elsevier Editorial System(tm) for
Radiotherapy and Oncology
Manuscript Draft

Manuscript Number: RO-D-18-01308R1

Title: Imaging of regional ventilation: Is CT Ventilation Imaging the answer? A Systematic Review of the Validation Data

Article Type: Review Article (by invitation)

Keywords: lung cancer; radiotherapy; pneumonitis; computed tomography; ventilation

Corresponding Author: Dr. Fiona Hegi-Johnson, FRANZCR, MBBS

Corresponding Author's Institution: Peter MacCallum

First Author: Fiona Hegi-Johnson, FRANZCR, MBBS

Order of Authors: Fiona Hegi-Johnson, FRANZCR, MBBS; Dirk De Ruysscher; Paul Keall; Lizza Hendriks; Yevgeniy Vinogradskiy; Tokihiro Yamamoto; Bilal Tahir; John Kipritidis, Northern Sydney Cancer Centre, St Leonards

Abstract: Computed Tomography Ventilation Imaging (CTVI) is an experimental imaging modality that derives regional lung function information from non-contrast respiratory-correlated CT datasets. Despite CTVI being extensively studied in cross-modality imaging comparisons, there is a lack of consensus on the state of its clinical validation in humans. This systematic review evaluates the CTVI clinical validation studies to date, highlights their common strengths and weaknesses and makes recommendations. We performed a PUBMED and EMBASE search of all English language papers on CTVI between 2000 and 2018. The results of these searches were filtered in accordance to a set of eligibility criteria and analysed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Guidelines. One hundred and forty-four records were identified, and 66 full text records were reviewed. After detailed assessment, twenty-three full text papers met the selection criteria and were included in the final review. This included thirteen prospective studies, with 579 human subjects. Studies used diverse methodologies, with a large amount of heterogeneity between different studies in terms of the reference ventilation imaging modality (e.g. nuclear medicine, hyperpolarised gas MRI), imaging parameters, DIR algorithm(s) used, and ventilation metric(s) applied. The most common ventilation metrics used deformable image registration to evaluate the exhale-to-inhale motion field Jacobian determinant (DIR-Jac) or changes in air volume content based on Hounsfield Units (DIR-HU). The strength of correlation between CTVI and the reference ventilation imaging modalities was moderate to strong when evaluated at the lobar or global level, with the average ± S.D.(number of studies) linear regression correlation coefficients were 0.73±0.25 (n= 6)and 0.86± 0.11(n=12) for DIR-Jac and DIR-HU respectively, and the SPC were 0.45±0.31(n=6) and 0.41±0.11(n=5) for DIR-Jac and DIR-HU respectively. We concluded that it is difficult to make a broad statement about the validity of CTVI due to the diverse methods used in the validation literature. Typically, CTVI appears to show reasonable cross-modality correlations at the lobar/whole lung level but poor correlations at the voxel level. Since CTVI is seeing new implementations in prospective trials, it is clear that refinement and standardization of the clinical validation methodologies are required. CTVI appears to be of relevance in radiotherapy planning, particularly in patients whose main pulmonary impairment is not a gas exchange problem but alternative imaging approaches may need to be considered in patients with other pulmonary diseases (i.e. restrictive or gas exchange problems).

Dr Fiona Hegi-Johnson
Peter MacCallum Cancer Centre
305 Grattan St
Melbourne 3000
Fiona.hegi-johnson@petermac.org

To the Editor-In-Chief
Professor Michael Baumann
Radiotherapy and Oncology

6th of November 2018

Dear Professor Baumann,

Re: Submission of new manuscript "Imaging of regional ventilation: Is CT Ventilation Imaging the answer? A Systematic Review of the Validation Data

Thank you for considering our paper for publication as a critical review. As we mentioned when proposing this work to you, this review reflects the cumulative expertise of several acknowledged experts in the field, such as authors Keall, Yamamoto and Vinogradskiy who have been instrumental in the development of this technology. This is coupled with the viewpoint of author De Ruysscher, who has extensive experience in thoracic oncology, to give a uniquely clinical assessment of this new imaging technology.

This work has not been previously published, and all authors have reviewed and approved the final manuscript and were involved in (1) the conception and design of the study, data acquisition and analysis and interpretation of data and (2) drafting and critically revising the article for intellectual content.

We look forward to the outcome of the review.

With kind regards,

Dr Fiona Hegi-Johnson and Dr John Kipritidis

***Point by Point detailed Response to Reviewer comments**

Dr Fiona Hegi-Johnson
The Sir Peter MacCallum Department of Oncology
University of Melbourne
305 Grattan St Melbourne

26th February 2019

Dear Professor van der Heide,

Thank you very much for your helpful feedback concerning our paper. We attach our amended version and our response to the reviewers' comments below.

We have increased the clarity of our discussion, in particular by streamlining the text to make it more focused, and improving the clarity of the figures and captions.

To address specific comments by the reviewers:

Reviewer #1: Well researched paper. However, it is covering a lot of different aspects and therefore the key points are difficult to abstract from the paper. Consider tightening the discussion to make it more concise. This will improve the understanding of the information you are presenting.

We have removed some of the detail from the discussion, see for example our discussion of pre and post processing of images which has been shortened considerably.
Other Aspects
Figure legend Figure 1 very long.
This has been shortened as suggested
Figure 2 b - dots etc. are not on the lines, legend figure 2a not homogenous in style
This has been amended
Check table numbers (Assessing Quality of the literature", should be 3 not 4
This has been addressed
Table 3: quite confusing with the head and the "actual table". Also, how do you define strength of recommendation?
Thank you for this comment. We have altered the formatting to try and improve clarity here. We acknowledge within the discussion the methodological issues in terms of assessing the quality of this literature. From the point of view of strength of evidence, there is no high level evidence to support the technical implementation of CTVI. Hence, most recommendations must be considered to be based on weak to moderate evidence.

Reviewer #2:

1.There is no statistical method in this review. Is it possible to apply some statistical method, such as meta-analysis to analyze these paper and data. We think it is needed.
We have discussed this criticism, and acknowledge that one limitation of this paper is that it is largely a descriptive analysis of the literature. However, we do not feel that a formal meta-regression would be possible due to the great heterogeneity across the studies, which are nearly all single arm and very small. To address this issue, and to clarify the presentation of results we have presented a Forest plot of the Voxel-based Spearman rank correlations, including mean Jacobian and HU metrics with standard errors. This will hopefully enable readers to get a feel for the small size of the studies, and the significant variability in results.
2.Some format should be noticed, such as Both "3He[39],[40]", "datasets [1],[31]", the comma was in

wrong position. These typographic issues have been addressed.
3.Table 2 was in a little mess, so some information could not be understand better. This table has been simplified significantly to make it easier to follow.
4.*Region of interest (ROI) was defined in the part of results and was not be redefined in the part of conclusion.* This has been addressed.

Please let us know if you require further information.

With kind regards,

Dr Fiona Hegi-Johnson

Imaging of regional ventilation: Is CT Ventilation Imaging the answer?

A Systematic Review of the Validation Data

Fiona Hegi-Johnson[1,2,3,4]; Dirk de Ruysscher[5,6]: Paul Keall[3]; Lizza Hendriks[7]; Yevgeniy Vinogradskiy[8]; Tokihiro Yamamoto[9]; Bilal Tahir[10.11]; John Kipritidis[3,12]

1. Sir Peter MacCallum Department of Oncology, University of Melbourne
2. Department of Radiation Oncology, Peter MacCallum Cancer Centre, 305 Grattan St, Melbourne, Australia 3000
3. ACRF Image X Institute, University of Sydney, Camperdown, NSW, Australia
4. School of Mathematical and Physical Sciences, The University of Newcastle (UON), University Drive, Callaghan NSW 2308
5. Maastricht University Medical Center, Dept. Radiation Oncology (Maastro Clinic), GROW, The Netherlands, Dr. Tanslaan 12, 6229 Maastricht, The Netherlands
6. KU Leuven, Radiation Oncology, Herestraat 49, 3000 Leuven, Belgium
7. Department of Pulmonary Diseases GROW - School for oncology and developmental biology, Maastricht University Medical Center+, Maastricht, the Netherlands
8. Department of Radiation Oncology, University of Colorado, Denver, Colorado USA
9. Radiation Oncology, UC Davis School Medicine, Sacramento, California, USA
10. Academic Unit of Clinical Oncology, University of Sheffield, United Kingdom
11. POLARIS, Academic Radiology, University of Sheffield, United Kingdom
12. Northern Sydney Cancer Centre, St Leonards, NSW Australia

Corresponding author:

Dr Fiona Hegi-Johnson

Peter MacCallum Cancer Centre

305 Grattan St Melbourne

Vic 3000 AUSTRALIA

Fiona.Hegi-Johnson@petermac.org

Abstract

Computed Tomography Ventilation Imaging (CTVI) is an experimental imaging modality that derives regional lung function information from non-contrast respiratory-correlated CT datasets. Despite CTVI being extensively studied in cross-modality imaging comparisons, there is a lack of consensus on the state of its clinical validation in humans. This systematic review evaluates the CTVI clinical validation studies to date, highlights their common strengths and weaknesses and makes recommendations. We performed a PUBMED and EMBASE search of all English language papers on CTVI between 2000 and 2018. The results of these searches were filtered in accordance to a set of eligibility criteria and analysed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Guidelines. One hundred and forty-four records were identified, and 66 full text records were reviewed. After detailed assessment, twenty-three full text papers met the selection criteria and were included in the final review. This included thirteen prospective studies, with 579 human subjects. Studies used diverse methodologies, with a large amount of heterogeneity between different studies in terms of the reference ventilation imaging modality (e.g. nuclear medicine, hyperpolarised gas MRI), imaging parameters, DIR algorithm(s) used, and ventilation metric(s) applied. The most common ventilation metrics used deformable image registration to evaluate the exhale-to-inhale motion field Jacobian determinant (DIR-Jac) or changes in air volume content based on Hounsfield Units (DIR-HU). The strength of correlation between CTVI and the reference ventilation imaging modalities was moderate to strong when evaluated at the lobar or global level, with the average ± S.D.(*number of studies*) linear regression correlation coefficients were 0.73±0.25 (*n= 6*)and 0.86± 0.11(*n=12*) for DIR-Jac and DIR-HU respectively, and the SPC were 0.45±0.31(*n=6*) and 0.41±0.11(*n=5*) for DIR-Jac and DIR-HU respectively. We concluded that it is difficult to make a broad statement about the validity of CTVI due to the diverse methods used in the validation literature. Typically, CTVI appears to show reasonable cross-modality correlations at the lobar/whole lung level but poor correlations at the voxel level. Since CTVI is seeing new implementations in prospective trials, it is clear that refinement and standardization of the clinical validation methodologies are required. CTVI appears to be of relevance in radiotherapy planning, particularly in patients whose main pulmonary impairment is not a gas exchange problem but alternative imaging approaches may need to be considered in patients with other pulmonary diseases (i.e. restrictive or gas exchange problems).

Introduction:

The idea of deriving information about regional pulmonary function from respiratory-correlated computed tomography (CT), especially 4-Dimensional Computed Tomography (4DCT) and inhale/exhale breath-hold CT (BHCT), without exogenous contrast is highly attractive. In the context of radiotherapy treatment planning, respiratory-correlated thoracic CT scans are acquired routinely for lung cancer patients, a population with significant impairment of respiratory function, and breast cancer patients, where radiation-induced lung toxicity remains a major dose-limiting factor. CT Ventilation Imaging (CTVI) is a method for visualizing regional air volume changes in the lung[1],[2] combining 4DCT or BHCT scans with deformable image registration (DIR) to visualise the breathing-induced change in air volume, or "ventilation," an important component of blood-gas exchange. CTVI is currently the subject of a number of clinical trials, which are integrating CTVI data into radiotherapy planning[3],[4] with the goal of minimising irradiation of functional lung and potentially minimising pulmonary toxicity. Attempts have been made to validate CTVI against a wide range of clinical and experimental ventilation imaging modalities including ^{99m}Tc-labeled diethylenetriamine pentacetate (DTPA) V-SPECT[4],[5], ^{68}Ga (Galligas) PET[6], ^{3}He MRI[7], ^{129}Xe MRI[8], 81m-Kr[9] and Technegas V-SPECT[10].

Almost all CTVI methods involve the application of DIR between the 4DCT or BHCT exhale and inhale phase images, with the DIR motion field then used to compute breathing-induced ventilation "metrics" at the voxel level. These are mainly based on regional lung volume changes as quantified by the DIR motion field Jacobian determinant ("DIR-Jac" methods), or evaluation of air volume changes as indicated by changes in the CT number or Hounsfield Units ("DIR-HU" methods). There are many sources of variation for studies comparing CTVI to other lung function imaging including: the CT acquisition protocol and breathing manoeuvre[11], the type of DIR method used for evaluating lung motion, the type of ventilation metric employed, the presence (or not) of image pre/post processing, and the choice of metrics used to evaluate the cross-modality correlation. Some of the most salient findings are that the 4DCT or BHCT image quality can significantly impact on CTVI generation[10]; DIR based metrics in particular, are highly sensitive to image artefacts, which may impair the ability to generate accurate CTVI images in the presence of 4DCT motion artefacts due to irregular breathing. There is also heterogeneity in the methods used to define “high function” or “low function” lung; some studies apply semi-automated thresholding approaches, whereas others perform a subjective clinical assessment of the image. The use of different types of “reference” ventilation imaging modalities, such as SPECT, PET, hyperpolarised gas MRI and Xenon-CT, introduces an additional complexity in that all of these imaging modalities operate on different (if complimentary) contrast mechanisms. Similarly, the various published CTVI metrics (mainly, dealing with lung volume or density change), are all related yet clearly distinct.

Although there have been 2 recent reviews of the literature of functional lung imaging in thoracic radiotherapy, these have focused more broadly on the application of different imaging types in clinical practice by looking at the integration of functional imaging into radiotherapy planning and the benefit of reducing the dose to normal lung[12],[13]. Our paper focuses on the technical details of the CTVI validation methodology; an understanding of this is crucial to define the utility and limitations of different approaches in assessing different kinds of pulmonary pathology, and standardization of these technical details is essential if we are to move forward and validate the integration of CTVI-based radiotherapy planning in clinical trials. The CTVI literature concentrates on the assessment of regional ventilation, but we also review alternative imaging approaches and discuss whether CTVI is the most appropriate modality for imaging pulmonary physiology in thoracic radiotherapy patients with pulmonary disease other than obstructive diseases, such as pulmonary vascular and interstitial lung disease.

Hence our purpose is twofold: 1) to summarize, and assess the quality of the validation literature for CTVI using the methodology of a systematic review and 2) to compare alternative imaging modalities for assessing regional pulmonary pathology, which may provide guidance as to the use of appropriate imaging for future studies in thoracic radiotherapy patients.

Methods:

The systematic review of the CTVI validation literature review was performed following the preferred reporting items for systematic reviews and meta-analyses (PRISMA)- statement reporting standard[14]. This consensus statement defines the process and items deemed essential for transparent reporting of a systematic review. Table 1 presents our research questions in the patients, intervention, comparison, outcome, study design (PICOS) approach.

Search Strategy

Between the 5th and 7th of September 2017 searches were performed on PUBMED and EMBASE using the search term "CT Ventilation". Further studies were identified by handsearching of references and identification of studies that could possibly meet the selection criteria, as well as by direct input from the authors of the study.

These references were exported to the Systematic Review Data Repository (SRDR), an online and freely available resource provided by the US National Institute of Health (NIH) for the management of data in systematic reviews, which is available at www.srdr.ahrq.gov. The SRDR software was used to exclude duplicate studies and to assess if studies met selection criteria.

Table 1: PICOS Table for Study Question

Patient/participants	Human or animal subjects undergoing 4DCT or BHCT
Intervention	Generation of CTVI
Comparison	Accepted contrast-based imaging for regional lung function, including γ-scintigraphy, ventilation SPECT or PET, Hyperpolarized gas MRI, single or dual energy CT OR Accepted pulmonary function tests for global lung function, including spirometry and measurements for static lung function parameters
Outcome	Correlation of CTVI parameters (evaluated at the voxel, sub-organ or whole-organ level) versus clinical function (imaging or spirometry)
Study	Retrospective or prospective study quantified animal and human studies

Paper Selection and Data Extraction

Two reviewers (FH and JK) reviewed the papers independently to assess if they met study selection criteria. Study details were collected and are available in the Systematic Review Data Repository (www.srdr.ahrq.gov). If there was discordance in the assessment a third reviewer (DR) reviewed the papers. Paper quality was assessed by using the STARD Quality Dimensions for Diagnostic Tests by FH and JK[15] and the QUADAS-2[16]. The process of study selection and details of studies excluded at each step are outlined in Figure 1.

The following data were extracted from each paper:

1. Type of CT protocol (4DCT or breath hold)

2. Type of comparative or “reference” ventilation imaging modality (Ventilation PET or SPECT, Xenon CT or Hyperpolarized gas MRI)
3. Use of breathing guidance for each scan, for example audiovisual (AV) biofeedback or ventilation under anaesthesia in animal studies)
4. Details of comparison metric used (Spearman correlations, Dice similarity coefficient, or linear correlation of CoV / lobar function values etc.)
5. Details of CTVI algorithms by: DIR type (if any), Whether masking was used, What functional quantification metric(s) were used (e.g. DIR-Jac, DIR-HU, other hybrid methods that combine the Jacobian and HU changes, DIR-Hy, as well as the use of various model-based scaling factors as described in the Results section).
6. Image smoothing/filtering used at any stage in the process (for example, pre-smoothing of the input 4DCT phase images, or application of a box filter to pixel values in the output CTVI).
7. Details of the DIR assessment (if any), based on the techniques recommended by the report of the AAPM Task Group 132. This includes visual inspection of deformed images and/or motion fields, evaluation of target registration error (TRE) using expert selected anatomic landmarks, or determination of the presence of any negative values of the DIR motion field Jacobian determinant, which indicates non-physical motion.

Selection Criteria

Studies were accepted if they:

(1) Quantitatively correlated CTVI against an accepted clinical reference for measuring clinical function (either clinical/experimental imaging or spirometry)

(2) Generated CTVI from either 4DCT or BHCT without the use of a radioactive, iodinated or other imaging contrast other than air.

(3) Reported in the English language.

(4) Published in a peer-reviewed journal between the years 2000–2018. The start date was chosen as the year 2000 since the review by Simon et al. is often taken as an originating paper for the DIR-HU formulation[17].

(5) Intra-patient imaging/spirometry measurements were acquired within a reasonable timeframe (e.g. <3 months) without pulmonary intervention (namely, surgical resection or radiation therapy).

(6) Report detailed methodology for generation of CTVI images for example CT post-processing, image registration methodology, and/or other relevant algorithm/acquisition parameters.

Human and animal studies were both acceptable.

Even if meeting the above criteria, studies were rejected which:

1. Did not have an intrapatient comparison to compare the standard functional imaging or accepted pulmonary function tests such as spirometry and other measures of static lung function against CTVI.
2. Used PET/CT, SPECT/CT or contrast-enhanced CT as the ventilation imaging test modality (these will be not be classified as “CTVI” for the purposes of this paper).
3. Lacked a statement of statistical significance.
4. Did not describe, or reference to an article, in sufficient detail the method of generation of the CTVI scan.
5. Did not describe, in sufficient detail, the level of spatial detail used for the cross-modality comparison.
6. Was not a scientific paper (for example, conference abstract, patent, book, conference proceeding).
7. Was not a new investigation (for example, a review or editorial).

Statistical Methods

Given the great heterogeneity across the studies included in our analysis, which are nearly all single arm and very small a formal meta-regression was not considered meaningful. However, just over half of studies presented voxel-based Spearman rank correlations, and to facilitate cross comparison of these studies we have calculated the standard errors and present these results here. All statistical analyses were performed in Microsoft Excel (Office 2016) and Graphpad Prism 8 (Version 8.02, 2019).

Results

One hundred and forty-two records were identified through searching of Pubmed and Embase and handsearching. After the exclusion of 34 duplicate records, the abstracts of 108 records were reviewed. Twenty-eight records were rejected after abstract review for not meeting eligibility criteria. The majority of these were conference abstracts.

Paper Selection

Sixty-nine full text records were reviewed, and after detailed assessment twenty-three full text papers met selection criteria and were included in the final review (see Figure 1in the Supplementary material for details of the papers selected and eliminated at each stage). Forty-six records were rejected for 1) failing to include an intrapatient comparison with an accepted gold standard (24 records) 2) using contrast-based methods to assess ventilation (5) 3) being conference abstracts not full papers (13 records) 4) being review papers with no original data (4 records). Four papers were reported in animals and 19 papers were reported in humans. There were fourteen prospective studies. Altogether 579 human subjects were included, averaging 25.2 participants per study.

Technical aspects of 4DCT ventilation methods

Defining “normal” ventilation within a diseased lung is challenging as most assessments of high functioning lung are based on normalized values. In view of this, it is interesting that only 1 human study included normal subjects[18] and only 5 papers assess the accuracy of validation of both high and low functioning lung (labelled as “High” and “Defect” in Fig. 2)[19], [6], [20], [21], [10].

The majority of papers used deformable image registration (DIR) based methods for CTVI generation[2],[22],[23],[24],[1],[19],[25],[7],[18],[20], [26],[27],[28],[29], [4], [9],[6],[30],[10] although some papers evaluated both DIR and non-DIR based methods[6],[10] (see Table 2 for details). The most common algorithms were DIR-HU and DIR-Jac (16 and 15 papers respectively).

Seventeen out of 23 papers used smoothing, to reduce CT noise prior to computing HU based CT images. Sixteen papers used masking to reduce the impact of image artefacts from DTPA deposition upon image assessment. Twelve authors using DIR-Jac metrics included some form of mass density correction to correct for respiratory induced changes in blood mass within the lung. Please refer to Figure 1 for an overview of papers and see the discussion for details of pre and post-processing techniques used in the papers reviewed.

Diverse methods were used to analyse the relationship between CTVI and the “gold-standard”, the commonest being Spearman Rank correlations (SPC), Dice Similarity Coefficients (DSC) and linear regression (LR) methods. The Spearman r values are defined in the range [-1, 1] and indicate the degree of monotonicity of values in spatially matched voxels within the whole lung ROI with 1indicating a perfect positive correlation. The Dice similarity coefficient (DSC) describes the fractional volume overlap between two regions (in our case, ventilation/perfusion defect regions or non-defect regions) and takes a value in the range [0, 1].

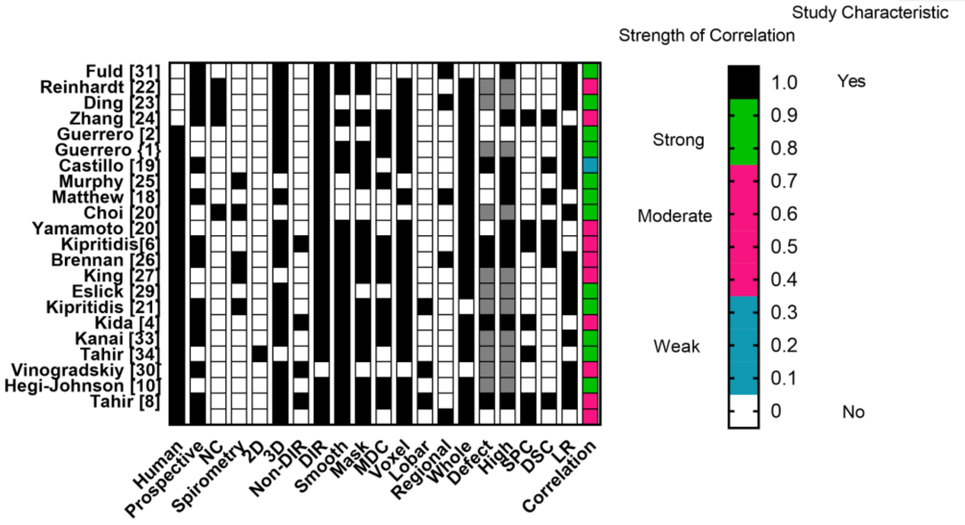


Figure 1: Characteristics of papers. This table provides an overview of study cohort and describes the gold standard correlation. Other details of the metrics such as post-processing and the methodology for comparison is described, whether this be voxel-based (Voxel)or based on a functional unit of lung (such as the lobe or whole lung) [Lobar, Regional and Whole], and whether respiratory defects (Defect) or high functioning lung (High) was evaluated. In this table the strongest average correlation reported in that paper is recorded for any type of correlation. Grey boxes indicate characteristics which are not applicable to these papers.

Abbreviations: Normal Comparator (NC), Deformable Image Registration (DIR), Dice Similarity Coefficients (DSC) and linear regression (LR) methods, smoothing (Smooth) or masking (Mask), mass density corrected metric (MDC), Spearman Rank Correlation (SPC), Dice Similarity Coefficient (DSC), Linear Regression (LR) are also described.

Table 2a Animal Studies and Table 2b Human Studies: Summary of CTVI studies. CTVI generation is defined as non-DIR and DIR based, which includes DIR Jacobian (DIR-Jac), DIR-HU and other hybrid DIR approaches (DIR-Hy). Papers used various metrics for assessing the strength of correlation, including linear regression (LR), Spearman Rank correlations (SPC) and Dice Similarity Coefficients (DSC).

Abbreviations: Coefficient of Variance (CoV), forced expiratory volume in 1 second (FEV1), FVC forced vital capacity, Δ Vol ventilation calculation of ventilation based on change in volume, Region of Interest (ROI).

Table 2a: Animal Studies of CT Ventilation

Study author (year)	Subjects No.	Comparative Imaging Modality	CTVI Generation	Type of correlation	Highest level of correlation	Range of correlation
Fuld et al (2008)**[31]**	4 sheep	Xenon-CT	Change in volume as assessed by change in HU units	Voxel based LR	LR *R*2 of 0.76	0.56–0.76
Reinhardt et al (2008)**[22]**	5 sheep	Xenon-CT	DIR-Jac	Voxel Based LR	LR *R*2 of 0.80	0.64–0.80
Ding et al 2012 **[23]**	4 sheep	Xenon-CT	2 variants of DIR-Jac (SAJ, SACJ) and DIR-HU	Voxel based LR	SAJ *r=0.97* SACJ *r=0.994* DIR-HU *r=0.952*	0.836-0.97 0.888-0.994 0.893-0.952
Zhang et al (2016)**[24]**	4 sheep	Xenon CT	3 metrics tested: 1. DIR Δ Vol 2.DIR-Jac 3. DIR- HU	Voxel based SPC	Δ Vol 0.61 Jacobian 0.61 HU 0.42	0.29-0.61 0.31-0.61 0.17–0.42

Table 2b: Human studies of CT ventilation

Study author (year)	Subjects No.	Comparative Imaging Modality	CTVI Generation	Type of Correlation	Highest average level of correlation	Range
Guerrero 2005**[2]**	22	Measured tidal volume on CT	DIR-HU	LR of lung volumes and tidal volumes.	DIR HU LR *r*= 0.985	LR *r* ranged from 0.982 -0.985
Guerrero 2006**[1]**	3 lung cancer RT patients	Measured tidal volume on CT	DIR-HU	Whole lung based	DIR-HU LR *r* = 0.985	NR

Castillo et al 2010**[19]**	7 thoracic oncology patients	DTPA-SPECT	DIR- HU DIR-Jac	Voxel based DSC	Highest average DSC was for 0-20% voxels	DSC ranged from 0.2-0.35
Murphy et al 2012**[25]**	216 patients with COPD	Spirometry	DIR-HU	Whole lung and lobar assessment LR of GOLD stage FEV1 and FEV1/FVC	Median *r* value of 0.87 for whole lung	0.85-0.91
Matthew et al 2012**[32]**	11 lung cancer patients	^{3}HE hyperpolarized MRI	DIR- HU	DSC to compare ventilated volume (VV) in both whole lung and lung ipsilateral to and contralateral to cancer	DSC 0.89±.01	0.69-0.95
Choi et al 2013**[18]**	30 asthma patients, 14 control subjects	PFT's comparison of Total Lung Capacity (TLC) and Air volume (AV) at exhale.	DIR-Jac derived from breath-hold CT	Global lung function LR	*R=0.87* for Total lung volume in severe asthmatics	*r* =0.78 to 0.87
Yamamoto et al 2013**[20]**	9 patients with thoracic cancer.	DTPA-SPECT	DIR- Jac	Voxel based SPC, DSC for segmented low-functional lung regions	Best Spearman rank 0.80 Best DSC 0.8	Average 0.69±0.26 0.71±
Kipritidis et al 2014**[6]**	12 lung cancer patients	PET-Galligas	DIR-HU and DIR-Jac with and without density scaling	Voxel based SPC	Density-scaled HU Spearman $\bar{r}$ = 0.28 ± 0.13 and DSC (lowest 20%) = 0.52 ± 0.09	DIR-Jac 0.25±0.17
Yamamoto et al 2014**[5]**	18 patients-all with thoracic cancer	DTPA-SPECT and PFT's (spirometry and	DIR- HU DIR-Jac .	Voxel based DSC to quantify overlap betweenV4D	DSC for DIR- HU 0.39	Average DSCs were: DIR-HU 0.39 ± 0.11

		measurement of DLCO)		CT and VSPECT defect regions Pearson correlation with FEV1 and FEV1/FVC		DIR-Jac 0.36 ±0.13 FEV1/FVC strongly correlated with 25% voxel value (0.73) and strongly negatively correlated with defective lung (-0.63).
Brennan et al 2015[26]	98 patients with lung cancer	Spirometry	DIR- HU and DIR-Jac	LR to compare spirometry and CTVI CoV V20, and visually defined defects.	Correlation coefficient ~ 0.7 for HU	DIR-HU CoV CC between: DIR-HU FEV1 0.72 FEV1/FVC 0.67 DIR-Jac FEV1 0.40 FEV1/FVC 0.38
King et al. 2015[27]	30 Thoracic radiotherapy patients	Tidal volume from 4DCT and CTVI compared	DIR-Jac DIR-Hy	Whole lung based Pearson correlation coefficient	TVInt and TVCT was 0.92 (P<.01) TVJac and TVCT was 0.97 (P<.01).	Not reported
Eslick et al 2015[29]	11 lung cancer patients	PET Galligas	DIR- HU	Lobar volumes and ventilation compared with LR	*r*-value 0.96 for comparison of lobar ventilation from CTVI and PET-Galligas.	Average *r*-value between CTVI analysed lobar ventilation vs. lobar volumes was 0.78
Kipritidis et al 2016[21]	25 lung cancer patients	PET-Galligas	Non-DIR based HU DIR-HU DIR-Jac	Voxel based SPC	CTVI-HU 0.50	Mean ±SD correlation with Galligas PET was r = (0.50±0.17), (0.42±0.20),

						and (0.19±0.23) for the CTVI-HU, DIR-HU, and DIR-Jac methods
Kida et al 2016**[4]**	8 thoracic cancer patients	DTPA SPECT and Spirometry	DIR-HU and DIR-Jac	Radiotherapy lung metrics compared with Pearson correlation and LR	0.94 for DIR-HU	DIR-HU Pearson R=0.94 and linear regression = 0.71 DIR-Jac R=0.85; slope=0.5
Kanai et al 2016**[33]**	11 lung cancer patients	Planar Kr images	DIR-HU and DIR-Jac	Voxel based SPC	HU 0.875	Mean ± SD: DIR-HU 0.875±0.07 DIR-Jac 0.803±0.114
Tahir et al 2016**[34]**	30 patients with sputum eosinophilia and asthma	Hyperpolarized ^{3}He MRI	Breath-hold CT at total lung capacity and functional residual capacity used to assess change in volume of lobes	Lobar Pearson correlation of all lobar regions	0.65	Range of Pearson correlations not stated
Vinogradskiy et al 2017**[30]**	16 lung cancer patients	DTPA SPECT and spirometry	DIR- HU	Global lung function ROC analysis to compare, 4DCT-ventilation-based preop FEV1 vs. SPECT based preop FEV1.	0.99 correlation coefficient for prediction of ventilation changes after lobectomy using CTVI	Pneumonectomy: Correlation coefficient .80(0.81for nuclear medicine-ventilation And 0.78 for nuclear medicine-perfusion). Lobectomy: Correlation coefficient was 0.99 for CTVI.
Hegi-Johnson et al 2017**[10]**	11 lung cancer patients	Technegas SPECT	CTVI-HU, DIR-HU and DIR-Jac	Voxel based Spearman Rank and DSC, lobar based Pearson Correlation	Non-defect regions: CTVI HU, DIR-HU and	Defect regions mean DSC were 0.39, 0.33, and 0.44. Spearman r

				and whole lung CoV	DIR-Jac mean DSC of 0.69,0.68, and 0.54.	:0.26, 0.18 and -0.02 for CTVIHU, DIR-HU and DIR-Jac respectively
Tahir et al 2018**[8]**	11 lung cancer patients	^{129}Xe and ^{3}He MRI	DIR-HU, DIR-Jac, Specific gas volume change	Spearman rank correlations of different ROI sizes	DIR-HU SPC *R*= 0.37, DIR-Jac SPC 0.31, Specific gas volume 0.34	Voxel-level: 0.1-0.8 ROI 20x20 voxels: *R=0.2-0.9*

CTVI metrics

Evaluating the precision of DIR

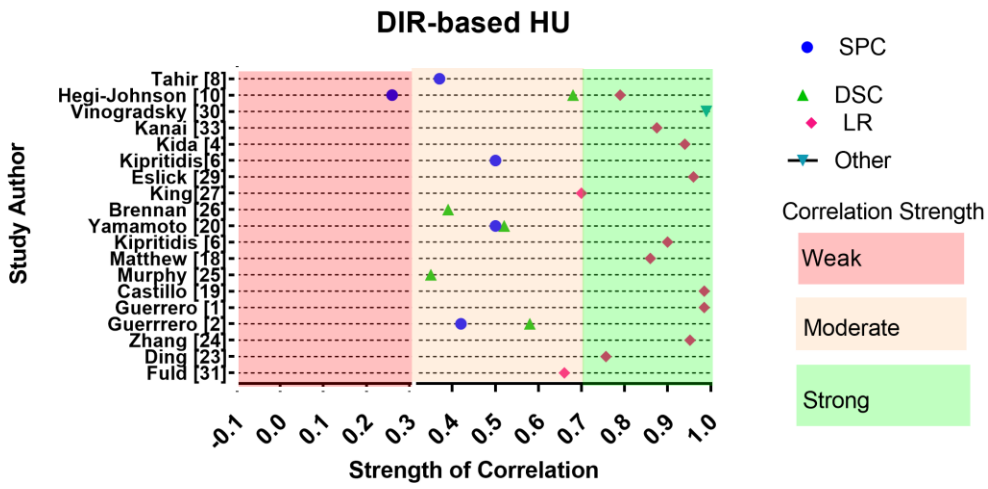
The majority of papers selected (22 out of 23 papers) used a DIR based approach to CTVI generation, although several of these also tested non-DIR based approaches on the same subject group. The commonest DIR approach was B-spline based with 12 papers using some variation of B-spline algorithms [1],[2],[6],[8] ,[10],[18],[21],[23],[24],[25] ,[27], [29]. Just over half of papers (13 out of 23) either discussed the method used to assess registration accuracy or referenced this within the text. Target registration errors (TRE) were quantified or referenced for DIR methods in 11 papers, with the other papers using visual assessment (2) and semi-automated landmark analysis (1). Most papers reported TRE of <1.5 mm indicating that DIR was accurate. However, the lack of reporting in 9 of the papers using a DIR CTVI approach is potentially problematic, as this is critical for the accurate calculation of regional ventilation. Please see our supplementary files for details of image registration methodology and registration accuracy assessment in individual papers.

Comparing ventilation metrics

Eighteen papers evaluated HU based metrics, with all papers evaluating DIR-HU; within this group, 3 papers evaluated non-DIR HU metrics in addition to DIR based approaches. Fifteen papers evaluated DIR-based Jacobian metrics. For details of the comparison methodology and strength of correlation, please refer to Figures 3a and 3b.

Average ± S.D. linear regression correlation coefficients were 0.73±0.24 and 0.86±0.11 for DIR-Jac and DIR-HU respectively, and the SPC were 0.45±0.31and 0.41±0.11 for DIR-Jac and DIR-HU respectively.

Figure 2a



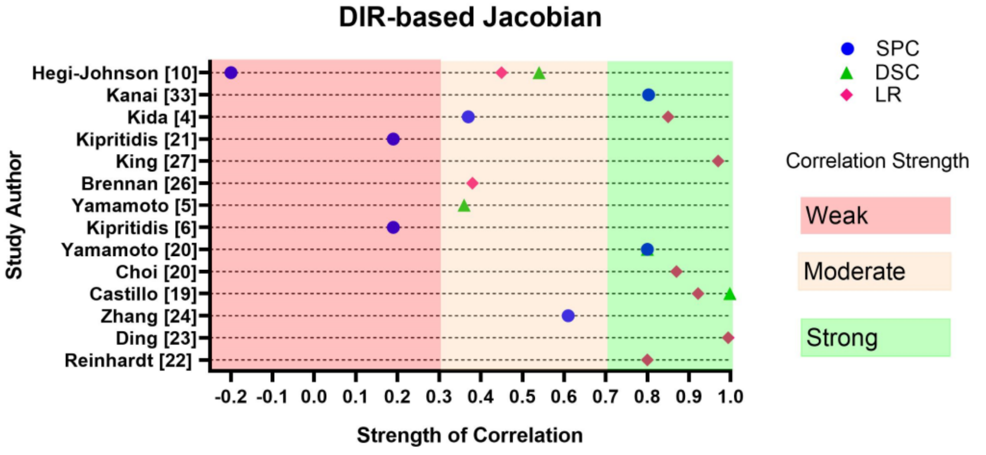


Figure 2b

Figure 2a and 2b: SPC, DSC and LR for DIR-HU (Fig. 3a) and DIR-Jacobian (Fig, 3B) respectively. Correlations are graded as either weak (0-0.3), moderate (<0.3-0.7) or strong (<0.7-1.0).

The number of papers consistently reporting on the size of region of interest (ROI) used in comparison was too small to draw strong conclusions, although in 2 papers which compared small ROI's (<1cm^3) vs. large ROI's (4 mm slices of the whole lung from top to bottom) the linear regression correlation coefficients were quite different, being 0.56±0.49 and 0.84±0.16 respectively.

Comment [f1]: This figure has been reformatted in accordance with suggestions from our reviewers.

Discussion

The primary function of the lung is gas exchange of which ventilation, perfusion and diffusion are fundamental components; oxygen from the air and carbon dioxide excreted by the body and dissolved in the blood is exchanged across the alveolar membrane. Even in healthy lungs this process is dynamic, with heterogeneous ventilation throughout the lung, partly because of mechanical issues such as the difference in pressure between the top and bottom of the lung, and partly because of rapid changes within the pulmonary vasculature. In patients with lung disease the local pathology can differ according to the underlying aetiology. Patients with severe COPD have obstructive pulmonary function test results and have large areas of the lung that are not ventilated due to flow limitation/hyperinflation. In patients with emphysema, spirometry can be remarkably normal, but diffusion is impaired due to destruction of alveoli. In asthmatic patients, obstruction can be reversible, but in severe cases it can be irreversible as well. In pulmonary vascular disease, ventilation may be adequate, but the ventilated lung is inadequately perfused. Finally, in interstitial lung disease, there is a restrictive pulmonary function, with both ventilation (although in a restrictive pattern: i.e. normal or increased FEV1/FVC ratio, but decreased tidal volume, decreased TLC and FVC) and perfusion present in the diseased areas of lung, but gas diffusion may be impaired depending on the cause of the restrictive pattern (interstitial lung disease vs. thoracic wall pathology) To make matters more complicated, patterns can overlap with multiple pathologies present in patients with severe lung disease.

All of the papers included in this review investigate the use of imaging technologies to define regional ventilation, although some have included perfusion scans as well[6],[10]. Given the dependence of gas exchange on perfusion, ventilation and diffusion, can imaging technologies which do not assess perfusion or diffusion provide useful information? We know from previous studies that perfusion scans can demonstrate changes after radiotherapy, and there is some evidence that they may strongly correlate with pulmonary pathology[35],[36]. Therefore, it seems likely that CTVI’s inherent focus on ventilation has limited its applicability to patients with mainly gas exchange problems without an impairment in ventilation. In this review, we have chosen to limit our assessment to papers focusing on ventilation for two reasons. Firstly, CT scans are now used as part of the standard workflow in radiotherapy planning and the assessment of surgical resection candidates. They are cheap and widely available making CTVI a highly accessible technology. Secondly, CTVI has rapidly progressed into clinical trials, and we wished to assess the robustness of the literature, which has focused largely on technical rather than clinical validation measures.

Defining the "gold standard': Alternative Imaging Methods to Assess Regional Ventilation with hyperpolarized gas MRI and nuclear imaging.

Nuclear Medicine Assessments of Pulmonary Function

V/Q SPECT and V/Q SPECT-CT are established modalities for the assessment of regional pulmonary function. Isotopes in common use include 99mTc-labeled particulate aerosols such as 99mTc-diethylenetriaminepentaacetic acid (99mT c-DTPA) or the ultrafine carbon–labelled nanoparticle 99mTc- Technegas (Cyclomedica). Subsequently, 99mTc-macroaggregated albumin is administered and a perfusion scan acquired[37]. V/Q PET-CT using Gallium-68 (^{68}Ga) has been developed, allowing higher resolution imaging of radioisotope uptake [38], but is unlikely to be widely clinically implemented due to the need for a ^{68}Ga- generator within departments. All these radioisotope techniques suffer to some extent from issues of clumping in the central airways, although peripheral airway distribution can be improved by using smaller particles such as Technegas or Galligas, and by careful ventilation of the patient with deep tidal breathing during radioisotope inhalation to ensure even distribution throughout the lung parenchyma. Their great advantage is the ability to simultaneously image perfusion, and they have been shown to be sensitive and specific for the diagnosis of pulmonary emboli, although they are not routinely used for the diagnosis of other pulmonary pathologies[37].

Spirometry based pulmonary function tests (PFT)

Spirometry remains the most common PFT performed in the assessment of lung cancer patients, but is limited to the assessment of obstructive pulmonary disease (and although it may be abnormal in restrictive pulmonary pathology, is not diagnostic) and does not provide information about the function of the lung parenchyma where gas exchange occurs. Also, spirometry does not measure hyperinflation in the lung; these areas of hyperinflated lung, as represented by the large residual volume (RV) in patients with severe obstructive COPD do not contribute to gas exchange. Spirometry is highly dependent on respiratory effort, and incorrectly performed spirometry may be non-diagnostic. This introduces great variability into their performance by individual patients who may have similar pathological profiles within the lung parenchyma.

MRI using hyperpolarized gases and the assessment of regional lung function

Hyperpolarized gas MRI gives detailed information about both the lung microstructure and regional ventilation. Both ^{3}He[39],[40] and ^{129}Xe[41],[42] have been used to assess regional pulmonary physiology. Studies with ^{3}He have shown strong correlations to spirometry in patients with a variety of pulmonary pathologies including COPD and asthma[43],[44]. ^{129}Xe diffuses through the alveolar membrane and into the red blood cells in the blood stream, producing distinct resonant signal frequencies in each vascular compartment and has great potential for studying gas diffusion between the lungs and the blood stream.

Diffusion-weighted ^{3}He and ^{129}Xe MRI can be used to evaluate the apparent diffusion coefficient (ADC) of lung parenchyma, which can give highly detailed information about lung microstructure at the alveolar level. These studies have been shown to correlate with early ultrastructural changes seen on MRI.

Comment [f2]: This section has been simplified in accordance with the recommendation fo our reviewers.

Hyperpolarized gas MRI static ventilation images show the regional distribution of inhaled noble gas and have been used by Tahir et al, who demonstrated moderate to strong correlations between them and CTVI at the lobar and voxel levels [34], [8], [45]. In many respects, hyperpolarized gas MRI provides a level of anatomical and physiological detail that is not yet available in 4DCT. Initially, implementation was limited by the high cost of ^{3}He but the maturation of ^{129}Xe MRI has made this into an accessible technology for future studies.

Common elements and limitations in CT Ventilation Studies

CT Acquisition and Issues of CT quality

Initial steps towards the development of CTVI occurred in the early 2000’s with the publications of methods to derive information about regional ventilation from CT datasets[1],[31]. The three key steps in the generation of a DIR-based CTVI image are essentially identical: (1) acquisition of a respiratory-correlated CT scan, most commonly 4DCT or sometimes BHCT, (2) application of DIR typically between the exhale and inhale phase images, and (3) computation of a ventilation metric either directly on the DIR motion field (e.g. DIR-Jac) or using the motion field to process changes in HU values for spatially registered voxels in the inhale and exhale images (DIR-HU) or both (DIR-Hy).

The first 10 years of development of the CTVI methodology have been marked by the wide variety of technical processes that have been explored and are currently in use to generate CTVI images (see Figure 3).

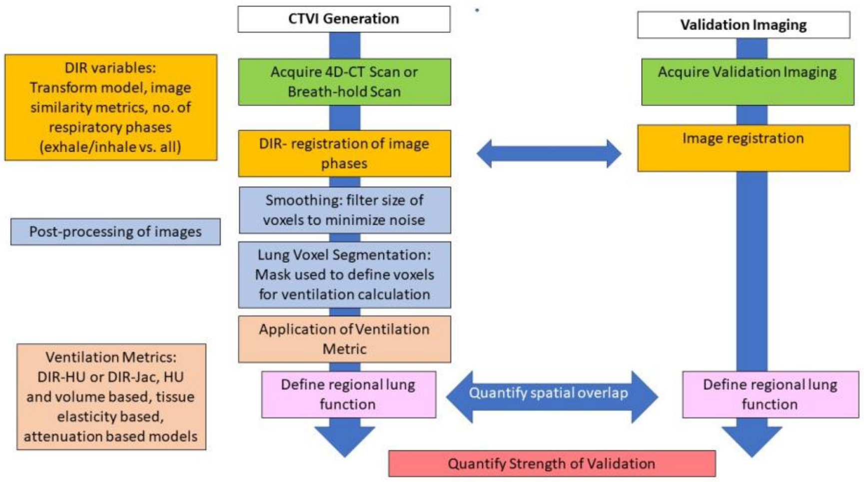


Figure 3: Overview of typical workflow for a DIR-based CTVI validation study including common steps for CTVI generation (left arm) and validation against contrast-based ventilation imaging (right arm). This demonstrates the multiple variables that should be considered when reporting and assessing CTVI validation studies. These variables include details of the DIR methodology (orange boxes), including the metrics used to perform DIR and the details of the respiratory phases included in the registration. Post-processing (blue boxes) may occur both before or after the CTVI image is generated (please see the text for details of this). Finally, the metrics that may be applied are diverse (brown boxes), although DIR-Jac and DIR-HU are the commonest. Common steps in the CTVI validation workflow are in the same colour as the CTVI generation pathway.

The majority of CTVI validation studies have focused on the use of 4DCT, which involves image reconstruction into 5-10 different respiratory “phase bins” based on the synchronous acquisition of CT projection data and a breathing motion signal [46]. In clinical human studies, the 4DCT scan is often performed under free-breathing (FB), which can lead to the well-known problems of anatomic truncation, duplication and blurring artefacts that arise due to irregular breathing motion, for example, coughing or changes in breathing period / amplitude during the scan. It has been reported that up to 90% of clinical 4DCT scans suffer anatomic imaging artefacts of magnitude >4mm[47].

Some CTVI studies have attempted to overcome the issue of irregular breathing using some form of AV biofeedback to increase the regularity of breathing[28], [29], or by using different methods of sorting the CT by matching bins based on anatomical features rather than the phase of respiration[20]. CTVI can also be derived from exhale/inhale image pairs acquired during breath-hold[2],[29],[18], although acquiring these images in thoracic cancer patients with impaired respiratory function can be challenging. By comparison, in animal studies the subjects are anaesthetized during the 4DCT scan, resulting in highly regular breathing motion with minimal 4DCT reconstruction artefacts. Other interesting methods to reduce image noise included using anatomic sorting rather than phase based sorting to reduce artefacts within 4DCT images[20].

Pre/Post-Processing of Images

The majority of studies in this review applied smoothing to either the input 4DCT images, the resulting CTVIs, the corresponding contrast-based ventilation images, or some combination of these. Various smoothing techniques were applied including applying a Gaussian, median or averaging filter to minimize noise within the raw 4DCT [1], [4], [10] [25],[24], [19],[25], [18], [26],[29], [21],[33],[30], or CTVI image[32],[33], [10] or by applying smoothing functions during the generation of the CTVI image[18],[20]. There is some evidence to suggest that the method of filtering can affect the strength of the correlation coefficients, although this appears to plateau at a median filter radius of 3x3x3 voxels[8].

Comment [f3]: The discussion on lung masking has been omitted in order to simplify and streamline this discussion

Ventilation Metric

Ventilation metrics may be classified into DIR-based vs. non-DIR based algorithms. The commonest DIR metrics evaluate vowel-wise HU changes between spatially aligned images (DIR-HU)[1],[2] or regional volume changes based on the Jacobian determinant of the DIR motion field (DIR-Jac)[22]. A small number of "hybrid" metrics have also been investigated, which may combine information about HU and volume changes to model lung elasticity as an alternate surrogate for lung function (DIR-Hy)[8] ,[27]. Non-DIR HU metrics ($_{\text{non}}$DIR-HU) have been found to be potentially robust against 4DCT motion artefacts, and use average HU values to model blood-gas exchange in the lung parenchyma[10],[21]. Physiological ventilation is a process of blood-gas exchange with diffusion of oxygen and carbon dioxide across the alveolar membrane. The HU values of a voxel reflect the air and tissue content of that particular part of the lung parenchyma, and in the nonDIR-HU model this is used as a surrogate for the capacity of blood-gas exchange at each voxel. As this approach relies on the average intensity projection of the 4DCT, it is closer to the average scans acquired in nuclear medicine imaging, and also reduces the impact of respiratory artefacts on CTVI image quality.

Statistical methods used for cross-modality comparisons

A number of methods have been used to analyse the accuracy of CTVI with respect to paired contrast-based comparator ventilation scans, including the SPC, DSC and Linear regression methods.

The size of the region of interest (ROI), relative to the pixel or voxel dimensions, has been seen to affect the strength of cross-modality correlation values obtained[8], with stronger correlations seen for larger ROIs. It is thought that averaging out ventilation values over larger ROIs can mitigate errors created by mis-registration between the CTVI and contrast-based comparator scans, errors in the DIR process or imaging artefacts in either the 4DCT or contrast-based comparator scans [10], [31]. This is particularly relevant in studies that have used clinically acquired imaging and hence, although voxel based comparisons are important to benchmark different CTVI and DIR methodologies, CTVI may be most robust when defining loco-regional function at larger anatomical distance scales, such as lung lobes[10],[29].

Different semi-automated approaches exist for thresholding lung ventilation images into high or low-function lung zones; however, there exists no consensus on which is the best approach. Since CTVI is amenable to quantitative analysis, a number of CTVI studies have defined low function lung as referring to those lung voxels with ventilation values less than the 20th-30th percentile ventilation for that patient. In contrast, clinical assessment of nuclear medicine ventilation images is usually based on the visual analysis of scans by a physician. Relatively little work has been carried out to validate semi-automated thresholding of lung ventilation images against clinical assessment [21]. It is encouraging that the correlation of functional dose with toxicity outcomes (e.g. Grade 3+ pneumonitis) may be relatively stable despite different methods for weighting the ventilation values in CTVI [48].

Assessing the Quality of the Literature

CTVI is a promising imaging technology, however the CTVI literature is hampered by the heterogeneity in the methodology for CTVI generation, the choice of reference modality and method of cross-modality comparison and issues of study quality, with only 14 small prospective studies. These issues of study quality and heterogeneity explain the wide variation in the strength of correlation seen. Indeed, even comparing a single parameter across the literature is difficult, with SPC varying between SPC were 0.45±0.31and 0.41±0.11 for DIR-Jac and DIR-HU respectively. The wide spread in the standard deviations highlights the small numbers of papers, low patient numbers and the variability in methods of analysis in this literature. The small sample sizes and variability in results is well illustrated in Figure 4, in which we present a Forest Plot of the voxel-based SPC results.

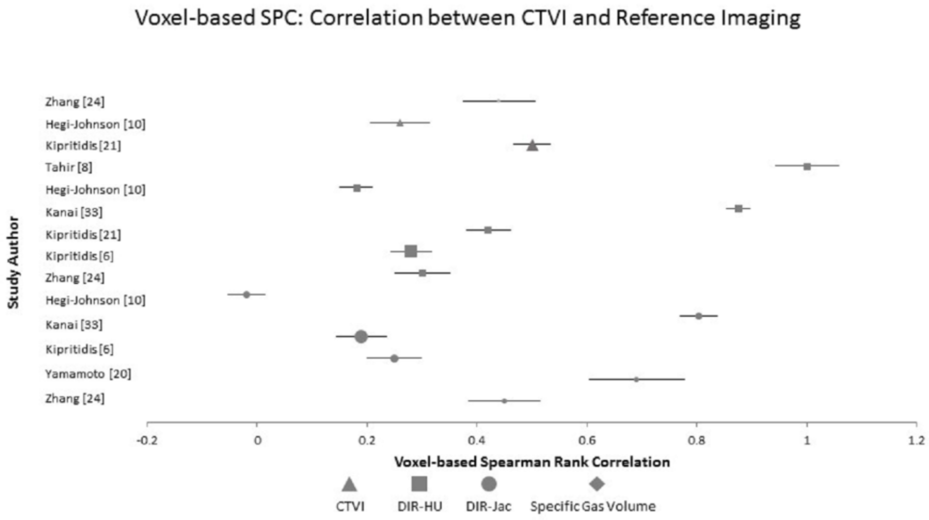


Figure 4: Forest plot illustrating Mean±Standard Error for Studies that present voxel-based SPC analyses for different ventilation metrics. Note the relatively small number of studies that present similar results, and the wide error bars. The size of the icon is proportional to patient numbers.

Comment [f4]: New figure to emphasize the statistical limitations of the current evidence base.

The use of comparators at the lobar or whole-lung level can mask gross errors at the voxel level. Most papers using PFTs as the comparators relied on spirometry[25],[18], [28], [30], and although it may be argued that spirometry provides an adequate measure of ventilation and is hence a valid modality for comparison with CTVI, it is likely to underestimate the impact of pulmonary diseases that are not adequately reflected by spirometry, such as emphysema and pulmonary vascular disease.

In Table 3, we provide a summary of the key aspects of the CTVI validation literature. Both DIR and non-DIR CTVI ventilation metrics have achieved robust correlations, but particular approaches may suit different datasets. For example, as they use the average projection of the 4DCT to generate the CTVI image, non-DIR approaches may be more robust when using clinically acquired 4DCT with significant respiratory artefacts. On the other hand, very high-quality images may be acquired clinically by using BH approaches. Although, BH was only investigated in 4 human studies[2],[8],[25],[29] this may help to overcome the quality issues created by irregular breathing motion. Other strategies, such as AV biofeedback may also be useful, but there is insufficient information in the current literature to advocate it for routine implementation. Similarly, the heterogeneity in the methodologies used for post-processing make it difficult to make recommendations.

Table 3: Summary of Findings Table

Key Findings

Question: Can CTVI defined regional ventilation provide the same information as contrast-based ventilation imaging and pulmonary function tests?

Population: Patients having 4DCT or BHCT, the majority of whom are undergoing radiotherapy planning

Reference standard: contrast-based imaging of regional lung function and/or spirometry

Findings:

- LR (Average ± S.D.): 0.73±0.24 and 0.86± 0.11 for DIR-Jac and DIR-HU
- SPC (Average ± S.D.): 0.45±0.31and 0.41±0.11 for DIR-Jac and DIR-HU

Heterogeneity in reporting standards reduces the ability to assess statistical significance.

Recommendation		No of relevant studies	Strength
DIR vs. Non-DIR	There is insufficient evidence to recommend DIR over non-DIR approaches. If DIR based CTVI metrics are used the quality of the DIR should be assessed	Only 4 non-DIR studies	Moderate Weak to moderate
Selection of metric	There is insufficient evidence to recommend one CTVI metric over another and further head to head is required.	19 DIR-HU 15 DIR-Jac	Moderate
CT acquisition and processing	CTVI based on both 4DCT and BHCT have shown moderate to strong correlations with clinical gold standards. There is insufficient evidence to guide the use of one method of acquisition above another.	4 BH studies	Moderate to strong

	There is insufficient evidence to support the use of AV biofeedback routinely in the acquisition of CT images: studies using AV biofeedback have similar strengths of validation compared to non-AV biofeedback studies.	2 AV biofeedback studies	Weak
	Post-processing such as smoothing and masking may influence the quality of the CTVI image, but there is insufficient evidence to recommend for or against their use.	18 papers used post-processing, but methodology was heterogeneous.	Weak
Clinical Validation Modality	The majority of validation studies have used 3D-imaging and this should be used for future validation studies for them to be benchmarked against existing methodologies.	19 papers used 3D imaging	Moderate to strong
	Standardized criteria for clinical assessment (e.g. GOLD criteria) should be used for validation modalities to increase the clinical relevance of future validation studies.	5 papers reported validation against spirometry	Moderate

This lack of standardization makes it difficult to determine whether CTVI is robust enough to implement in clinical practice. To improve the standardization of future validation studies, we recommend: 1) the use of high quality, prospectively collected datasets and that, where possible, these should include normal and disease cohorts 2) documentation of respiratory function of individual subjects as assessed by pulmonary function tests which are recognized clinical standards (spirometry or cardio-pulmonary exercise testing), 3) incorporation of high quality ventilation imaging as a comparator, 4) report detailed methodology for their CTVI imaging, including the details of CT image acquisition, use of post-processing techniques such as smoothing, masking, methodology of DIR and an assessment of its accuracy 5) report the strength of validation across both high functioning and low functioning lung using a cohort of standardized statistical assessments. We would suggest that these tests should include voxel-based assessments using the SPC and DSC as a minimum to allow new modalities to be benchmarked against existing studies as well as more clinically relevant regional volumes. We would also encourage the investigation of other comparative methodologies. For example, although outside of this review, it would be powerful to compare CTVI defined regional ventilation with histopathological specimens and this could facilitate the development of imaging surrogates for different types of COPD, increasing the non-invasive options to diagnose these diseases.

Finally, given the complexity of pulmonary pathology found in thoracic oncology patients, it is possible that CTVI may be most useful in patients who are affected by obstructive lung diseases such as COPD, but alternative imaging modalities may be required in other patients. For example, hyperpolarized dissolved-phase ^{129}Xe MRI may be particularly useful in patients with interstitial lung disease[49], and V/Q SPECT and PET in patients with pulmonary vascular pathology.

Judgement on the quality of the CTVI validation literature will ultimately be determined by the outcome of prospective clinical studies investigating CTVI implementation in radiotherapy patients (NCT02528942, NCT02308709, NCT02843568) [50], [48]. The awaited publication of the VAMPIRE challenge, which uses a range of clinical imaging datasets to compare DIR and CTVI methodologies will also provide new insights into the variabilities and uncertainties associated with this technology[51].

CTVI has shown moderate to strong voxel-based correlations in most human studies. However, CTVI is being increasingly incorporated into the clinical workflow of thoracic radiotherapy and is undergoing clinical validation. Our results show that further refinement and standardization of CTVI methodology will enable better comparative studies and a more robust application of this technology in clinical practice. CTVI appears to be of relevance in radiotherapy planning, particularly in patients whose main pulmonary impairment is not a gas exchange problem, and awaits clinical validation in prospective clinical trials.

References

[1] Guerrero T, Sanders K, Castillo E, Zhang Y, Bidaut L, Pan T, et al. Dynamic ventilation imaging from four-dimensional computed tomography. Phys Med Biol 2006;51:777–91.

[2] Guerrero T, Sanders K, Noyola-Martinez J, Castillo E, Zhang Y, Tapia R, et al. Quantification of regional ventilation from treatment planning CT. Int J Radiat Oncol Biol Phys 2005;62:630–4.

[3] Yamamoto T, Kabus S, Bal M, Keall P, Benedict S, Daly M. The first patient treatment of computed tomography ventilation functional image-guided radiotherapy for lung cancer. Radiother Oncol 2016;118.

[4] Kida S, Bal M, Kabus S, Negahdar M, Shan X, Loo BW, et al. CT ventilation functional image-based IMRT treatment plans are comparable to SPECT ventilation functional image-based plans. Radiother Oncol 2016;118:521–7.

[5] Yamamoto T, Kabus S, Lorenz C, Mittra E, Hong JC, Chung M, et al. Pulmonary ventilation imaging based on 4-dimensional computed tomography: Comparison with pulmonary function tests and SPECT ventilation images. Int J Radiat Oncol Biol Phys 2014;90:414–22.

[6] Kipritidis J, Siva S, Hofman MS, Callahan J, Hicks RJ, Keall PJ. Validating and improving CT ventilation imaging by correlating with ventilation 4D-PET/CT using 68Ga-labeled nanoparticles. Med Phys 2014;41:011910.

[7] Mathew L, VanDyk J, Etemad-Rezai R, Rodrigues G, Parraga G. Hyperpolarized (3) He pulmonary functional magnetic resonance imaging prior to radiation therapy. Med Phys 2012;39:4284.

[8] Tahir BA, Hughes PJC, Robinson SD, Marshall H, Stewart NJ, Norquay G, et al. Spatial comparison of CT-based surrogates of lung ventilation with hyperpolarized Helium-3 and Xenon-129 gas MRI in patients undergoing radiation therapy. Int J Radiat Oncol Biol Phys n.d.

[9] Kanai T, Kadoya N, Ito K, Kishi K, Dobashi S, Yamamoto T, et al. Evaluation of four-dimensional computed tomography (4D-CT)-based pulmonary ventilation: The high correlation between 4D-CT ventilation and 81m Kr-planar images was found. Radiother Oncol 2016.

[10] Hegi-Johnson F, Keall P, Barber J, Bui C, Kipritidis J. Evaluating the Accuracy of 4D-CT Ventilation Imaging: First Comparison with Technegas SPECT Ventilation. Med Phys 2017;44:4045–55.

[11] Mistry NN, Diwanji T, Shi X, Pokharel S, Feigenberg S, Scharf SM, et al. Evaluation of fractional regional ventilation using 4D-CT and effects of breathing maneuvers on ventilation. Int J Radiat Oncol Biol Phys 2013;87:825–31.

[12] Ireland RH, Tahir BA, Wild JM, Lee CE, Hatton MQ. Functional image-guided radiotherapy planning for normal lung avoidance. Clin Oncol 2016;28:695–707.

[13] Bucknell NW, Hardcastle N, Bressel M, Hofman MS, Kron T, Ball D, et al. Functional lung imaging in radiation therapy for lung cancer: A systematic review and meta-analysis. Radiother Oncol 2018.

[14] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151:264–9.

[15] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Radiology 2015;277:826–32.

[16] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36.

[17] Simon BA. Non-invasive imaging of regional lung function using x-ray computed tomography. J Clin Monit Comput 2000;16:433–42.

[18] Choi S, Hoffman EA, Wenzel SE, Tawhai MH, Yin Y, Castro M, et al. Registration-based assessment of regional lung function via volumetric CT images of normal subjects vs. severe asthmatics. J Appl Physiol 2013;115:730–42.

[19] Castillo R, Castillo E, Martinez J, Guerrero T. Ventilation from four-dimensional computed tomography: density versus Jacobian methods. Phys Med Biol 2010;55:4661.

[20] Yamamoto T, Kabus S, Lorenz C, Johnston E, Maxim PG, Diehn M, et al. 4D CT lung ventilation images are affected by the 4D CT sorting method. Med Phys 2013;40.

[21] Kipritidis J, Hofman MS, Siva S, Callahan J, Le Roux P-Y, Woodruff HC, et al. Estimating lung ventilation directly from 4D CT Hounsfield unit values. Med Phys 2016;43:33–43.

[22] Reinhardt JM, Ding K, Cao K, Christensen GE, Hoffman EA, Bodas SV. Registration-based estimates of local lung tissue expansion compared to xenon CT measures of specific ventilation. Med Image Anal 2008;12:752–63.

[23] Ding K, Cao K, Fuld MK, Du K, Christensen GE, Hoffman EA, et al. Comparison of image registration based measures of regional lung ventilation from dynamic spiral CT with Xe-CT. Med Phys 2012;39:5084–98.

[24] Zhang GG, Latifi K, Du K, Reinhardt JM, Christensen GE, Ding K, et al. Evaluation of the ΔV 4D CT ventilation calculation method using in vivo xenon CT ventilation data and comparison to other methods. JACMP 2016;17:550–60.

[25] Murphy K, Pluim JP, van Rikxoort EM, de Jong PA, de Hoop B, Gietema HA, et al. Toward automatic regional analysis of pulmonary function using inspiration and expiration thoracic CT. Med Phys 2012;39:1650–62.

[26] Brennan D, Schubert L, Diot Q, Castillo R, Castillo E, Guerrero T, et al. Clinical validation of 4-dimensional computed tomography ventilation with pulmonary function test data. Int J Radiat Oncol Biol Phys 2015;92:423–9.

[27] King MT, Maxim PG, Diehn M, Loo BW, Xing L. Analysis of long-term 4-dimensional computed tomography regional ventilation after radiation therapy. Int J Radiat Oncol Biol Phys 2015;92:683–90.

[28] Yamamoto T, Kabus S, Lorenz C, Mittra E, Hong JC, Chung M, et al. Pulmonary ventilation imaging based on 4-dimensional computed tomography: comparison with pulmonary function tests and SPECT ventilation images. Int J Radiat Oncol Biol Phys 2014;90:414–22.

[29] Eslick EM, Bailey DL, Harris B, Kipritidis J, Stevens M, Li BT, et al. Measurement of preoperative lobar lung function with computed tomography ventilation imaging: progress towards rapid stratification of lung cancer lobectomy patients with abnormal lung function. Eur J Cardiothorac Surg 2015;49:1075–82.

[30] Vinogradskiy Y, Jackson M, Schubert L, Jones B, Castillo R, Castillo E, et al. Assessing the use of 4DCT-ventilation in pre-operative surgical lung cancer evaluation. Med Phys 2017;44:200–8.

[31] Fuld MK, Easley RB, Saba OI, Chon D, Reinhardt JM, Hoffman EA, et al. CT-measured regional specific volume change reflects regional ventilation in supine sheep. J Appl Physiol 2008;104:1177–84.

[32] Mathew L, Wheatley A, Castillo R, Castillo E, Rodrigues G, Guerrero T, et al. Hyperpolarized 3 He magnetic resonance imaging: comparison with four-dimensional x-ray computed tomography imaging in lung cancer. Acad Radiol 2012;19:1546–53.

[33] Kanai T, Kadoya N, Ito K, Kishi K, Dobashi S, Yamamoto T, et al. Evaluation of four-dimensional computed tomography (4D-CT)-based pulmonary ventilation: The high correlation between 4D-CT ventilation and 81m Kr-planar images was found. Radiother Oncol 2016;119:444–8.

[34] Tahir BA, Van Holsbeke C, Ireland RH, Swift AJ, Horn FC, Marshall H, et al. Comparison of CT-based lobar ventilation with 3He MR imaging ventilation measurements. Radiology 2016;278:585–92.

[35] Seppenwoolde Y, De Jaeger K, Boersma LJ, Belderbos JSA, Lebesque JV. Regional differences in lung radiosensitivity after radiotherapy for non-small-cell lung cancer. Int J Radiat Oncol Biol Phys 2004;60:748–58.

[36] Marks LB. Physiology-based studies of radiation-induced normal tissue injury. Radiother Oncol J Eur Soc Ther Radiol Oncol 1999;51:101.

[37] Roach PJ, Schembri GP, Bailey DL. V/q scanning using spect and spect/ct. J Nucl Med 2013;54:1588–96.

[38] Andreeff M, Wunderlich G. PET aerosol lung scintigraphy using Galligas. Eur J Nucl Med Mol Imaging 2010;37:175.

[39] Ebert M, Grossmann T, Heil W, Otten EW, Surkau R, Thelen M, et al. Nuclear magnetic resonance imaging with hyperpolarised helium-3. The Lancet 1996;347:1297–9.

[40] MacFall JR, Charles HC, Black RD, Middleton H, Swartz JC, Saam B, et al. Human lung air spaces: potential for MR imaging with hyperpolarized He-3. Radiology 1996;200:553–8.

[41] Mugler III JP, Driehuys B, Brookeman JR, Cates GD, Berr SS, Bryant RG, et al. MR imaging and spectroscopy using hyperpolarized 129Xe gas: preliminary human results. Magn Reson Med 1997;37:809–15.

[42] Albert MS, Cates GD, Driehuys B, Happer W, Saam B, Springer Jr CS, et al. Biological magnetic resonance imaging using laser-polarized 129Xe. Nature 1994;370:199.

[43] Fain SB, Panth SR, Evans MD, Wentland AL, Holmes JH, Korosec FR, et al. Early emphysematous changes in asymptomatic smokers: detection with 3He MR imaging. Radiology 2006;239:875–83.

[44] Kirby M, Mathew L, Wheatley A, Santyr GE, McCormack DG, Parraga G. Chronic obstructive pulmonary disease: longitudinal hyperpolarized 3He MR imaging. Radiology 2010;256:280–9.

[45] Tahir BA, Marshall H, Hughes PJC, Brightling CE, Collier G, Ireland RH, et al. Comparison of CT ventilation imaging and hyperpolarised gas MRI: effects of breathing manoeuvre. Phys Med Biol 2019.

[46] Sonke J-J, Lebesque J, Van Herk M. Variability of four-dimensional computed tomography patient models. Int J Radiat Oncol Biol Phys 2008;70:590–8.

[47] Yamamoto T, Langner U, Loo BW, Shen J, Keall PJ. Retrospective analysis of artifacts in four-dimensional CT images of 50 abdominal and thoracic radiotherapy patients. Int J Radiat Oncol Biol Phys 2008;72:1250–8.

[48] Faught AM, Yamamoto T, Castillo R, Castillo E, Zhang J, Miften M, et al. Evaluating which dose-function metrics are most critical for functional-guided radiotherapy with CT ventilation imaging. Int J Radiat Oncol Biol Phys 2017.

[49] Wang JM, Robertson SH, Wang Z, He M, Virgincar RS, Schrank GM, et al. Using hyperpolarized 129Xe MRI to quantify regional gas transfer in idiopathic pulmonary fibrosis. Thorax 2018;73:21–8.

[50] Waxweiler T, Schubert L, Diot Q, Faught A, Stuhr K, Castillo R, et al. A complete 4DCT-ventilation functional avoidance virtual trial: Developing strategies for prospective clinical trials. J Appl Clin Med Phys 2017;18:144–52.

[51] Kipritidis J. TU-H-202-04: The VAMPIRE Challenge: Preliminary Results From a Multi-Institutional Study of CT Ventilation Image Accuracy. Med Phys 2016;43:3771–3771.

Table

Table 1: PICOS Table for Study Question

Patient/participants	Human or animal subjects undergoing 4DCT or BHCT
Intervention	Generation of CTVI
Comparison	Accepted contrast-based imaging for regional lung function, including γ-scintigraphy, ventilation SPECT or PET, Hyperpolarized gas MRI, single or dual energy CT OR Accepted pulmonary function tests for global lung function, including spirometry and measurements for static lung function parameters
Outcome	Correlation of CTVI parameters (evaluated at the voxel, sub-organ or whole-organ level) versus clinical function (imaging or spirometry)
Study	Retrospective or prospective study quantified animal and human studies

Table

Table 2a Animal Studies and Table 2b Human Studies: Summary of CTVI studies. CTVI generation is defined as non-DIR and DIR based, which includes DIR Jacobian (DIR-Jac), DIR-HU and other hybrid DIR approaches (DIR-Hy). Papers used various metrics for assessing the strength of correlation, including linear regression (LR), Spearman Rank correlations (SPC) and Dice Similarity Coefficients (DSC).

Abbreviations: Coefficient of Variance (CoV), forced expiratory volume in 1 second (FEV1), FVC forced vital capacity, Δ Vol ventilation calculation of ventilation based on change in volume, Region of Interest (ROI).

Table 2a: Animal Studies of CT Ventilation

Study author (year)	Subjects No.	Comparative Imaging Modality	CTVI Generation	Type of correlation	Highest level of correlation	Range of correlation
Fuld et al (2008)**[31]**	4 sheep	Xenon-CT	Change in volume as assessed by change in HU units	Voxel based LR	LR *R*2 of 0.76	0.56–0.76
Reinhardt et al (2008)**[22]**	5 sheep	Xenon-CT	DIR-Jac	Voxel Based LR	LR *R*2 of 0.80	0.64–0.80
Ding et al 2012 **[23]**	4 sheep	Xenon-CT	2 variants of DIR-Jac (SAJ, SACJ) and DIR-HU	Voxel based LR	SAJ *r=0.97* SACJ *r=0.994* DIR-HU *r=0.952*	0.836-0.97 0.888-0.994 0.893-0.952
Zhang et al (2016)**[24]**	4 sheep	Xenon CT	3 metrics tested: 1. DIR Δ Vol 2.DIR-Jac 3. DIR- HU	Voxel based SPC	Δ Vol 0.61 Jacobian 0.61 HU 0.42	0.29-0.61 0.31-0.61 0.17–0.42

Table 2b: Human studies of CT ventilation

Study author (year)	Subjects No.	Comparative Imaging Modality	CTVI Generation	Type of Correlation	Highest average level of correlatio	Range

					n	
Guerrero 2005**[2]**	22	Measured tidal volume on CT	DIR-HU	LR of lung volumes and tidal volumes.	DIR HU LR *r*= 0.985	LR *r* ranged from 0.982 -0.985
Guerrero 2006**[1]**	3 lung cancer RT patients	Measured tidal volume on CT	DIR-HU	Whole lung based	DIR-HU LR *r* = 0.985	NR
Castillo et al 2010**[19]**	7 thoracic oncology patients	DTPA-SPECT	DIR- HU DIR-Jac	Voxel based DSC	Highest average DSC was for 0-20% voxels	DSC ranged from 0.2-0.35
Murphy et al 2012**[25]**	216 patients with COPD	Spirometry	DIR-HU	Whole lung and lobar assessment LR of GOLD stage FEV1 and FEV1/FVC	Median *r* value of 0.87 for whole lung	0.85-0.91
Matthew et al 2012**[32]**	11 lung cancer patients	^{3}HE hyperpolarized MRI	DIR- HU	DSC to compare ventilated volume (VV) in both whole lung and lung ipsilateral to and contralateral to cancer	DSC 0.89±.01	0.69-0.95
Choi et al 2013**[18]**	30 asthma patients, 14 control subjects	PFT’s comparison of Total Lung Capacity (TLC) and Air volume (AV) at exhale.	DIR-Jac derived from breath-hold CT	Global lung function LR	*R=0.87* for Total lung volume in severe asthmatics	*r* =0.78 to 0.87
Yamamoto et al 2013**[20]**	9 patients with thoracic cancer.	DTPA-SPECT	DIR- Jac	Voxel based SPC, DSC for segmented low-functional lung regions	Best Spearman rank 0.80 Best DSC 0.8	Average 0.69±0.26 0.71±
Kipritidis et al 2014**[6]**	12 lung cancer patients	PET-Galligas	DIR-HU and DIR-Jac with and without density	Voxel based SPC	Density-scaled HU Spearman $\bar{r}$ = 0.28 ± 0.13 and DSC	DIR-Jac 0.25±0.17

			scaling		(lowest 20%) = 0.52 ± 0.09	
Yamamoto et al 2014**[5]**	18 patients- all with thoracic cancer	DTPA-SPECT and PFT's (spirometry and measurement of DLCO)	DIR- HU DIR-Jac .	Voxel based DSC to quantify overlap betweenV4DCT and VSPECT defect regions Pearson correlation with FEV1 and FEV1/FVC	DSC for DIR- HU 0.39	Average DSCs were: DIR-HU 0.39 ± 0.11 DIR-Jac 0.36 ±0.13 FEV1/FVC strongly correlated with 25% voxel value (0.73) and strongly negatively correlated with defective lung (-0.63).
Brennan et al 2015**[26]**	98 patients with lung cancer	Spirometry	DIR- HU and DIR-Jac	LR to compare spirometry and CTVI CoV V20, and visually defined defects.	Correlation coefficient ~ 0.7 for HU	DIR-HU CoV CC between: DIR-HU FEV1 0.72 FEV1/FVC 0.67 DIR-Jac FEV1 0.40 FEV1/FVC 0.38
King et al. 2015**[27]**	30 Thoracic radiotherapy patients	Tidal volume from 4DCT and CTVI compared	DIR-Jac DIR-Hy	Whole lung based Pearson correlation coefficient	TVInt and TVCT was 0.92 (P<.01) TVJac and TVCT was 0.97 (P<.01).	Not reported
Eslick et al 2015**[29]**	11 lung cancer patients	PET Galligas	DIR- HU	Lobar volumes and ventilation compared with LR	*r*-value 0.96 for comparison of lobar ventilation from CTVI and	Average *r*-value between CTVI analysed lobar ventilation vs. lobar volumes was 0.78

					PET-Galligas.	
Kipritidis et al 2016**[21]**	25 lung cancer patients	PET-Galligas	Non-DIR based HU DIR-HU DIR-Jac	Voxel based SPC	CTVI-HU 0.50	Mean ±SD correlation with Galligas PET was r = (0.50±0.17), (0.42±0.20), and (0.19±0.23) for the CTVI-HU, DIR-HU, and DIR-Jac methods
Kida et al 2016**[4]**	8 thoracic cancer patients	DTPA SPECT and Spirometry	DIR-HU and DIR-Jac	Radiotherapy lung metrics compared with Pearson correlation and LR	0.94 for DIR-HU	DIR-HU Pearson R=0.94 and linear regression = 0.71 DIR-Jac R=0.85; slope=0.5
Kanai et al 2016**[33]**	11 lung cancer patients	Planar Kr images	DIR-HU and DIR-Jac	Voxel based SPC	HU 0.875	Mean ± SD: DIR-HU 0.875±0.07 DIR-Jac 0.803±0.114
Tahir et al 2016**[34]**	30 patients with sputum eosinophilia and asthma	Hyperpolarized ^{3}He MRI	Breath-hold CT at total lung capacity and functional residual capacity used to assess change in volume of lobes	Lobar Pearson correlation of all lobar regions	0.65	Range of Pearson correlations not stated
Vinogradskiy et al 2017**[30]**	16 lung cancer patients	DTPA SPECT and spirometry	DIR- HU	Global lung function ROC analysis to compare, 4DCT-ventilation-based preop FEV1 vs. SPECT based preop FEV1.	0.99 correlation coefficient for prediction of ventilation changes after	Pneumonectomy: Correlation coefficient .80(0.81for nuclear medicine-ventilation And 0.78 for nuclear medicine-perfusion). Lobectomy:

					lobectomy using CTVI	Correlation coefficient was 0.99 for CTVI.
Hegi-Johnson et al 2017**[10]**	11 lung cancer patients	Technegas SPECT	CTVI-HU, DIR-HU and DIR-Jac	Voxel based Spearman Rank and DSC, lobar based Pearson Correlation and whole lung CoV	Non-defect regions: CTVI HU, DIR-HU and DIR-Jac mean DSC of 0.69,0.68, and 0.54.	Defect regions mean DSC were 0.39, 0.33, and 0.44. Spearman r :0.26, 0.18 and -0.02 for CTVIHU, DIR-HU and DIR-Jac respectively
Tahir et al 2018**[8]**	11 lung cancer patients	^{129}Xe and ^{3}He MRI	DIR-HU, DIR-Jac, Specific gas volume change	Spearman rank correlations of different ROI sizes	DIR-HU SPC *R*= 0.37, DIR-Jac SPC 0.31, Specific gas volume 0.34	Voxel-level: 0.1-0.8 ROI 20x20 voxels: *R=0.2-0.9*

~~Table 2a Animal Studies and Table 2b Human Studies: Summary of CTVI studies. CTVI generation is defined as non-DIR and DIR based, which includes DIR Jacobian (DIR-Jac), DIR-HU and other hybrid DIR approaches (DIR-Hy). Papers used various metrics for assessing the strength of correlation, including linear regression (LR), Spearman Rank correlations (SPC) and Dice Similarity Coefficients (DSC).~~

~~Abbreviations: Coefficient of Variance (CoV), forced expiratory volume in 1 second (FEV1), FVC forced vital capacity, Δ Vol ventilation calculation of ventilation based on change in volume.~~

~~Table 2a: Animal Studies of CT Ventilation~~

~~Study author (year)~~	~~Subjects No.~~	~~Comparative Imaging Modality~~	~~CTVI Generation~~	~~Type of correlation~~	~~Highest level of correlation~~	~~Range of correlation~~
~~Fuld et al (2008)~~**~~[31]~~**	~~4 sheep~~	~~Xenon CT~~	~~Change in volume as assessed by change in HU units~~	~~Voxel based LR~~	~~LR *R2* of 0.76~~	~~0.56–0.76~~
~~Reinhardt et al (2008)~~**~~[22]~~**	~~5 sheep~~	~~Xenon CT~~	~~DIR-Jac~~	~~Voxel Based LR~~	~~LR *R2* of 0.80~~	~~0.64–0.80~~
~~Ding et al~~	~~4 sheep~~	~~Xenon CT~~	~~2 variants of DIR-Jac~~	~~Voxel based LR~~	~~SAJ *r=0.97*~~	~~0.836-0.97~~

~~2012 **[23]**~~			~~(SAJ, SACJ) and DIR-HU~~		~~SACJ *r=0.994*~~ ~~DIR-HU *r=0.952*~~	~~0.888-0.994~~ ~~0.893-0.952~~
~~Zhang et al (2016)**[24]**~~	~~4 sheep~~	~~Xenon CT~~	~~3 metrics tested:~~ ~~1. DIR Δ Vol~~ ~~2.DIR-Jac~~ ~~3. DIR- HU~~	~~Voxel based SPC~~	~~Δ Vol 0.61~~ ~~Jacobian 0.61~~ ~~HU 0.42~~	~~0.29-0.61~~ ~~0.31-0.61~~ ~~0.17- 0.42~~

~~Table 2b: Human studies of CT ventilation~~

~~Study author (year)~~	~~Subjects No.~~	~~Comparative Imaging Modality~~	~~CTVI Generation~~	~~Type of Correlation~~	~~Highest average level of correlation~~	~~Range~~
~~Guerrero 2005**[2]**~~	~~22~~	~~Measured tidal volume on CT~~	~~DIR-HU~~	~~LR of lung volumes and tidal volumes.~~	~~DIR-HU LR *r*= 0.985~~	~~LR *r* ranged from 0.982 -0.985~~
~~Guerrero 2006**[1]**~~	~~3 lung cancer RT patients~~	~~Measured tidal volume on CT~~	~~DIR-HU~~	~~Whole lung based~~	~~DIR-HU LR *r* = 0.985~~	~~NR~~
~~Castillo et al 2010**[19]**~~	~~7 thoracic oncology patients~~	~~DTPA-SPECT~~	~~DIR- HU~~ ~~DIR-Jac~~	~~Voxel based DSC~~	~~Highest average DSC was for 0-20% voxels~~	~~DSC ranged from 0.2-0.35~~
~~Murphy et al 2012**[25]**~~	~~216 patients with COPD~~	~~Spirometry~~	~~DIR-HU~~	~~Whole lung and lobar assessment LR of GOLD stage FEV1 and FEV1/FVC~~	~~Median *r* value of 0.87 for whole lung~~	~~0.85-0.91~~
~~Matthew et al 2012**[32]**~~	~~11 lung cancer patients~~	~~^{3}HE hyperpolarized MRI~~	~~DIR- HU~~	~~DSC to compare ventilated volume (VV) in both whole lung and lung ipsilateral to~~	~~DSC 0.89±.01~~	~~0.69-0.95~~

				~~and contralateral to cancer~~		
~~Choi et al 2013~~**~~[18]~~**	~~30 asthma patients, 14 control subjects~~	~~PFT's comparison of Total Lung Capacity (TLC) and Air volume (AV) at exhale.~~	~~DIR-Jac derived from breath-hold CT~~	~~Global lung function LR~~	~~*R=0.87* for Total lung volume in severe asthmatics~~	~~*r* =0.78 to 0.87~~
~~Yamamoto et al 2013~~**~~[20]~~**	~~9 patients with thoracic cancer.~~	~~DTPA-SPECT~~	~~DIR–Jac~~	~~Voxel based SPC, DSC for segmented low-functional lung regions~~	~~Best Spearman rank 0.80~~ ~~Best DSC 0.8~~	~~0.1-0.8~~ ~~0.4-0.8.~~
~~Kipritidis et al 2014~~**~~[6]~~**	~~12 lung cancer patients~~	~~PET-Galligas~~	~~DIR-HU and DIR-Jac with and without density scaling~~	~~Voxel based SPC~~	~~Density-scaled HU Spearman~~ $\bar{r}$ ~~= 0.50 ± 0.17 and DSC (lowest 20%) = 0.52 ± 0.09~~	~~DIR-Jac 0.19±0.23~~ ~~DIR-HU 0.50 ± 0.17~~
~~Yamamoto et al 2014~~**~~[5]~~**	~~18 patients-all with thoracic cancer~~	~~DTPA-SPECT and PFT's (spirometry and measurement of DLCO)~~	~~DIR–HU~~ ~~DIR-Jac~~ ~~.~~	~~Voxel based DSC to quantify overlap betweenV4D CT and VSPECT defect regions Pearson correlation with FEV1 and FEV1/FVC~~	~~DSC for DIR–HU 0.39~~	~~Average DSCs were:~~ ~~DIR-HU 0.39 ± 0.11~~ ~~DIR-Jac 0.36 ±0.13~~ ~~FEV1/FVC strongly correlated with 25% voxel value (0.73) and strongly negatively correlated with defective lung (-0.63).~~
~~Brennan et al 2015~~**~~[26]~~**	~~98 patients with lung cancer~~	~~Spirometry~~	~~DIR–HU and DIR-Jac~~	~~LR to compare spirometry and CTVI CoV V20, and visually defined~~	~~Correlation coefficient ~ 0.7 for HU~~	~~DIR-HU CoV CC between:~~ ~~DIR-HU~~ ~~FEV1 0.72~~ ~~FEV1/FVC 0.67~~

				~~defects.~~		~~DIR-Jac FEV1 0.40 FEV1/FVC 0.38~~
~~King et al. 2015**[27]**~~	~~30 Thoracic radiotherapy patients~~	~~Tidal volume from 4DCT and CTVI compared~~	~~DIR-Jac DIR-Hy~~	~~Whole lung based Pearson correlation coefficient~~	~~TVInt and TVCT was 0.92 (P<.01)~~ ~~TVJac and TVCT was 0.97 (P<.01).~~	~~Not reported~~
~~Eslick et al 2015**[29]**~~	~~11 lung cancer patients~~	~~PET Galligas~~	~~DIR- HU~~	~~Lobar volumes and ventilation compared with LR~~	~~*r*-value 0.96 for comparison of lobar ventilation from CTVI and PET-Galligas.~~	~~Average *r*-value between CTVI analysed lobar ventilation vs. lobar volumes was 0.78~~
~~Kipritidis et al 2016**[21]**~~	~~25 lung cancer patients~~	~~PET-Galligas~~	~~Non-DIR based HU~~ ~~DIR-HU~~ ~~DIR-Jac~~	~~Voxel based SPC~~	~~CTVI-HU 0.50~~	~~Mean ±SD correlation with Galligas PET was r = (0.50±0.17), (0.42±0.20), and (0.19±0.23) for the CTVI-HU, DIR-HU, and DIR-Jac methods~~
~~Kida et al 2016**[4]**~~	~~8 thoracic cancer patients~~	~~DTPA SPECT and Spirometry~~	~~DIR-HU and DIR-Jac~~	~~Radiotherapy lung metrics compared with Pearson correlation and LR~~	~~0.94 for DIR-HU~~	~~DIR-HU Pearson R=0.94 and linear regression = 0.71~~ ~~DIR-Jac R=0.85; slope=0.5~~
~~Kanai et al 2016**[33]**~~	~~11 lung cancer patients~~	~~Planar Kr images~~	~~DIR-HU and DIR-Jac~~	~~Voxel based SPC~~	~~HU 0.875~~	~~Mean ± SD: DIR-HU 0.875±0.07 DIR-Jac~~

						~~0.803±0.114~~
~~Tahir et al 2016**[34]**~~	~~30 patients with sputum eosinophilia and asthma~~	~~Hyperpolarized ^{3}He MRI~~	~~Breath-hold CT at total lung capacity and functional residual capacity used to assess change in volume of lobes~~	~~Lobar Pearson correlation of all lobar regions~~	~~0.65~~	~~Range of Pearson correlations not stated~~
~~Vinogradskiy et al 2017**[30]**~~	~~16 lung cancer patients~~	~~DTPA SPECT and spirometry~~	~~DIR- HU~~	~~Global lung function ROC analysis to compare, 4DCT-ventilation-based preop FEV1 vs. SPECT based preop FEV1.~~	~~0.99 correlation coefficient for prediction of ventilation changes after lobectomy using CTVI~~	~~Pneumonectomy: Correlation coefficient .80(0.81for nuclear medicine-ventilation And 0.78 for nuclear medicine-perfusion). Lobectomy: Correlation coefficient was 0.99 for CTVI.~~
~~Hegi-Johnson et al 2017**[10]**~~	~~11 lung cancer patients~~	~~Technegas SPECT~~	~~CTVI-HU, DIR-HU and DIR-Jac~~	~~Voxel based Spearman Rank and DSC, lobar based Pearson Correlation and whole lung CoV~~	~~Non-defect regions: CTVI HU, DIR-HU and DIR-Jac mean DSC of 0.69,0.68, and 0.54.~~	~~Defect regions mean DSC were 0.39, 0.33, and 0.44. Spearman r :0.26, 0.18 and -0.02 for CTVIHU, DIR-HU and DIR-Jac respectively~~
~~Tahir et al 2018**[8]**~~	~~11 lung cancer patients~~	~~^{129}Xe and ^{3}He MRI~~	~~DIR-HU, DIR-Jac, Specific gas volume change~~	~~Spearman rank correlations of different ROI sizes~~	~~DIR-HU SPD *R*= 0.37, DIR-Jac SPD 0.31, Specific gas volume 0.34~~	~~Voxel level: 0.1-0.8~~ ~~ROI 20x20 voxels: *R=0.2-0.9*~~

Table

Table 3: Summary of Findings Table

Key Findings
Question: Can CTVI defined regional ventilation provide the same information as contrast-based ventilation imaging and pulmonary function tests?

Population: Patients having 4DCT or BHCT, the majority of whom are undergoing radiotherapy planning

Reference standard: contrast-based imaging of regional lung function and/or spirometry

Findings:

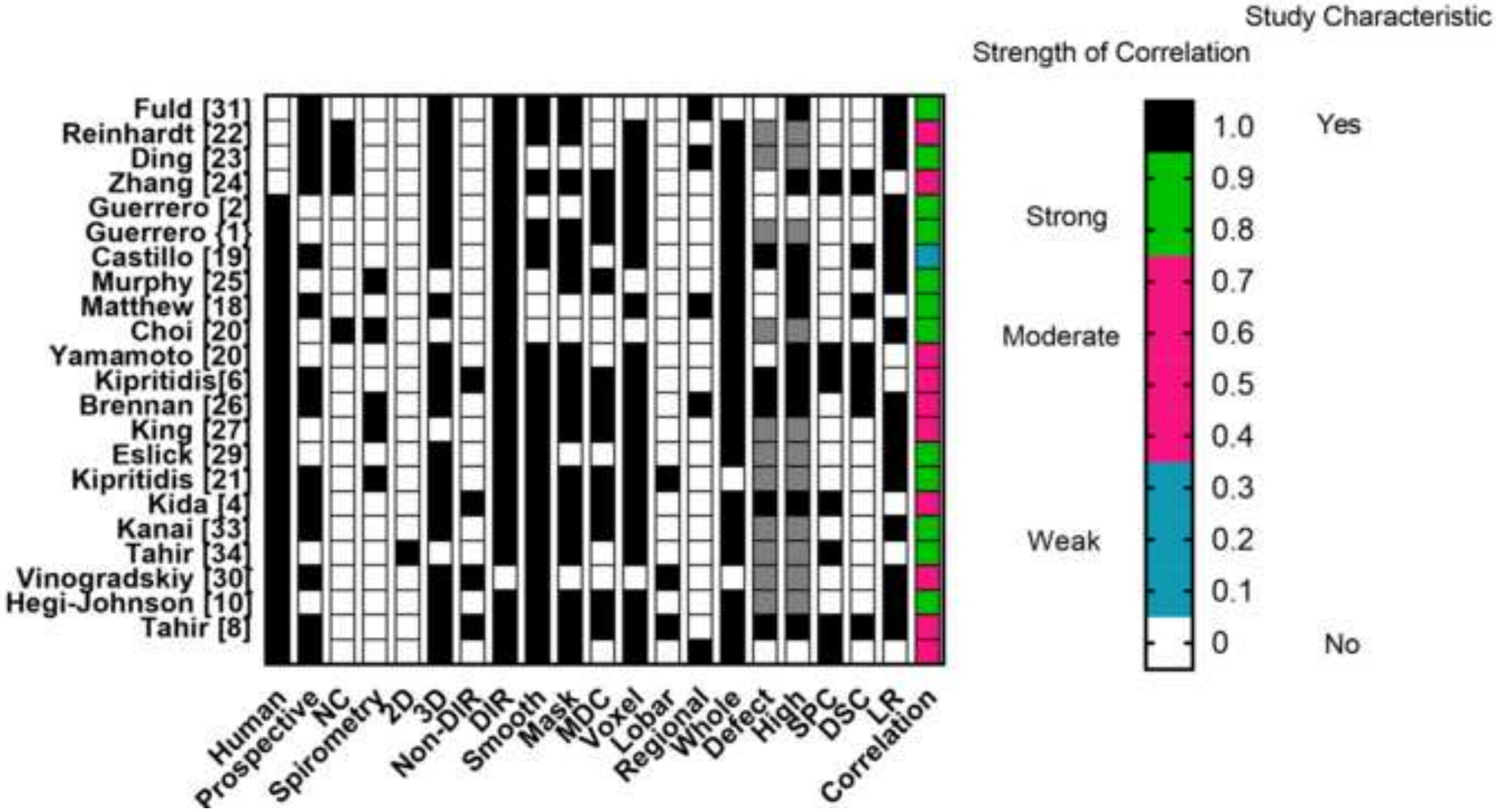
- LR (Average ± S.D.): 0.73±0.24 and 0.86± 0.11 for DIR-Jac and DIR-HU
- SPC (Average ± S.D.): 0.45±0.31and 0.41±0.11 for DIR-Jac and DIR-HU

Heterogeneity in reporting standards reduces the ability to assess statistical significance.

Recommendation		No of relevant studies	Strength
DIR vs. Non-DIR	There is insufficient evidence to recommend DIR over non-DIR approaches.	Only 4 non-DIR studies	Moderate
	If DIR based CTVI metrics are used the quality of the DIR should be assessed		Weak to moderate
Selection of metric	There is insufficient evidence to recommend one CTVI metric over another and further head to head testing is required.	19 DIR-HU 15 DIR-Jac	Moderate
CT acquisition and processing	CTVI based on both 4DCT and BHCT have shown moderate to strong correlations with clinical gold standards. There is insufficient evidence to guide the use of one method of acquisition above another.	4 BH studies	Moderate to strong
	There is insufficient evidence to support the use of AV biofeedback routinely in the acquisition of CT images: studies using AV biofeedback have similar strengths of validation compared to non-AV biofeedback studies.	2 AV biofeedback studies	Weak
	Post-processing such as smoothing and masking may influence the quality of the CTVI image, but there is insufficient evidence to recommend for or against their use.	18 papers used post-processing, but methodology was heterogeneous.	Weak
Clinical Validation	The majority of validation studies have used 3D-imaging and this should be	19 papers used 3D imaging	Moderate to strong

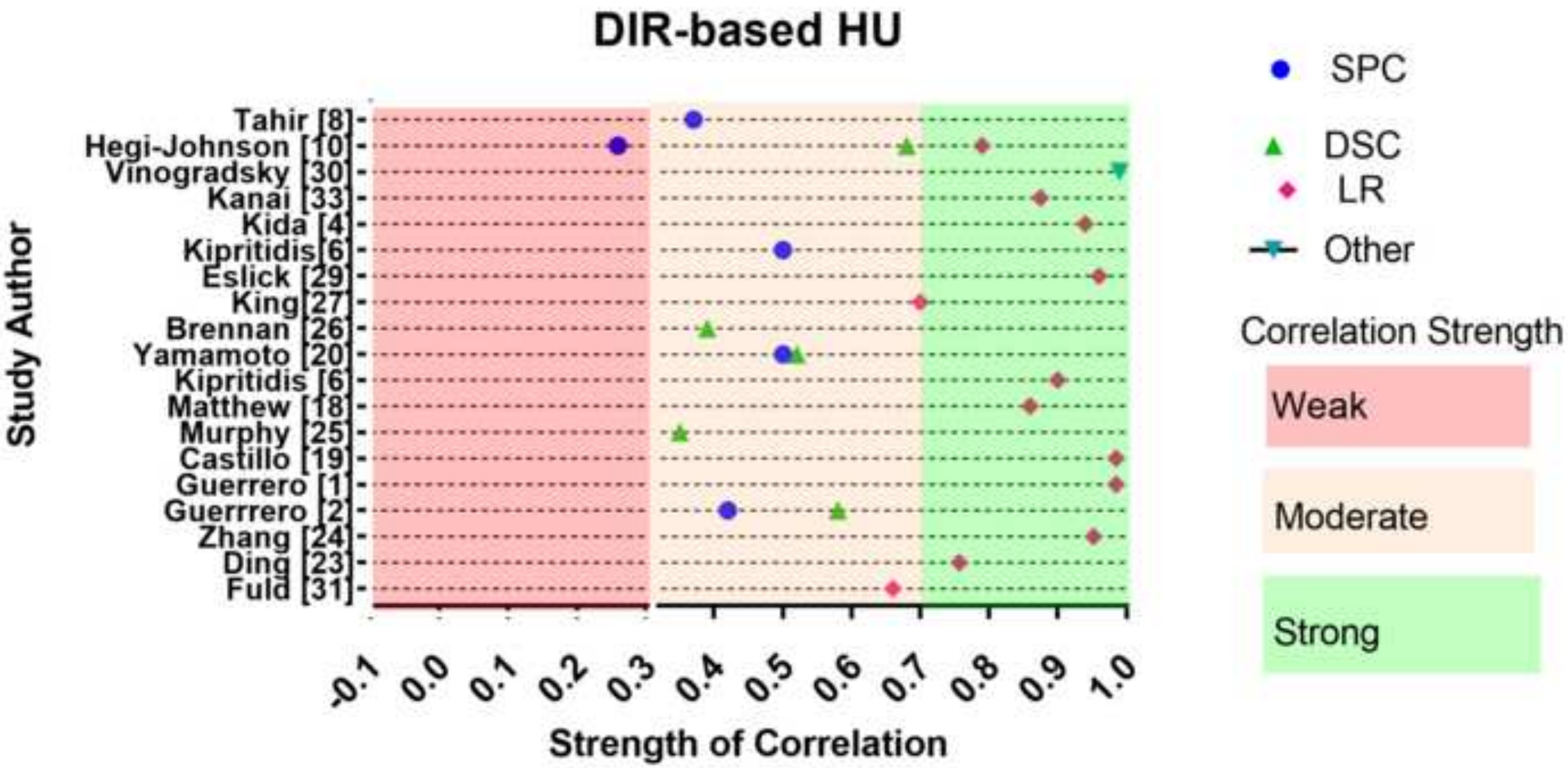
Modality	used for future validation studies for them to be benchmarked against existing methodologies. Standardized criteria for clinical assessment (e.g. GOLD criteria) should be used for validation modalities to increase the clinical relevance of future validation studies.	 5 papers reported validation against spirometry	 Moderate

Figure
Click here to download high resolution image



Figure

Click here to download high resolution image



Figure

Click here to download high resolution image

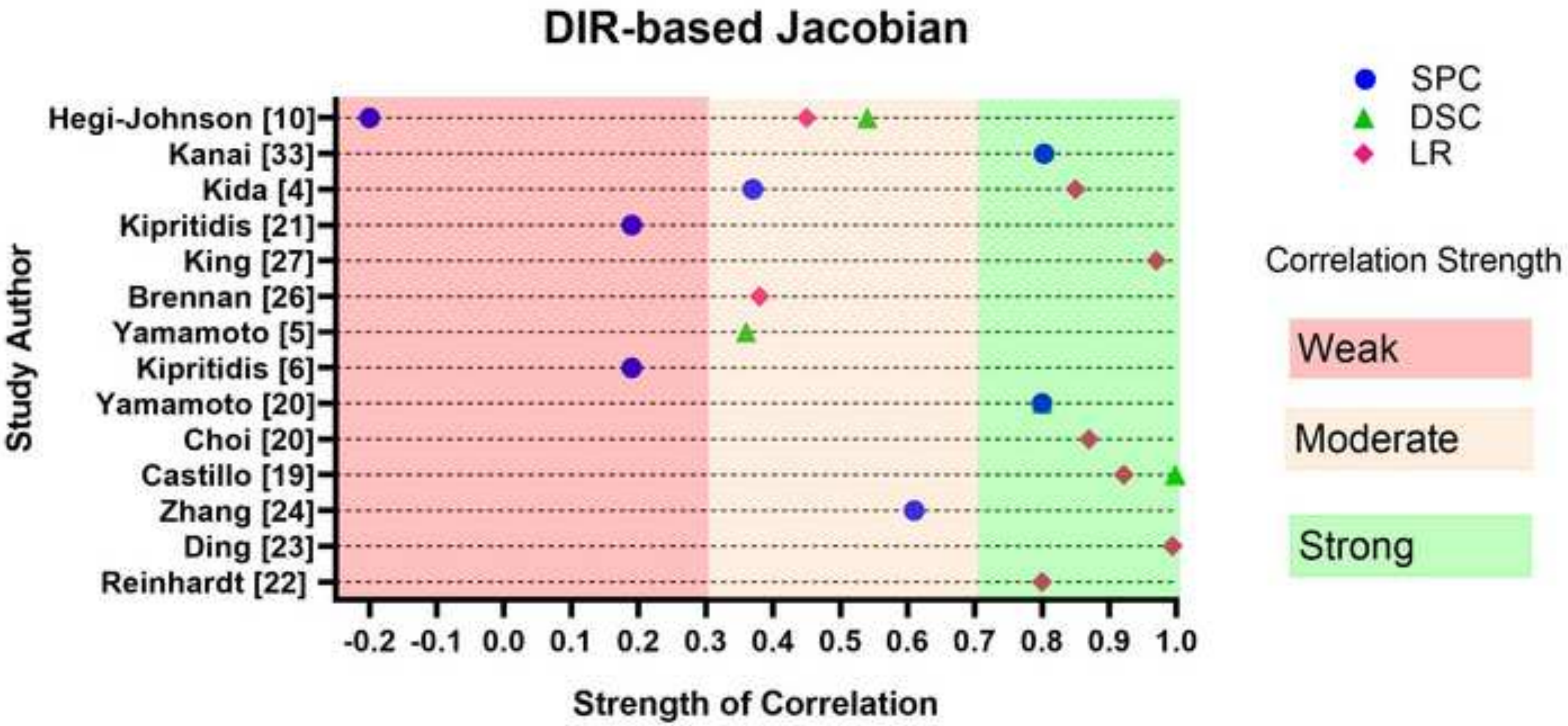
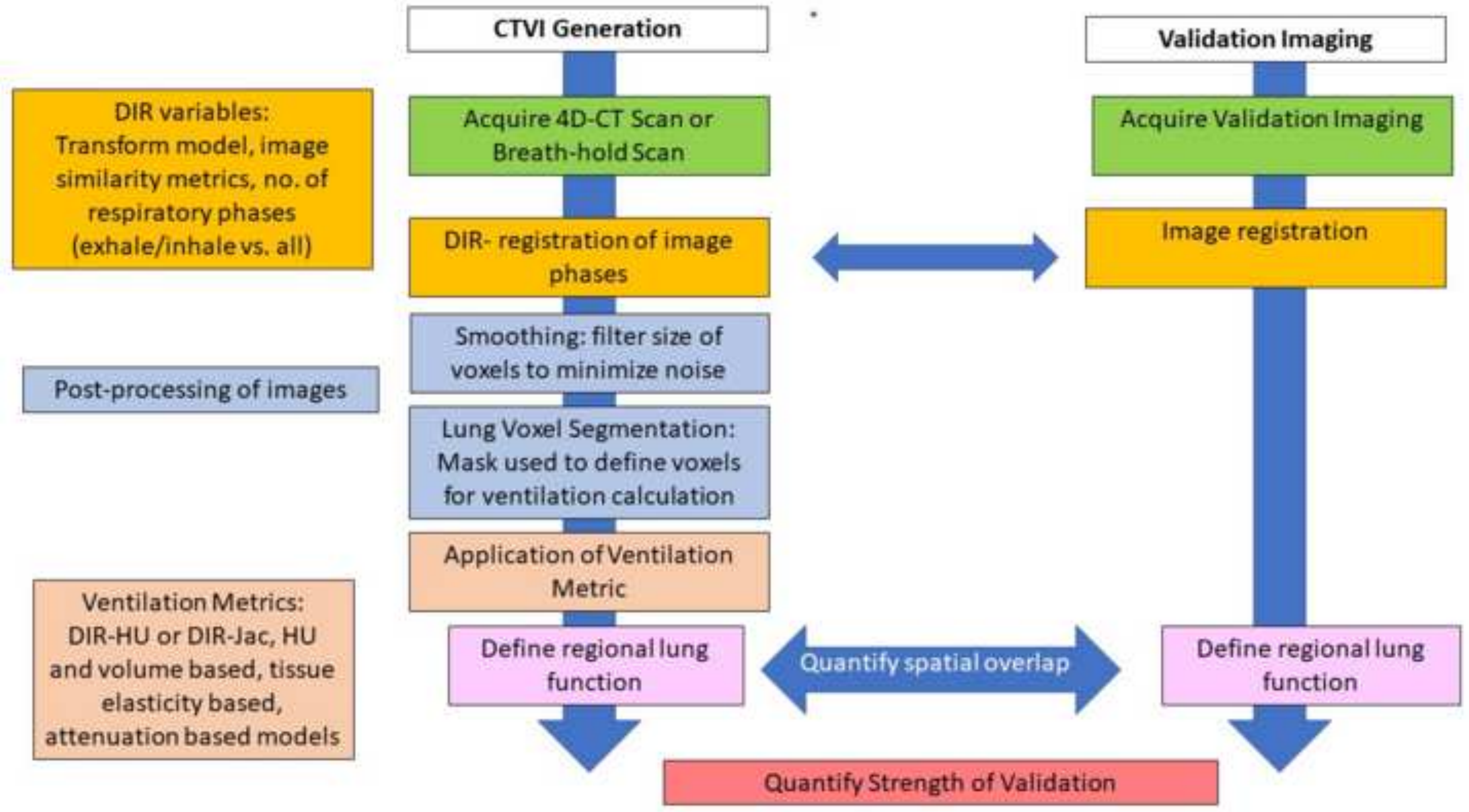


Figure
Click here to download high resolution image



Figure

Click here to download high resolution image

Figure 1: Characteristics of papers. This table provides an overview of study cohort and describes the gold standard correlation. Other details of the metrics such as post-processing and the methodology for comparison is described, whether this be voxel-based (Voxel)or based on a functional unit of lung (such as the lobe or whole lung) [Lobar, Regional and Whole], and whether respiratory defects (Defect) or high functioning lung (High) was evaluated. In this table the strongest average correlation reported in that paper is recorded for any type of correlation. Grey boxes indicate characteristics which are not applicable to these papers.

Abbreviations: Normal Comparator (NC), Deformable Image Registration (DIR), Dice Similarity Coefficients (DSC) and linear regression (LR) methods, smoothing (Smooth) or masking (Mask), mass density corrected metric (MDC), Spearman Rank Correlation (SPC), Dice Similarity Coefficient (DSC), Linear Regression (LR) are also described. .

Figure 2a and 2b: SPC, DSC and LR for DIR-HU (Fig. 3a) and DIR-Jacobian (Fig, 3B) respectively. Correlations are graded as either weak (0-0.3), moderate (<0.3-0.7) or strong (<0.7-1.0).

Figure 3: Overview of typical workflow for a DIR-based CTVI validation study including common steps for CTVI generation (left arm) and validation against contrast-based ventilation imaging (right arm). This demonstrates the multiple variables that should be considered when reporting and assessing CTVI validation studies. These variables include details of the DIR methodology (orange boxes), including the metrics used to perform DIR and the details of the respiratory phases included in the registration. Post-processing (blue boxes) may occur both before or after the CTVI image is generated (please see the text for details of this). Finally, the metrics that may be applied are diverse (brown boxes), although DIR-Jac and DIR-HU are the commonest. Common steps in the CTVI validation workflow are in the same colour as the CTVI generation pathway.

Figure 4: Forest plot illustrating Mean±Standard Error for Studies that present voxel-based SPC analyses for different ventilation metrics. Note the relatively small

number of studies that present similar results, and the wide error bars. The size of the icon is proportional to patient numbers.

Supplementary Files

Click here to download Supplementary Files: Supplementary Files.docx

***Conflict of Interest Statement**

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

Windows is either a registered trademark or a trademark of Microsoft Corporation in the United States and/or other countries. Mac is a trademark of Apple Inc., registered in the United States and other countries. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries.

***Conflict of Interest Statement**

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

Windows is either a registered trademark or a trademark of Microsoft Corporation in the United States and/or other countries. Mac is a trademark of Apple Inc., registered in the United States and other countries. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries.

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

Windows is either a registered trademark or a trademark of Microsoft Corporation in the United States and/or other countries. Mac is a trademark of Apple Inc., registered in the United States and other countries. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries.

***Conflict of Interest Statement**

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

Windows is either a registered trademark or a trademark of Microsoft Corporation in the United States and/or other countries. Mac is a trademark of Apple Inc., registered in the United States and other countries. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries.

***Conflict of Interest Statement**

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

Windows is either a registered trademark or a trademark of Microsoft Corporation in the United States and/or other countries. Mac is a trademark of Apple Inc., registered in the United States and other countries. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries.

***Conflict of Interest Statement**

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

Windows is either a registered trademark or a trademark of Microsoft Corporation in the United States and/or other countries. Mac is a trademark of Apple Inc., registered in the United States and other countries. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries.

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

Windows is either a registered trademark or a trademark of Microsoft Corporation in the United States and/or other countries. Mac is a trademark of Apple Inc., registered in the United States and other countries. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries.

***Conflict of Interest Statement**

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

Windows is either a registered trademark or a trademark of Microsoft Corporation in the United States and/or other countries. Mac is a trademark of Apple Inc., registered in the United States and other countries. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries.

Highlights

- We reviewed papers up to 2018 to see if CTVI can image regional ventilation.
- Voxel-based correlations were moderate(~0.4), and strong for larger volumes (~0.8).
- CTVI images obstructive COPD well, but may miss restrictive/gas diffusion issues.
- CTVI validation methodologies need standardization to help clinical implementation.

*Manuscript
Click here to view linked References

Imaging of regional ventilation: Is CT Ventilation Imaging the answer?

A Systematic Review of the Validation Data

Fiona Hegi-Johnson[1,2,3,4]; Dirk de Ruysscher[5,6]: Paul Keall[3]; Lizza Hendriks[7]; Yevgeniy Vinogradskiy[8]; Tokihiro Yamamoto[9]; Bilal Tahir[10.11]; John Kipritidis[3,12]

1. Sir Peter MacCallum Department of Oncology, University of Melbourne
2. Department of Radiation Oncology, Peter MacCallum Cancer Centre, 305 Grattan St, Melbourne, Australia 3000
3. ACRF Image X Institute, University of Sydney, Camperdown, NSW, Australia
4. School of Mathematical and Physical Sciences, The University of Newcastle (UON), University Drive, Callaghan NSW 2308
5. Maastricht University Medical Center, Dept. Radiation Oncology (Maastro Clinic), GROW, The Netherlands, Dr. Tanslaan 12, 6229 Maastricht, The Netherlands
6. KU Leuven, Radiation Oncology, Herestraat 49, 3000 Leuven, Belgium
7. Department of Pulmonary Diseases GROW - School for oncology and developmental biology, Maastricht University Medical Center+, Maastricht, the Netherlands
8. Department of Radiation Oncology, University of Colorado, Denver, Colorado USA
9. Radiation Oncology, UC Davis School Medicine, Sacramento, California, USA
10. Academic Unit of Clinical Oncology, University of Sheffield, United Kingdom
11. POLARIS, Academic Radiology, University of Sheffield, United Kingdom
12. Northern Sydney Cancer Centre, St Leonards, NSW Australia

Corresponding author:

Dr Fiona Hegi-Johnson

Peter MacCallum Cancer Centre

305 Grattan St Melbourne

Vic 3000 AUSTRALIA

Fiona.Hegi-Johnson@petermac.org

Abstract

Computed Tomography Ventilation Imaging (CTVI) is an experimental imaging modality that derives regional lung function information from non-contrast respiratory-correlated CT datasets. Despite CTVI being extensively studied in cross-modality imaging comparisons, there is a lack of consensus on the state of its clinical validation in humans. This systematic review evaluates the CTVI clinical validation studies to date, highlights their common strengths and weaknesses and makes recommendations. We performed a PUBMED and EMBASE search of all English language papers on CTVI between 2000 and 2018. The results of these searches were filtered in accordance to a set of eligibility criteria and analysed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Guidelines. One hundred and forty-four records were identified, and 66 full text records were reviewed. After detailed assessment, twenty-three full text papers met the selection criteria and were included in the final review. This included thirteen prospective studies, with 579 human subjects. Studies used diverse methodologies, with a large amount of heterogeneity between different studies in terms of the reference ventilation imaging modality (e.g. nuclear medicine, hyperpolarised gas MRI), imaging parameters, DIR algorithm(s) used, and ventilation metric(s) applied. The most common ventilation metrics used deformable image registration to evaluate the exhale-to-inhale motion field Jacobian determinant (DIR-Jac) or changes in air volume content based on Hounsfield Units (DIR-HU). The strength of correlation between CTVI and the reference ventilation imaging modalities was moderate to strong when evaluated at the lobar or global level, with the average ± S.D.(*number of studies*) linear regression correlation coefficients were 0.73±0.25 ($n= 6$)and 0.86± 0.11($n=12$) for DIR-Jac and DIR-HU respectively, and the SPC were 0.45±0.31($n=6$) and 0.41±0.11($n=5$) for DIR-Jac and DIR-HU respectively. We concluded that it is difficult to make a broad statement about the validity of CTVI due to the diverse methods used in the validation literature. Typically, CTVI appears to show reasonable cross-modality correlations at the lobar/whole lung level but poor correlations at the voxel level. Since CTVI is seeing new implementations in prospective trials, it is clear that refinement and standardization of the clinical validation methodologies are required. CTVI appears to be of relevance in radiotherapy planning, particularly in patients whose main pulmonary impairment is not a gas exchange problem but alternative imaging approaches may need to be considered in patients with other pulmonary diseases (i.e. restrictive or gas exchange problems).

Introduction:

The idea of deriving information about regional pulmonary function from respiratory-correlated computed tomography (CT), especially 4-Dimensional Computed Tomography (4DCT) and inhale/exhale breath-hold CT (BHCT), without exogenous contrast is highly attractive. In the context of radiotherapy treatment planning, respiratory-correlated thoracic CT scans are acquired routinely for lung cancer patients, a population with significant impairment of respiratory function, and breast cancer patients, where radiation-induced lung toxicity remains a major dose-limiting factor. CT Ventilation Imaging (CTVI) is a method for visualizing regional air volume changes in the lung[1],[2] combining 4DCT or BHCT scans with deformable image registration (DIR) to visualise the breathing-induced change in air volume, or “ventilation,” an important component of blood-gas exchange. CTVI is currently the subject of a number of clinical trials, which are integrating CTVI data into radiotherapy planning[3],[4] with the goal of minimising irradiation of functional lung and potentially minimising pulmonary toxicity. Attempts have been made to validate CTVI against a wide range of clinical and experimental ventilation imaging modalities including ^{99m}Tc-labeled diethylenetriamine pentacetate (DTPA) V-SPECT[4],[5], ^{68}Ga (Galligas) PET[6], ^{3}He MRI[7], ^{129}Xe MRI[8], 81m-Kr[9] and Technegas V-SPECT[10].

Almost all CTVI methods involve the application of DIR between the 4DCT or BHCT exhale and inhale phase images, with the DIR motion field then used to compute breathing-induced ventilation “metrics” at the voxel level. These are mainly based on regional lung volume changes as quantified by the DIR motion field Jacobian determinant (“DIR-Jac” methods), or evaluation of air volume changes as indicated by changes in the CT number or Hounsfield Units (“DIR-HU” methods). There are many sources of variation for studies comparing CTVI to other lung function imaging including: the CT acquisition protocol and breathing manoeuvre[11], the type of DIR method used for evaluating lung motion, the type of ventilation metric employed, the presence (or not) of image pre/post processing, and the choice of metrics used to evaluate the cross-modality correlation. Some of the most salient findings are that the 4DCT or BHCT image quality can significantly impact on CTVI generation[10]; DIR based metrics in particular, are highly sensitive to image artefacts, which may impair the ability to generate accurate CTVI images in the presence of 4DCT motion artefacts due to irregular breathing. There is also heterogeneity in the methods used to define “high function” or “low function” lung; some studies apply semi-automated thresholding approaches, whereas others perform a subjective clinical assessment of the image. The use of different types of “reference” ventilation imaging modalities, such as SPECT, PET, hyperpolarised gas MRI and Xenon-CT, introduces an additional complexity in that all of these imaging modalities operate on different (if complimentary) contrast mechanisms. Similarly, the various published CTVI metrics (mainly, dealing with lung volume or density change), are all related yet clearly distinct.

Although there have been 2 recent reviews of the literature of functional lung imaging in thoracic radiotherapy, these have focused more broadly on the application of different imaging types in clinical practice by looking at the integration of functional imaging into radiotherapy planning and the benefit of reducing the dose to normal lung[12],[13]. Our paper focuses on the technical details of the CTVI validation methodology; an understanding of this is crucial to define the utility and limitations of different approaches in assessing different kinds of pulmonary pathology, and standardization of these technical details is essential if we are to move forward and validate CTVI in clinical trials. The CTVI literature concentrates on the assessment of regional ventilation, but we also review alternative imaging approaches and discuss whether CTVI is the most appropriate modality for imaging pulmonary physiology in thoracic radiotherapy patients with pulmonary disease other than obstructive diseases, such as pulmonary vascular and interstitial lung disease.

Hence our purpose is twofold: 1) to summarize, and assess the quality of the validation literature for CTVI using the methodology of a systematic review and 2) to compare alternative imaging modalities for assessing regional pulmonary pathology, which may provide guidance as to the use of appropriate imaging for future studies in thoracic radiotherapy patients.

Methods:

The systematic review of the CTVI validation literature review was performed following the preferred reporting items for systematic reviews and meta-analyses (PRISMA)- statement reporting standard[14]. This consensus statement defines the process and items deemed essential for transparent reporting of a systematic review. Table 1 presents our research questions in the patients, intervention, comparison, outcome, study design (PICOS) approach.

Search Strategy

Between the 5th and 7th of September 2017 searches were performed on PUBMED and EMBASE using the search term “CT Ventilation”. Further studies were identified by handsearching of references and identification of studies that could possibly meet the selection criteria, as well as by direct input from the authors of the study.

These references were exported to the Systematic Review Data Repository (SRDR), an online and freely available resource provided by the US National Institute of Health (NIH) for the management of data in systematic reviews, which is available at www.srdr.ahrq.gov. The SRDR software was used to exclude duplicate studies and to assess if studies met selection criteria.

Table 1: PICOS Table for Study Question

Patient/participants	Human or animal subjects undergoing 4DCT or BHCT
Intervention	Generation of CTVI
Comparison	Accepted contrast-based imaging for regional lung function, including γ-scintigraphy, ventilation SPECT or PET, Hyperpolarized gas MRI, single or dual energy CT OR Accepted pulmonary function tests for global lung function, including spirometry and measurements for static lung function parameters
Outcome	Correlation of CTVI parameters (evaluated at the voxel, sub-organ or whole-organ level) versus clinical function (imaging or spirometry)
Study	Retrospective or prospective study quantified animal and human studies

Paper Selection and Data Extraction

Two reviewers (FH and JK) reviewed the papers independently to assess if they met study selection criteria. Study details were collected and are available in the Systematic Review Data Repository (www.srdr.ahrq.gov). If there was discordance in the assessment a third reviewer (DR) reviewed the papers. Paper quality was assessed by using the STARD Quality Dimensions for Diagnostic Tests by FH and JK[15] and the QUADAS-2[16]. The process of study selection and details of studies excluded at each step are outlined in Figure 1.

The following data were extracted from each paper:

1. Type of CT protocol (4DCT or breath hold)
2. Type of comparative or “reference” ventilation imaging modality (Ventilation PET or SPECT, Xenon CT or Hyperpolarized gas MRI)

3. Use of breathing guidance for each scan, for example audiovisual (AV) biofeedback or ventilation under anaesthesia in animal studies)
4. Details of comparison metric used (Spearman correlations, Dice similarity coefficient, or linear correlation of CoV / lobar function values etc.)
5. Details of CTVI algorithms by: DIR type (if any), Whether masking was used, What functional quantification metric(s) were used (e.g. DIR-Jac, DIR-HU, other hybrid methods that combine the Jacobian and HU changes, DIR-Hy, as well as the use of various model-based scaling factors as described in the Results section).
6. Image smoothing/filtering used at any stage in the process (for example, pre-smoothing of the input 4DCT phase images, or application of a box filter to pixel values in the output CTVI).
7. Details of the DIR assessment (if any), based on the techniques recommended by the report of the AAPM Task Group 132. This includes visual inspection of deformed images and/or motion fields, evaluation of target registration error (TRE) using expert selected anatomic landmarks, or determination of the presence of any negative values of the DIR motion field Jacobian determinant, which indicates non-physical motion.

Selection Criteria

Studies were accepted if they:

(1) Quantitatively correlated CTVI against an accepted clinical reference for measuring clinical function (either clinical/experimental imaging or spirometry)

(2) Generated CTVI from either 4DCT or BHCT without the use of a radioactive, iodinated or other imaging contrast other than air.

(3) Reported in the English language.

(4) Published in a peer-reviewed journal between the years 2000–2018. The start date was chosen as the year 2000 since the review by Simon et al. is often taken as an originating paper for the DIR-HU formulation[17].

(5) Intra-patient imaging/spirometry measurements were acquired within a reasonable timeframe (e.g. <3 months) without pulmonary intervention (namely, surgical resection or radiation therapy).

(6) Report detailed methodology for generation of CTVI images for example CT post-processing, image registration methodology, and/or other relevant algorithm/acquisition parameters.

Human and animal studies were both acceptable.

Even if meeting the above criteria, studies were rejected which:

(1) Did not have an intrapatient comparison to compare the standard functional imaging or accepted pulmonary function tests such as spirometry and other measures of static lung function against CTVI.
(2) Used PET/CT, SPECT/CT or contrast-enhanced CT as the ventilation imaging test modality (these will be not be classified as “CTVI” for the purposes of this paper).
(3) Lacked a statement of statistical significance.
(4) Did not describe, or reference to an article, in sufficient detail the method of generation of the CTVI scan.
(5) Did not describe, in sufficient detail, the level of spatial detail used for the cross-modality comparison.
(6) Was not a scientific paper (for example, conference abstract, patent, book, conference proceeding).
(7) Was not a new investigation (for example, a review or editorial).

Results

One hundred and forty-two records were identified through searching of Pubmed and Embase and handsearching. After the exclusion of 34 duplicate records, the abstracts of 108 records were reviewed. Twenty-eight records were rejected after abstract review for not meeting eligibility criteria. The majority of these were conference abstracts.

Paper Selection

Sixty-nine full text records were reviewed, and after detailed assessment twenty-three full text papers met selection criteria and were included in the final review (see Figure 1in the Supplementar material for details of the papers selected and eliminated at each stage). Forty-six records were rejected for 1) failing to include an intrapatient comparison with an accepted gold standard (24 records) 2) using contrast-based methods to assess ventilation (5) 3) being conference abstracts not full papers (13 records) 4) being review papers with no original data (4 records). Four papers were reported in animals and 19 papers were reported in humans. There were fourteen prospective studies. Altogether 579 human subjects were included, averaging 25.2 participants per study.

Technical aspects of 4DCT ventilation methods

Defining “normal” ventilation within a diseased lung is challenging as most assessments of high functioning lung are based on normalized values. In view of this, it is interesting that only 1 human study included normal subjects[18] and only 5 papers assess the accuracy of validation of both high and low functioning lung (labelled as “High” and “Defect” in Fig. 2)[19], [6], [20], [21], [10].

The majority of papers used deformable image registration (DIR) based methods for CTVI generation[2],[22],[23],[24],[1],[19],[25],[7],[18],[20], [26],[27],[28],[29], [4], [9],[6],[30],[10] although some papers evaluated both DIR and non-DIR based methods[6]'[10] (see Table 2 for details). The most common algorithms were DIR-HU and DIR-Jac (16 and 15 papers respectively).

Seventeen out of 23 papers used smoothing, to reduce CT noise prior to computing HU based CT images. Sixteen papers used masking to reduce the impact of image artefacts from DTPA deposition upon image assessment. Twelve authors using DIR-Jac metrics included some form of mass density correction to correct for respiratory induced changes in blood mass within the lung. Please refer to Figure 1 for an overview of papers and see the discussion for details of pre and post-processing techniques used in the papers reviewed.

Diverse methods were used to analyse the relationship between CTVI and the “gold-standard”, the commonest being Spearman Rank correlations (SPC), Dice Similarity Coefficients (DSC) and linear regression (LR) methods. The Spearman *r* values are defined in the range [-1, 1] and indicate the degree of monotonicity of values in spatially matched voxels within the whole lung ROI with 1indicating a perfect positive correlation. The Dice similarity coefficient (DSC) describes the fractional volume overlap between two regions (in our case, ventilation/perfusion defect regions or non-defect regions) and takes a value in the range [0, 1].

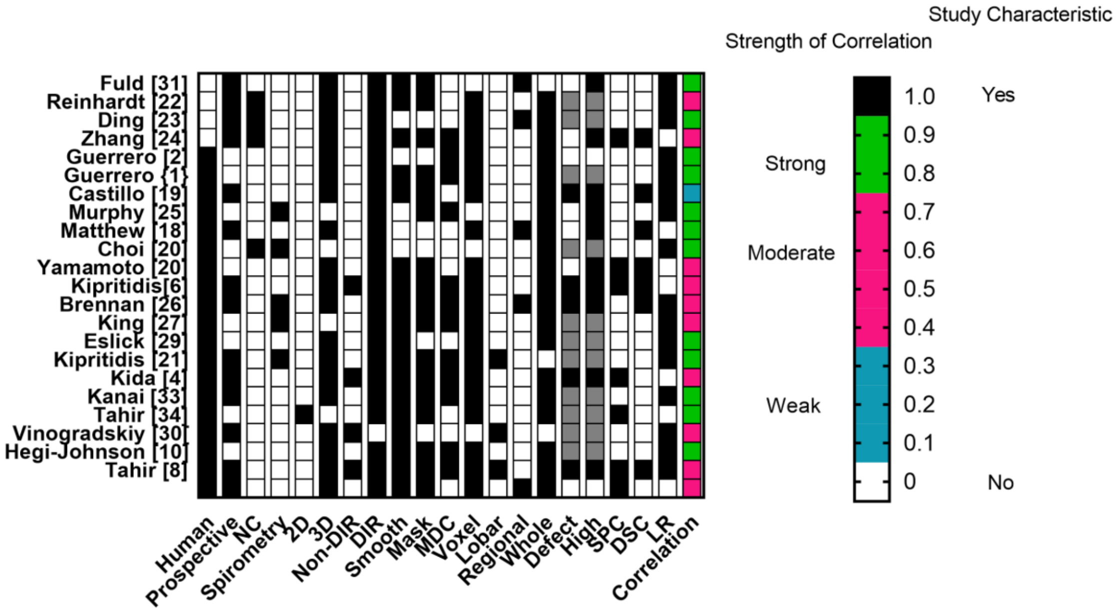


Figure 1: Characteristics of papers. This table provides an overview of study cohort (animal or human, prospective or retrospective and the inclusion of normal comparator (NC) subjects). The gold standard correlation is described, and may include spirometry or imaging (either 2 or 3D). Other details of the metrics such as post-processing with smoothing (Smooth) or masking (Mask), and the use of mass density corrected (MDC) metrics are tabulated. Finally, the methodology for comparison is described, whether this be voxel-based (Voxel)or based on a functional unit of lung (such as the lobe or whole lung) [Lobar, Regional and Whole], and whether respiratory defects (Defect) or high functioning lung (High) was evaluated. The statistical methodology used to evaluate the strength of correlation with the Spearman Rank Correlation (SPC), Dice Similarity Coefficient (DSC) or Linear Regression (LR) are also described. In this table the strongest average correlation reported in that paper is recorded for any type of correlation (includes SPC, DSC and LR). Grey boxes indicate characteristics which are not applicable to these papers.

Abbreviations: Deformable Image Registration (DIR), **Dice** Similarity Coefficients (DSC) and linear regression (LR) methods.

Table 2a Animal Studies and Table 2b Human Studies: Summary of CTVI studies. CTVI generation is defined as non-DIR and DIR based, which includes DIR Jacobian (DIR-Jac), DIR-HU and other hybrid DIR approaches (DIR-Hy). Papers used various

metrics for assessing the strength of correlation, including linear regression (LR), Spearman Rank correlations (SPC) and Dice Similarity Coefficients (DSC).

Abbreviations: Coefficient of Variance (CoV), forced expiratory volume in 1 second (FEV1), FVC forced vital capacity, Δ Vol ventilation calculation of ventilation based on change in volume.

Table 2a: Animal Studies of CT Ventilation

Study author (year)	Subjects No.	Comparative Imaging Modality	CTVI Generation	Type of correlation	Highest level of correlation	Range of correlation
Fuld et al (2008)**[31]**	4 sheep	Xenon-CT	Change in volume as assessed by change in HU units	Voxel based LR	LR *R*2 of 0.76	0.56–0.76
Reinhardt et al (2008)**[22]**	5 sheep	Xenon-CT	DIR-Jac	Voxel Based LR	LR *R*2 of 0.80	0.64–0.80
Ding et al 2012 **[23]**	4 sheep	Xenon-CT	2 variants of DIR-Jac (SAJ, SACJ) and DIR-HU	Voxel based LR	SAJ *r=0.97* SACJ *r=0.994* DIR-HU *r=0.952*	0.836-0.97 0.888-0.994 0.893-0.952
Zhang et al (2016)**[24]**	4 sheep	Xenon CT	3 metrics tested: 1. DIR Δ Vol 2.DIR-Jac 3. DIR- HU	Voxel based SPC	Δ Vol 0.61 Jacobian 0.61 HU 0.42	0.29-0.61 0.31-0.61 0.17–0.42

Table 2b: Human studies of CT ventilation

Study author (year)	Subjects No.	Comparative Imaging Modality	CTVI Generation	Type of Correlation	Highest average level of correlation	Range
Guerrero 2005**[2]**	22	Measured tidal volume on CT	DIR-HU	LR of lung volumes and tidal volumes.	DIR HU LR *r*= 0.985	LR *r* ranged from 0.982 -0.985

Guerrero 2006**[1]**	3 lung cancer RT patients	Measured tidal volume on CT	DIR-HU	Whole lung based	DIR-HU LR *r* = 0.985	NR
Castillo et al 2010**[19]**	7 thoracic oncology patients	DTPA-SPECT	DIR- HU DIR-Jac	Voxel based DSC	Highest average DSC was for 0-20% voxels	DSC ranged from 0.2-0.35
Murphy et al 2012**[25]**	216 patients with COPD	Spirometry	DIR-HU	Whole lung and lobar assessment LR of GOLD stage FEV1 and FEV1/FVC	Median *r* value of 0.87 for whole lung	0.85-0.91
Matthew et al 2012**[32]**	11 lung cancer patients	^{3}HE hyperpolarized MRI	DIR- HU	DSC to compare ventilated volume (VV) in both whole lung and lung ipsilateral to and contralateral to cancer	DSC 0.89±.01	0.69-0.95
Choi et al 2013**[18]**	30 asthma patients, 14 control subjects	PFT's comparison of Total Lung Capacity (TLC) and Air volume (AV) at exhale.	DIR-Jac derived from breath-hold CT	Global lung function LR	*R=0.87* for Total lung volume in severe asthmatics	*r* =0.78 to 0.87
Yamamoto et al 2013**[20]**	9 patients with thoracic cancer.	DTPA-SPECT	DIR- Jac	Voxel based SPC, DSC for segmented low-functional lung regions	Best Spearman rank 0.80 Best DSC 0.8	0.1-0.8 0.4-0.8.
Kipritidis et al 2014**[6]**	12 lung cancer patients	PET-Galligas	DIR-HU and DIR-Jac with and without density scaling	Voxel based SPC	Density-scaled HU Spearman $\bar{r}$ = 0.50 ± 0.17 and DSC (lowest 20%) = 0.52 ± 0.09	DIR-Jac 0.19±0.23 DIR-HU 0.50 ± 0.17

Yamamoto et al 2014**[5]**	18 patients- all with thoracic cancer	DTPA-SPECT and PFT's (spirometry and measurement of DLCO)	DIR- HU DIR-Jac.	Voxel based DSC to quantify overlap betweenV4DCT and VSPECT defect regions Pearson correlation with FEV1 and FEV1/FVC	DSC for DIR- HU 0.39	Average DSCs were: DIR-HU 0.39 ± 0.11 DIR-Jac 0.36 ±0.13 FEV1/FVC strongly correlated with 25% voxel value (0.73) and strongly negatively correlated with defective lung (-0.63).
Brennan et al 2015**[26]**	98 patients with lung cancer	Spirometry	DIR- HU and DIR-Jac	LR to compare spirometry and CTVI CoV V20, and visually defined defects.	Correlation coefficient ~ 0.7 for HU	DIR-HU CoV CC between: DIR-HU FEV1 0.72 FEV1/FVC 0.67 DIR-Jac FEV1 0.40 FEV1/FVC 0.38
King et al. 2015**[27]**	30 Thoracic radiotherapy patients	Tidal volume from 4DCT and CTVI compared	DIR-Jac DIR-Hy	Whole lung based Pearson correlation coefficient	TVInt and TVCT was 0.92 (P<.01) TVJac and TVCT was 0.97 (P<.01).	Not reported
Eslick et al 2015**[29]**	11 lung cancer patients	PET Galligas	DIR- HU	Lobar volumes and ventilation compared with LR	*r*-value 0.96 for comparison of lobar ventilation from CTVI and PET-Galligas.	Average *r*-value between CTVI analysed lobar ventilation vs. lobar volumes was 0.78
Kipritidis et	25 lung	PET-Galligas	Non-DIR	Voxel based	CTVI-HU	Mean ±SD

al 2016**[21]**	cancer patients		based HU DIR-HU DIR-Jac	SPC	0.50	correlation with Galligas PET was r = (0.50±0.17), (0.42±0.20), and (0.19±0.23) for the CTVI-HU, DIR-HU, and DIR-Jac methods
Kida et al 2016**[4]**	8 thoracic cancer patients	DTPA SPECT and Spirometry	DIR-HU and DIR-Jac	Radiotherapy lung metrics compared with Pearson correlation and LR	0.94 for DIR-HU	DIR-HU Pearson R=0.94 and linear regression = 0.71 DIR-Jac R=0.85; slope=0.5
Kanai et al 2016**[33]**	11 lung cancer patients	Planar Kr images	DIR-HU and DIR-Jac	Voxel based SPC	HU 0.875	Mean ± SD: DIR-HU 0.875±0.07 DIR-Jac 0.803±0.114
Tahir et al 2016**[34]**	30 patients with sputum eosinophilia and asthma	Hyperpolarized ^{3}He MRI	Breath-hold CT at total lung capacity and functional residual capacity used to assess change in volume of lobes	Lobar Pearson correlation of all lobar regions	0.65	Range of Pearson correlations not stated
Vinogradskiy et al 2017**[30]**	16 lung cancer patients	DTPA SPECT and spirometry	DIR- HU	Global lung function ROC analysis to compare, 4DCT-ventilation-based preop FEV1 vs. SPECT based preop FEV1.	0.99 correlation coefficient for prediction of ventilation changes after lobectomy using CTVI	Pneumonectomy: Correlation coefficient .80(0.81for nuclear medicine-ventilation And 0.78 for nuclear medicine-perfusion). Lobectomy: Correlation coefficient was 0.99 for CTVI.
Hegi-	11 lung	Technegas	CTVI-	Voxel based	Non-	Defect regions

Johnson et al 2017**[10]**	cancer patients	SPECT	HU, DIR-HU and DIR-Jac	Spearman Rank and DSC, lobar based Pearson Correlation and whole lung CoV	defect regions: CTVI HU, DIR-HU and DIR-Jac mean DSC of 0.69,0.68, and 0.54.	mean DSC were 0.39, 0.33, and 0.44. Spearman r :0.26, 0.18 and -0.02 for CTVIHU, DIR-HU and DIR-Jac respectively
Tahir et al 2018**[8]**	11 lung cancer patients	^{129}Xe and ^{3}He MRI	DIR-HU, DIR-Jac, Specific gas volume change	Spearman rank correlations of different ROI sizes	DIR-HU SPC *R*= 0.37, DIR-Jac SPC 0.31, Specific gas volume 0.34	Voxel-level: 0.1-0.8 ROI 20x20 voxels: *R=0.2-0.9*

CTVI metrics

Evaluating the precision of DIR

The majority of papers selected (22 out of 23 papers) used a DIR based approach to CTVI generation, although several of these also tested non-DIR based approaches on the same subject group. The commonest DIR approach was B-spline based with 12 papers using some variation of B-spline algorithms [1],[2],[6],[8] ,[10],[18],[21],[23],[24],[25] ,[27], [29]. Just over half of papers (13 out of 23) either discussed the method used to assess registration accuracy or referenced this within the text. Target registration errors (TRE) were quantified or referenced for DIR methods in 11 papers, with the other papers using visual assessment (2) and semi-automated landmark analysis (1). Most papers reported TRE of <1.5 mm indicating that DIR was accurate. However, the lack of reporting in 9 of the papers using a DIR CTVI approach is potentially problematic, as this is critical for the accurate calculation of regional ventilation. Please see our supplementary files for details of image registration methodology and registration accuracy assessment in individual papers.

Comparing ventilation metrics

Eighteen papers evaluated HU based metrics, with all papers evaluating DIR-HU; within this group, 3 papers evaluated non-DIR HU metrics in addition to DIR based approaches. Fifteen papers evaluated DIR-based Jacobian metrics. For details of the comparison methodology and strength of correlation, please refer to Figures 3a and 3b.

Average ± S.D. linear regression correlation coefficients were 0.73±0.24 and 0.86± 0.11 for DIR-Jac and DIR-HU respectively, and the SPC were 0.45±0.31and 0.41±0.11 for DIR-Jac and DIR-HU respectively.

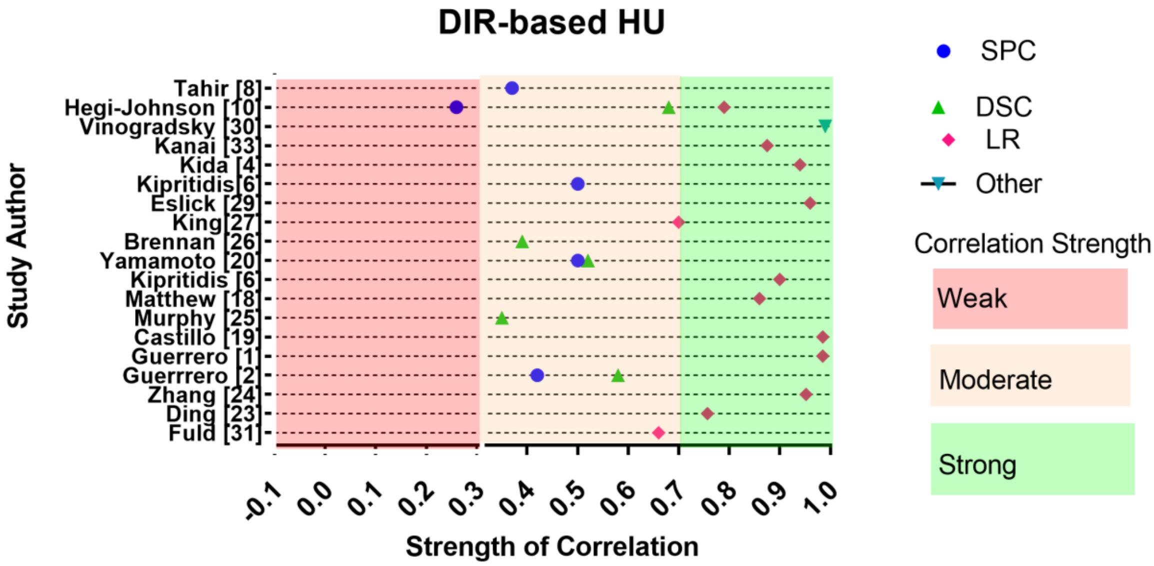


Figure 2a

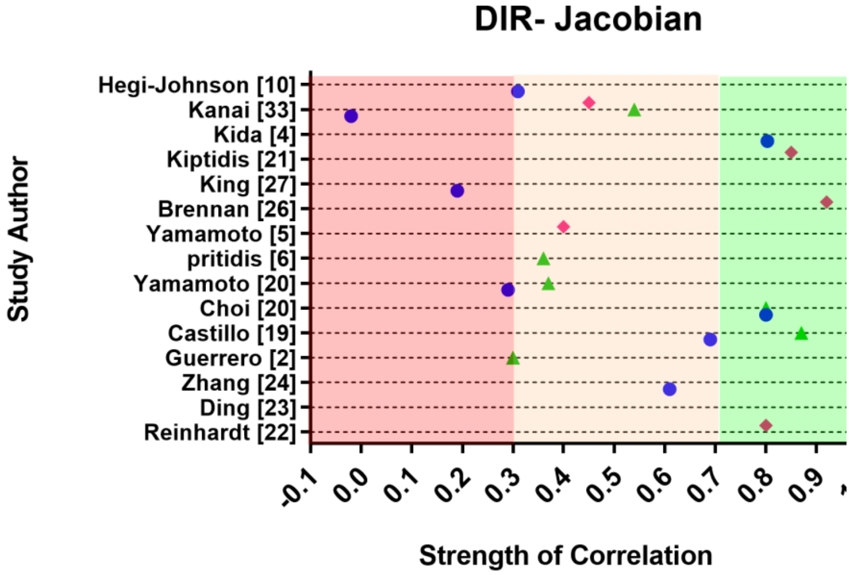


Figure 2b

Figure 2a and 2b: SPC, DSC and LR for DIR-HU (Fig. 3a) and DIR-Jacobian (Fig, 3B) respectively. Correlations are graded as either weak (0-0.3), moderate (<0.3-0.7) or strong (<0.7-1.0).

The number of papers consistently reporting on the size of region of interest (ROI) used in comparison was too small to draw strong conclusions, although in 2 papers which compared small ROI's (<1cm^3) vs. large ROI's (4 mm slices of the whole lung from top to bottom) the linear regression correlation coefficients were quite different, being 0.56±0.49 and 0.84±0.16 respectively.

Discussion

The primary function of the lung is gas exchange of which ventilation, perfusion and diffusion are fundamental components; oxygen from the air and carbon dioxide excreted by the body and dissolved in the blood is exchanged across the alveolar membrane. Even in healthy lungs this process is dynamic, with heterogeneous ventilation throughout the lung, partly because of mechanical issues such as the difference in pressure between the top and bottom of the lung, and partly because of rapid changes within the pulmonary vasculature. The pattern of local pathology can differ in patients with lung disease according to the underlying aetiology. Patients with severe COPD have obstructive pulmonary function test results and have large areas of the lung that are not ventilated due to flow limitation/hyperinflation. In patients with emphysema, spirometry can be remarkably normal, but diffusion is impaired due to destruction of alveoli. In asthmatic patients, obstruction can be reversible, but in severe cases it can be irreversible as well. In pulmonary vascular disease, ventilation may be adequate, but the ventilated lung is inadequately perfused. Finally, in interstitial lung disease, there is a restrictive pulmonary function, with both ventilation (although in a restrictive pattern: i.e. normal or increased FEV1/FVC ratio, but decreased tidal volume, decreased TLC and FVC) and perfusion present in the diseased areas of lung, but gas diffusion may be impaired depending on the cause of the restrictive pattern (interstitial lung disease vs. thoracic wall pathology) To make matters more complicated, patterns can overlap with multiple pathologies present in patients with severe lung disease.

All of the papers included in this review investigate the use of imaging technologies to define regional ventilation, although some have included perfusion scans as well[6],[10]. Given the dependence of gas exchange on perfusion, ventilation and diffusion, can imaging technologies which do not assess perfusion or diffusion provide useful information? We know from previous studies that perfusion scans can demonstrate changes after radiotherapy, and there is some evidence that they may strongly correlate with pulmonary pathology[35],[36]. Therefore, it seems likely that CTVI's inherent focus on ventilation has limited its applicability to patients with mainly gas exchange problems without an impairment in ventilation. In this review, we have chosen to limit our assessment to papers focusing on ventilation for two reasons. Firstly, CT scans are now used as part of the standard workflow in the radiotherapy planning and in the assessment of surgical resection candidates. They are cheap and widely available making CTVI a highly accessible technology. Secondly, CTVI has rapidly progressed into clinical trials, and we wished to assess the robustness of the literature, which has focused largely on technical rather than clinical validation measures.

Defining the "gold standard': Alternative Imaging Methods to Assess Regional Ventilation with hyperpolarized gas MRI and nuclear imaging.

Nuclear Medicine Assessments of Pulmonary Function

V/Q SPECT and V/Q SPECT-CT are established and widely available modalities for the assessment of regional pulmonary function. During V/Q SPECT acquisition, a ventilation scan is acquired, followed by a scan assessing perfusion. Several isotopes are commercially available to assess ventilation, but those in common use include 99mTc-labeled particulate aerosols such as 99mTc-diethylenetriaminepentaacetic acid (99mTc-DTPA) or the ultrafine carbon–labelled nanoparticle 99mTc- Technegas (Cyclomedica). Subsequently, 99mTc-macroaggregated albumin is administered and a perfusion scan acquired[37]. More recently, V/Q PET-CT using Gallium-68 (^{68}Ga), commonly referred to as Galligas, and ^{68}Ga- macroaggregated albumin has been developed, allowing higher resolution imaging of radioisotope uptake compared to SPECT[38]. However, ^{68}Ga is unlikely to be widely clinically implemented due to the need for a ^{68}Ga- generator within departments. All these radioisotope techniques suffer to some extent from issues of clumping in the central airways, although peripheral airway distribution can be improved by using smaller particles such as Technegas or Galligas, and by careful ventilation of the patient with deep tidal breathing during radioisotope inhalation to ensure even distribution throughout the lung parenchyma. Their great advantage is the ability to simultaneously image perfusion, and they have been shown to be sensitive and specific for the diagnosis of pulmonary emboli, although they are not routinely used for the diagnosis of other pulmonary pathologies[37].

Limitations of spirometry based pulmonary function tests

Spirometry remains the most common PFT performed in the assessment of lung cancer patients, but is limited to the assessment of obstructive pulmonary disease (and although it may be abnormal in restrictive pulmonary pathology, is not diagnostic) and does not provide information about the function of the lung parenchyma where gas exchange occurs. Also, spirometry does not measure hyperinflation in the lung; these areas of hyperinflated lung, as represented by the large residual volume (RV) in patients with severe obstructive COPD do not contribute to gas exchange. Spirometry is highly dependent on respiratory effort, and incorrectly performed spirometry may be non-diagnostic. This introduces great variability into their performance by individual patients who may have similar pathological profiles within the lung parenchyma.

MRI using hyperpolarized gases and the assessment of regional lung function

Hyperpolarized gas MRI, which gives detailed information about both the lung microstructure and regional ventilation, was developed in part because of the need for more detailed evaluation of parenchymal pathology than provided by spirometry and nuclear imaging. Both ^{3}He[39],[40] and ^{129}Xe[41],[42] have been used to assess regional pulmonary physiology. Studies with ^{3}He have shown strong correlations to conventional PFTs (i.e. spirometry) in patients with a variety of pulmonary pathologies including COPD and asthma[43],[44]. ^{129}Xe diffuses across the alveolar membrane, and can also be used to investigate gas diffusion into the blood stream. Although, this property also causes some anaesthetic side-effects, which initially limited clinical studies, advances in the methodology for polarization of ^{129}Xe have increased the quality of the imaging, and reduced the dose required for imaging to around 0.5-1.0L for each study, thereby significantly reducing this adverse effect. ^{129}Xe diffuses through the alveolar membrane and into the red blood cells in the blood stream, producing distinct resonant signal frequencies in each vascular compartment and has great potential for studying gas diffusion between the lungs and the blood stream.

Diffusion-weighted ^{3}He and ^{129}Xe MRI can be used to evaluate the apparent diffusion coefficient (ADC) of lung parenchyma, which can give highly detailed information about lung microstructure at the alveolar level. These studies have been shown to correlate with early ultrastructural changes seen on MRI.

Hyperpolarized gas MRI static ventilation images show the regional distribution of inhaled noble gas and have been used by Tahir et al, who demonstrated moderate correlations between them and CTVI at the lobar and voxel levels [34], [8]. In many respects, hyperpolarized gas MRI provides a level of anatomical and physiological detail that is not yet available in 4DCT. The high cost of ^{3}He has made it less than ideal for widespread implementation in clinical practice. However, with the maturation of ^{129}Xe MRI, clinical implementation is becoming feasible, making this into an accessible technology for future studies.

Common elements and limitations in CT Ventilation Studies

CT Acquisition and Issues of CT quality

Initial steps towards the development of CTVI occurred in the early 2000's with the publications of methods to derive information about regional ventilation from CT datasets[1]'[31]. The three key steps in the generation of a DIR-based CTVI image are essentially identical: (1) acquisition of a respiratory-correlated CT scan, most commonly 4DCT or sometimes BHCT, (2) application of DIR typically between the exhale and inhale phase images, and (3) computation of a ventilation metric either directly on the DIR motion field (e.g. DIR-Jac) or using the motion field to process changes in HU values for spatially registered voxels in the inhale and exhale images (DIR-HU) or both (DIR-Hy).

The first 10 years of development of the CTVI methodology have been marked by the wide variety of technical processes that have been explored and are currently in use to generate CTVI images (see Figure 3).

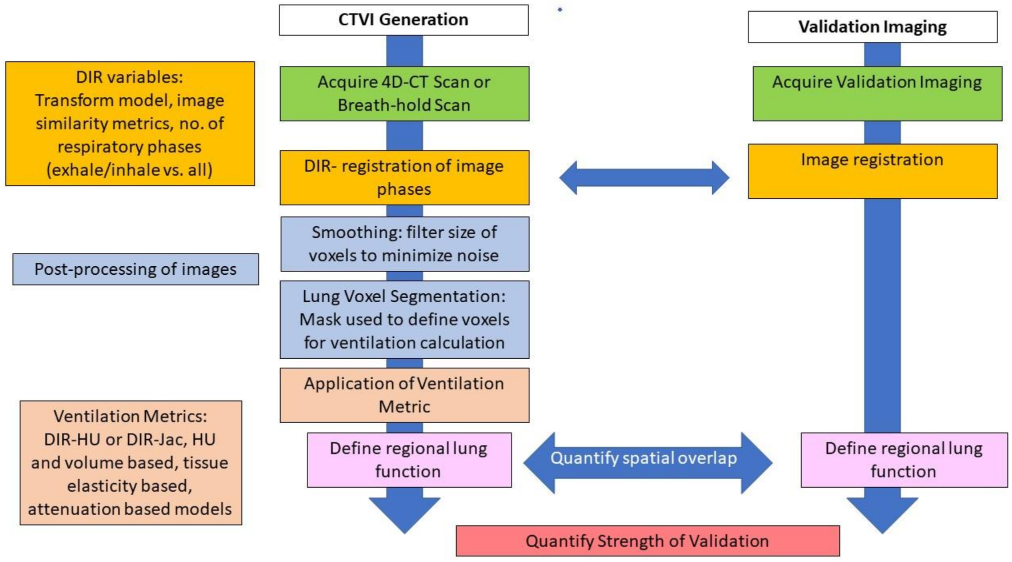


Figure 3: Overview of typical workflow for a DIR-based CTVI validation study including common steps for CTVI generation (left arm) and validation against contrast-based ventilation imaging (right arm). This demonstrates the multiple variables that should be considered when reporting and assessing CTVI validation studies. These variables include details of the DIR methodology (orange boxes), including the metrics used to perform DIR and the details of the respiratory phases included in the registration. Post-processing (blue boxes) may occur both before or after the CTVI image is generated (please see the text for details of this). Finally, the metrics that may be applied are diverse (brown boxes), although DIR-Jac and DIR-HU are the commonest. Common steps in the CTVI validation workflow are in the same colour as the CTVI generation pathway.

The majority of CTVI validation studies have focused on the use of 4DCT, which involves image reconstruction into 5-10 different respiratory “phase bins” based on the synchronous acquisition of CT projection data and a breathing motion signal [45]. In clinical human studies, the 4DCT scan is often performed under free-breathing (FB), which can lead to the well-known problems of anatomic truncation, duplication and blurring artefacts that arise due to irregular breathing motion, for example, coughing or changes in breathing period / amplitude during the scan. It has been reported that up to 90% of clinical 4DCT scans suffer anatomic imaging artefacts of magnitude >4mm[46].

Some CTVI studies have attempted to overcome the issue of irregular breathing using some form of AV biofeedback to increase the regularity of breathing[28], [29], or by using different methods of sorting the CT by matching bins based on anatomical features rather than the phase of respiration[20]. CTVI can also be derived from exhale/inhale image pairs acquired during breath-hold[2],[29],[18], although acquiring these images in thoracic cancer patients with impaired respiratory function can be challenging. By comparison, in animal studies the subjects are anaesthetized during the 4DCT scan, resulting in highly regular breathing motion with minimal 4DCT reconstruction artefacts. Other interesting methods to reduce image noise included using anatomic sorting rather than phase based sorting to reduce artefacts within 4DCT images[20].

Pre/Post-Processing: Smoothing and Masking Images

The majority of studies in this review applied smoothing to either the input 4DCT images, the resulting CTVIs, the corresponding contrast-based ventilation images, or some combination of these. Various smoothing techniques were applied including applying a Gaussian, median or averaging filter to minimize noise within the raw 4DCT [1], [4], [10] [25],[24], [19],[25], [18], [26],[29], [21],[33],[30], or CTVI image[32],[33], [10] or by applying smoothing functions during the generation of the CTVI image[18],[20]. There is some evidence to suggest that the method of filtering can affect the strength of the correlation coefficients, although this appears to plateau at a median filter radius of 3x3x3 voxels[8]. The majority of authors also apply some “masking” of lung voxels as a means of defining which voxels should be included in the validation analysis or to improve lung image registration accuracy by limiting the cost function to lung voxels only. Lung masking may be performed based on the 4DCT scan (e.g. using a threshold-based or region-growing segmentation or by manually delineating the lung parenchyma)[26], [30],[10], based on the ventilation scan (e.g. by setting a minimum cut-off for the radioisotope uptake in a SPECT or PET scan)[6],[21]or sometimes using information in both. The central airways are often a site of focal clumping artefacts in DTPA SPECT [19],[28] and even in PET-Galligas[6] and investigators dealt with this by manually excluding obvious areas of abnormally high signal[6], or using automated filters to exclude these voxels.

Ventilation Metric

Ventilation metrics may be classified into DIR-based vs. non-DIR based algorithms. The commonest DIR metrics evaluate vowel-wise HU changes between spatially aligned images

(DIR-HU)[1],[2] or regional volume changes based on the Jacobian determinant of the DIR motion field (DIR-Jac)[22]. A small number of “hybrid” metrics have also been investigated, which may combine information about HU and volume changes to model lung elasticity as an alternate surrogate for lung function (DIR-Hy)[8] ,[27]. Non-DIR HU metrics ($_{\text{non}}$DIR-HU) have been found to be potentially robust against 4DCT motion artefacts, and use average HU values to model blood-gas exchange in the lung parenchyma[10],[21]. Physiological ventilation is a process of blood-gas exchange with diffusion of oxygen and carbon dioxide across the alveolar membrane. The HU values of a voxel reflect the air and tissue content of that particular part of the lung parenchyma, and in the nonDIR-HU model this is used as a surrogate for the capacity of blood-gas exchange at each voxel. As this approach relies on the average intensity projection of the 4DCT, it is closer to the average scans acquired in nuclear medicine imaging, and also reduces the impact of respiratory artefacts on CTVI image quality.

Statistical methods used for cross-modality comparisons

A number of methods have been used to analyse the accuracy of CTVI with respect to paired contrast-based comparator ventilation scans, including the SPC, DSC and Linear regression methods.

The size of the region of interest (ROI), relative to the pixel or voxel dimensions, has been seen to affect the strength of cross-modality correlation values obtained[8], with stronger correlations seen for larger ROIs. It is thought that averaging out ventilation values over larger ROIs can mitigate errors created by mis-registration between the CTVI and contrast-based comparator scans, errors in the DIR process or imaging artefacts in either the 4DCT or contrast-based comparator scans [10], [31]. This is particularly relevant in studies that have used clinically acquired imaging and hence, although voxel based comparisons are important to benchmark different CTVI and DIR methodologies, CTVI may be most robust when defining loco-regional function at larger anatomical distance scales, such as lung lobes[10],[29].

Different semi-automated approaches exist for thresholding lung ventilation images into high or low-function lung zones; however, there exists no consensus on which is the best approach. Since CTVI is amenable to quantitative analysis, a number of CTVI studies have defined low function lung as referring to those lung voxels with ventilation values less than the 20th-30th percentile ventilation for that patient. In contrast, clinical assessment of nuclear medicine ventilation images is usually based on the visual analysis of scans by a physician.

Relatively little work has been carried out to validate semi-automated thresholding of lung ventilation images against clinical assessment [21]. It is encouraging that the correlation of functional dose with toxicity outcomes (e.g. Grade 3+ pneumonitis) may be relatively stable despite different methods for weighting the ventilation values in CTVI [47].

Assessing the Quality of the Literature

CTVI is a promising imaging technology, however the CTVI literature is hampered by the heterogeneity in the methodology for CTVI generation, the choice of reference modality and method of cross-modality comparison and issues of study quality, with only 14 small prospective studies. These issues of study quality and heterogeneity explain the wide variation in the strength of correlation seen. Indeed, even comparing a single parameter across the literature is difficult, with SPC varying between SPC were 0.45±0.31and 0.41±0.11 for DIR-Jac and DIR-HU respectively. The wide spread in the standard deviations highlights the small numbers of papers, low patient numbers and the variability in methods of analysis in this literature.

The use of comparators at the lobar or whole-lung level can mask gross errors at the voxel level. Most papers using PFTs as the comparators relied on spirometry[25],[18], [28], [30], and although it may be argued that spirometry provides an adequate measure of ventilation and is hence a valid modality for comparison with CTVI, it is likely to underestimate the impact of pulmonary diseases that are not adequately reflected by spirometry, such as emphysema and pulmonary vascular disease.

In Table 4, we provide a summary of the key aspects of the CTVI validation literature. Both DIR and non-DIR CTVI ventilation metrics have achieved robust correlations, but particular approaches may suit different datasets. For example, as they use the average projection of the 4DCT to generate the CTVI image, non-DIR approaches may be more robust when using clinically acquired 4DCT with significant respiratory artefacts. On the other hand, very high-quality images may be acquired clinically by using BH approaches. Although, BH was only investigated in 4 human studies[2],[8],[25],[29] this may help to overcome the quality issues created by irregular breathing motion. Other strategies, such as AV biofeedback may also be useful, but there is insufficient information in the current literature to advocate it for routine implementation. Similarly, the heterogeneity in the methodologies used for post-processing make it difficult to make recommendations.

Table 3: Summary of Findings Table

Question: Can CTVI defined regional ventilation provide the same information as contrast-based ventilation imaging and pulmonary function tests?

Population: Patients having 4DCT or BHCT, the majority of whom are undergoing radiotherapy planning

Reference standard: contrast-based imaging of regional lung function and/or spirometry

Findings:

- LR (Average ± S.D.): 0.73±0.24 and 0.86± 0.11 for DIR-Jac and DIR-HU
- SPC (Average ± S.D.): 0.45±0.31and 0.41±0.11 for DIR-Jac and DIR-HU

Heterogeneity in reporting standards reduces the ability to assess statistical significance.

Recommendation		No of relevant studies	Strength
DIR vs. Non-DIR	There is insufficient evidence to recommend DIR over non-DIR approaches. If DIR based CTVI metrics are used the quality of the DIR should be assessed	Only 4 non-DIR studies	Moderate Weak to moderate
Selection of metric	There is insufficient evidence to recommend one CTVI metric over another and further head to head is required.	19 DIR-HU 15 DIR-Jac	Moderate
CT acquisition and processing	CTVI based on both 4DCT and BHCT have shown moderate to strong correlations with clinical gold standards. There is insufficient evidence to guide the use of one method of acquisition above another.	4 BH studies	Moderate to strong
	There is insufficient evidence to support the use of AV biofeedback routinely in the acquisition of CT images: studies using AV biofeedback have similar strengths of validation compared to non-AV biofeedback studies.	2 AV biofeedback studies	Weak
	Post-processing such as smoothing and masking may influence the quality of the CTVI image, but there is insufficient evidence to recommend for or against their use.	18 papers used post-processing, but methodology was heterogeneous.	Weak
Clinical Validation Modality	The majority of validation studies have used 3D-imaging and this should be used for future validation studies for them to be benchmarked against existing methodologies.	19 papers used 3D imaging	Moderate to strong

	Standardized criteria for clinical assessment (e.g. GOLD criteria) should be used for validation modalities to increase the clinical relevance of future validation studies.	5 papers reported validation against spirometry	Moderate

This lack of standardization makes it difficult to determine whether CTVI is robust enough to implement in clinical practice. To improve the standardization of future validation studies, we recommend: 1) the use of high quality, prospectively collected datasets and that, where possible, these should include normal and disease cohorts 2) documentation of respiratory function of individual subjects as assessed by pulmonary function tests which are recognized clinical standards (spirometry or cardio-pulmonary exercise testing), 3) incorporation of high quality ventilation imaging as a comparator, 4) report detailed methodology for their CTVI imaging, including the details of CT image acquisition, use of post-processing techniques such as smoothing, masking, methodology of DIR and an assessment of its accuracy 5) report the strength of validation across both high functioning and low functioning lung using a cohort of standardized statistical assessments. We would suggest that these tests should include voxel-based assessments using the SPC and DSC as a minimum to allow new modalities to be benchmarked against existing studies as well as more clinically relevant regional volumes. We would also encourage the investigation of other comparative methodologies. For example, although outside of this review, it would be powerful to compare CTVI defined regional ventilation with histopathological specimens and this could facilitate the development of imaging surrogates for different types of COPD, increasing the non-invasive options to diagnose these diseases.

Finally, given the complexity of pulmonary pathology found in thoracic oncology patients, it is possible that CTVI may be most useful in patients who are affected by obstructive lung diseases such as COPD, but alternative imaging modalities may be required in other patients. For example, hyperpolarized dissolved-phase ^{129}Xe MRI may be particularly useful in patients with interstitial lung disease[48], and V/Q SPECT and PET in patients with pulmonary vascular pathology.

Judgement on the quality of the CTVI validation literature will ultimately be determined by the outcome of prospective clinical studies investigating CTVI implementation in radiotherapy patients (NCT02528942, NCT02308709, NCT02843568) [49], [47]. The awaited publication of the VAMPIRE challenge, which uses a range of clinical imaging datasets to compare DIR and CTVI methodologies will also provide new insights into the variabilities and uncertainties associated with this technology[50].

Conclusion

CTVI has shown moderate to strong voxel-based correlations in most human studies. However, CTVI is being increasingly incorporated into the clinical workflow of thoracic radiotherapy and is undergoing clinical validation. Our results show that further refinement and standardization of CTVI methodology will enable better comparative studies and a more robust application of this technology in clinical practice. CTVI appears to be of relevance in radiotherapy planning, particularly in patients whose main pulmonary impairment is not a gas exchange problem, and awaits clinical validation in prospective clinical trials.

References

[1] Guerrero T, Sanders K, Castillo E, Zhang Y, Bidaut L, Pan T, et al. Dynamic ventilation imaging from four-dimensional computed tomography. Phys Med Biol 2006;51:777–91.

[2] Guerrero T, Sanders K, Noyola-Martinez J, Castillo E, Zhang Y, Tapia R, et al. Quantification of regional ventilation from treatment planning CT. Int J Radiat Oncol Biol Phys 2005;62:630–4.

[3] Yamamoto T, Kabus S, Bal M, Keall P, Benedict S, Daly M. The first patient treatment of computed tomography ventilation functional image-guided radiotherapy for lung cancer. Radiother Oncol 2016;118:227–31.

[4] Kida S, Bal M, Kabus S, Negahdar M, Shan X, Loo BW, et al. CT ventilation functional image-based IMRT treatment plans are comparable to SPECT ventilation functional image-based plans. Radiother Oncol 2016;118:521–7.

[5] Yamamoto T, Kabus S, Lorenz C, Mittra E, Hong JC, Chung M, et al. Pulmonary ventilation imaging based on 4-dimensional computed tomography: Comparison with pulmonary function tests and SPECT ventilation images. Int J Radiat Oncol Biol Phys 2014;90:414–22.

[6] Kipritidis J, Siva S, Hofman MS, Callahan J, Hicks RJ, Keall PJ. Validating and improving CT ventilation imaging by correlating with ventilation 4D-PET/CT using 68Ga-labeled nanoparticles. Med Phys 2014;41:011910–1 to 011910–2.

[7] Mathew L, VanDyk J, Etemad-Rezai R, Rodrigues G, Parraga G. Hyperpolarized (3) He pulmonary functional magnetic resonance imaging prior to radiation therapy. Med Phys 2012;39:4284.

[8] Tahir BA, Hughes PJC, Robinson SD, Marshall H, Stewart NJ, Norquay G, et al. Spatial comparison of CT-based surrogates of lung ventilation with hyperpolarized Helium-3 and Xenon-129 gas MRI in patients undergoing radiation therapy. Int J Radiat Oncol Biol Phys n.d.

[9] Kanai T, Kadoya N, Ito K, Kishi K, Dobashi S, Yamamoto T, et al. Evaluation of four-dimensional computed tomography (4D-CT)-based pulmonary ventilation: The high correlation between 4D-CT ventilation and 81m Kr-planar images was found. Radiother Oncol 2016.

[10] Hegi-Johnson F, Keall P, Barber J, Bui C, Kipritidis J. Evaluating the Accuracy of 4D-CT Ventilation Imaging: First Comparison with Technegas SPECT Ventilation. Med Phys 2017;44:4045–55.

[11] Mistry NN, Diwanji T, Shi X, Pokharel S, Feigenberg S, Scharf SM, et al. Evaluation of fractional regional ventilation using 4D-CT and effects of breathing maneuvers on ventilation. Int J Radiat Oncol Biol Phys 2013;87:825–31.

[12] Ireland RH, Tahir BA, Wild JM, Lee CE, Hatton MQ. Functional image-guided radiotherapy planning for normal lung avoidance. Clin Oncol 2016;28:695–707.

[13] Bucknell NW, Hardcastle N, Bressel M, Hofman MS, Kron T, Ball D, et al. Functional lung imaging in radiation therapy for lung cancer: A systematic review and meta-analysis. Radiother Oncol 2018.

[14] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151:264–9.

[15] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Radiology 2015;277:826–32.

[16] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36.

[17] Simon BA. Non-invasive imaging of regional lung function using x-ray computed tomography. J Clin Monit Comput 2000;16:433–42.

[18] Choi S, Hoffman EA, Wenzel SE, Tawhai MH, Yin Y, Castro M, et al. Registration-based assessment of regional lung function via volumetric CT images of normal subjects vs. severe asthmatics. J Appl Physiol 2013;115:730–42.

[19] Castillo R, Castillo E, Martinez J, Guerrero T. Ventilation from four-dimensional computed tomography: density versus Jacobian methods. Phys Med Biol 2010;55:4661.

[20] Yamamoto T, Kabus S, Lorenz C, Johnston E, Maxim PG, Diehn M, et al. 4D CT lung ventilation images are affected by the 4D CT sorting method. Med Phys 2013;40.

[21] Kipritidis J, Hofman MS, Siva S, Callahan J, Le Roux P-Y, Woodruff HC, et al. Estimating lung ventilation directly from 4D CT Hounsfield unit values. Med Phys 2016;43:33–43.

[22] Reinhardt JM, Ding K, Cao K, Christensen GE, Hoffman EA, Bodas SV. Registration-based estimates of local lung tissue expansion compared to xenon CT measures of specific ventilation. Med Image Anal 2008;12:752–63.

[23] Ding K, Cao K, Fuld MK, Du K, Christensen GE, Hoffman EA, et al. Comparison of image registration based measures of regional lung ventilation from dynamic spiral CT with Xe-CT. Med Phys 2012;39:5084–98.

[24] Zhang GG, Latifi K, Du K, Reinhardt JM, Christensen GE, Ding K, et al. Evaluation of the ΔV 4D CT ventilation calculation method using in vivo xenon CT ventilation data and comparison to other methods. JACMP 2016;17:550–60.

[25] Murphy K, Pluim JP, van Rikxoort EM, de Jong PA, de Hoop B, Gietema HA, et al. Toward automatic regional analysis of pulmonary function using inspiration and expiration thoracic CT. Med Phys 2012;39:1650–62.

[26] Brennan D, Schubert L, Diot Q, Castillo R, Castillo E, Guerrero T, et al. Clinical validation of 4-dimensional computed tomography ventilation with pulmonary function test data. Int J Radiat Oncol Biol Phys 2015;92:423–9.

[27] King MT, Maxim PG, Diehn M, Loo BW, Xing L. Analysis of long-term 4-dimensional computed tomography regional ventilation after radiation therapy. Int J Radiat Oncol Biol Phys 2015;92:683–90.

[28] Yamamoto T, Kabus S, Lorenz C, Mittra E, Hong JC, Chung M, et al. Pulmonary ventilation imaging based on 4-dimensional computed tomography: comparison with pulmonary function tests and SPECT ventilation images. Int J Radiat Oncol Biol Phys 2014;90:414–22.

[29] Eslick EM, Bailey DL, Harris B, Kipritidis J, Stevens M, Li BT, et al. Measurement of preoperative lobar lung function with computed tomography ventilation imaging: progress towards rapid stratification of lung cancer lobectomy patients with abnormal lung function. Eur J Cardiothorac Surg 2015;49:1075–82.

[30] Vinogradskiy Y, Jackson M, Schubert L, Jones B, Castillo R, Castillo E, et al. Assessing the use of 4DCT-ventilation in pre-operative surgical lung cancer evaluation. Med Phys 2017;44:200–8.

[31] Fuld MK, Easley RB, Saba OI, Chon D, Reinhardt JM, Hoffman EA, et al. CT-measured regional specific volume change reflects regional ventilation in supine sheep. J Appl Physiol 2008;104:1177–84.

[32] Mathew L, Wheatley A, Castillo R, Castillo E, Rodrigues G, Guerrero T, et al. Hyperpolarized 3 He magnetic resonance imaging: comparison with four-dimensional x-ray computed tomography imaging in lung cancer. Acad Radiol 2012;19:1546–53.

[33] Kanai T, Kadoya N, Ito K, Kishi K, Dobashi S, Yamamoto T, et al. Evaluation of four-dimensional computed tomography (4D-CT)-based pulmonary ventilation: The high correlation between 4D-CT ventilation and 81m Kr-planar images was found. Radiother Oncol 2016;119:444–8.

[34] Tahir BA, Van Holsbeke C, Ireland RH, Swift AJ, Horn FC, Marshall H, et al. Comparison of CT-based lobar ventilation with 3He MR imaging ventilation measurements. Radiology 2016;278:585–92.

[35] Seppenwoolde Y, De Jaeger K, Boersma LJ, Belderbos JSA, Lebesque JV. Regional differences in lung radiosensitivity after radiotherapy for non-small-cell lung cancer. Int J Radiat Oncol Biol Phys 2004;60:748–58.

[36] Marks LB. Physiology-based studies of radiation-induced normal tissue injury. Radiother Oncol J Eur Soc Ther Radiol Oncol 1999;51:101.

[37] Roach PJ, Schembri GP, Bailey DL. V/q scanning using spect and spect/ct. J Nucl Med 2013;54:1588–96.

[38] Andreeff M, Wunderlich G. PET aerosol lung scintigraphy using Galligas. Eur J Nucl Med Mol Imaging 2010;37:175.

[39] Ebert M, Grossmann T, Heil W, Otten EW, Surkau R, Thelen M, et al. Nuclear magnetic resonance imaging with hyperpolarised helium-3. The Lancet 1996;347:1297–9.

[40] MacFall JR, Charles HC, Black RD, Middleton H, Swartz JC, Saam B, et al. Human lung air spaces: potential for MR imaging with hyperpolarized He-3. Radiology 1996;200:553–8.

[41] Mugler III JP, Driehuys B, Brookeman JR, Cates GD, Berr SS, Bryant RG, et al. MR imaging and spectroscopy using hyperpolarized 129Xe gas: preliminary human results. Magn Reson Med 1997;37:809–15.

[42] Albert MS, Cates GD, Driehuys B, Happer W, Saam B, Springer Jr CS, et al. Biological magnetic resonance imaging using laser-polarized 129Xe. Nature 1994;370:199.

[43] Fain SB, Panth SR, Evans MD, Wentland AL, Holmes JH, Korosec FR, et al. Early emphysematous changes in asymptomatic smokers: detection with 3He MR imaging. Radiology 2006;239:875–83.

[44] Kirby M, Mathew L, Wheatley A, Santyr GE, McCormack DG, Parraga G. Chronic obstructive pulmonary disease: longitudinal hyperpolarized 3He MR imaging. Radiology 2010;256:280–9.

[45] Sonke J-J, Lebesque J, Van Herk M. Variability of four-dimensional computed tomography patient models. Int J Radiat Oncol Biol Phys 2008;70:590–8.

[46] Yamamoto T, Langner U, Loo BW, Shen J, Keall PJ. Retrospective analysis of artifacts in four-dimensional CT images of 50 abdominal and thoracic radiotherapy patients. Int J Radiat Oncol Biol Phys 2008;72:1250–8.

[47] Faught AM, Yamamoto T, Castillo R, Castillo E, Zhang J, Miften M, et al. Evaluating which dose-function metrics are most critical for functional-guided radiotherapy with CT ventilation imaging. Int J Radiat Oncol Biol Phys 2017.

[48] Wang JM, Robertson SH, Wang Z, He M, Virgincar RS, Schrank GM, et al. Using hyperpolarized 129Xe MRI to quantify regional gas transfer in idiopathic pulmonary fibrosis. Thorax 2018;73:21–8.

[49] Waxweiler T, Schubert L, Diot Q, Faught A, Stuhr K, Castillo R, et al. A complete 4DCT-ventilation functional avoidance virtual trial: Developing strategies for prospective clinical trials. J Appl Clin Med Phys 2017;18:144–52.

[50] Kipritidis J. TU-H-202-04: The VAMPIRE Challenge: Preliminary Results From a Multi-Institutional Study of CT Ventilation Image Accuracy. Med Phys 2016;43:3771–3771.