Published in final edited form as:

Acad Radiol. 2022 February ; 29(Suppl 2): S145–S155. doi:10.1016/j.acra.2021.06.017.

Hyperpolarized ^{129}Xenon MRI Ventilation Defect Quantification via Thresholding and Linear Binning in Multiple Pulmonary Diseases

David J Roach[1], **Matthew M. Willmering**[1], **Joseph W. Plummer**[1], **Laura L. Walkup**[1,2], **Yin Zhang**[3], **Md Monir Hossain**[2,3], **Zackary I. Cleveland**[1,2], **Jason C. Woods**[1,2]

[1]Center for Pulmonary Imaging Research, Division of Pulmonary Medicine and Department of Radiology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio

[2]Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio

[3]Department of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio

Abstract

Rationale: There is no agreed upon method for quantifying ventilation defect percentage (VDP) with high sensitivity and specificity from hyperpolarized (HP) gas ventilation MR images in multiple pulmonary diseases for both pediatrics and adults, yet identifying such methods will be necessary for future multi-site trials. Most HP gas MRI ventilation research focuses on a specific pulmonary disease and utilizes one quantification scheme for determining VDP. Here we sought to determine the potential of different methods for quantifying VDP from HP ^{129}Xe images in multiple pulmonary diseases through comparison of the most utilized quantification schemes: linear binning and thresholding.

Materials and Methods: HP ^{129}Xe MRI was performed in a total of 176 pediatric (n=125) and adult (51) subjects (age 20.98±16.48 years) who were either healthy controls (n=23) or clinically diagnosed with cystic fibrosis (CF) (n=37), lymphangioleiomyomatosis (LAM) (n=29), asthma (n=22), systemic juvenile idiopathic arthritis (sJIA) (n=11), interstitial lung disease (ILD) (n=7), or were bone marrow transplant (BMT) recipients (n=47). HP ^{129}Xe ventilation images were acquired during a ≤16 second breath-hold using a 2D multi-slice gradient echo sequence on a 3T Philips scanner (TR/TE 8.0/4.0ms, FA 10–12°, FOV 300×300mm, voxel size≈3×3×15mm). Images were analyzed using four different methods to quantify VDPs: linear binning (histogram normalization with binning into 6 clusters) following either linear or a variant of a nonparametric nonuniform intensity normalization algorithm (N4ITK) rf-bias correction, thresholding ≤60% of the mean signal intensity with linear rf-bias-correction, and thresholding ≤60% and ≤75% of the mean signal intensity following N4ITK rf-bias-correction. Spirometry was successfully obtained in 84% of subjects.

Results: All quantification schemes were able to label visually identifiable ventilation defects in similar regions within all subjects. The VDPs of control subjects were significantly lower

*Correspondence to: J.C. Woods, Center for Pulmonary Imaging Research, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave, Cincinnati, OH 45229, USA. Jason.Woods@cchmc.org.

(P<0.05) compared to BMT, CF, LAM, and ILD subjects for most of the quantification methods. No one quantification scheme was better able to differentiate individual disease groups from the control group. Advanced statistical modeling of the VDP quantification schemes revealed that in comparing controls to the combined disease groups, N4ITK-biased corrected 60% thresholding had the highest predictive efficacy, sensitivity, and specificity at the VDP cut-point of 2.3%. However, compared to the thresholding quantification schemes, linear binning was able to capture and label subtle low-ventilation regions in subjects with milder obstruction, such as subjects with asthma.

Conclusions: The difference in VDP between healthy controls and patients varied between the different disease states for all quantification methods. Although N4ITK bias corrected 60% thresholding was superior in separating the combined diseased group from controls, linear binning is able to better label low-ventilation regions unlike the current, 60% thresholding scheme. For future clinical trials, a consensus will need to be reached on which VDP scheme to utilize, as there are subtle advantages for each for specific disease.

Keywords

Hyperpolarized ^{129}Xe MRI; ventilation defect percentage; linear binning; thresholding

Introduction:

Major achievements have been made in hyperpolarized (HP) noble (3He and 129Xe) gas magnetic resonance imaging (MRI) over the course of 25 years, which has now expanded to tens of sites world-wide (1,2). As such the quantification of regional obstruction of ventilation (so called “ventilation defects”) has evolved over time, with each site developing very similar and reasonable, but non-equivalent, quantification schemes. Initially, HP noble gas ventilation MRI relied on reader scoring to detect and quantify ventilation defects in subjects with cystic fibrosis and then asthma (3–7). Subsequently, diffusion HP gas MRI has been utilized within the lungs to measure regional airspace size, which provides additional lung structure information, particularly useful in quantifying enlarged airspaces caused by emphysema or alveolar simplification (8,9). Furthermore, Xe possesses the unique property of being slightly soluble in lung tissue and blood, therefore, allowing for regional mapping of gas exchange from the alveolar airspace to the capillary bed (10,11).

While more straightforward and routine than restricted diffusion or gas-exchange MRI, ^{129}Xe ventilation MRI has demonstrated direct clinical and translational relevance and very high sensitivity to early or mild obstruction. By combining HP gas and anatomical proton lung MR imaging, studies were able to quantify global lung ventilation using the percent lung-ventilated volume. This quantification scheme initially showed strong correlations to the reader scoring method, and further evolved into quantifying ventilation defect volume or ventilation defect percentage (VDP), with clear differences between chronic obstructive pulmonary disease, asthma, cystic fibrosis, or radiation induced lung injury and healthy subjects.(12–21).

From these foundations, more sophisticated and semi-automated methods for quantifying ventilation, with less human bias, have been implemented (22,23). The K-means clustering

algorithm was developed whereby the histogram intensity distribution was partitioned into four clusters based on an expert radiologist's interpretation of the clinical meaning of intensity differences. In short, K-means performs an initial histogram analysis to classify pixels into 4 clusters, which include no signal and hypointense pixels (cluster 1) to hyperintense pixels (cluster 4). Next an additional round of clustering is performed on cluster 1, representing signal voids and hypointense regions, to determine the ventilation defects from low-ventilated regions (22). This method has been successfully applied to asthma, chronic obstructive pulmonary disease (COPD), and cystic fibrosis (CF) subjects (22–29)

A "linear binning" scheme has been used recently, whereby the ventilation histogram is initially rescaled from 0 to 1 by using the 99^{th} percentile of the signal intensity distribution and then assigning pixels into different bins using predetermined thresholds (30–33). These thresholds are then used to classify pixels as having ventilation that is zero, low, medium, and high. This technique was further refined through the utilization of healthy controls to determine binning thresholds based on the control group's mean and standard deviation from their intensity distribution in HP gas ventilation MRI. As with K-means clustering, this method classifies pixels into different bins or clusters and has been successfully applied in healthy controls and subjects with either COPD or asthma (30–33). Furthermore, a comparison study of linear binning and K-means clustering demonstrated there was a significant correlation between these two quantification schemes in determining VDP in both healthy controls and asthmatics (33).

Our group has routinely applied a simpler VDP quantification scheme via mean signal intensity thresholding, where an individual's mean signal intensity is utilized in determining which pixels are classified into 3 categories: either ventilation defect, normal ventilation, or hyperventilation. A threshold cutoff for ventilation defects was initially empirically determined by calculating the maximum difference in VDP between adult asthma subjects and healthy controls (17). This maximum difference between controls and patients with lung disease was also found to be at a threshold near 60% mean whole lung signal when applied to pediatrics and adults with other pulmonary diseases such as asthma, CF, and LAM (17,34–36). This technique has been successfully implemented at two separate sites utilizing the same subject data, demonstrating very high inter-site reproducibility in analysis (37).

Overall, individual sites across the HP gas MRI community utilize one of these major ventilation defect quantification schemes (histogram linear binning, K-means clustering, or mean signal intensity thresholding), yet each site tends focus on a select few pulmonary diseases and/or age groups. As such, there is no yet agreed upon method used for quantifying VDP from HP noble gas ventilation MR images across different pulmonary diseases in both pediatric and adult subjects. While differences in analysis are subtle, identifying which VDP method provides both high sensitivity and specificity for an individual pulmonary disease is necessary for future multi-site trials and the eventual transition into clinical practice. In an attempt to provide additional objective data for future consensus within the community, here we compare the performance of linear binning and mean signal thresholding using two different B_1 inhomogeneity bias correction techniques, linear and a variant of a nonparametric nonuniform intensity normalization algorithm

(N4ITK) bias-field correction (17,38). We sought to determine the optimum method for quantifying VDP from HP ^{129}Xe images of pediatric and adult subjects over a range of pulmonary diseases.

Methods:

This study was approved by our institutional review board (IRB approval number 2014–5279) following United States Food and Drug Administration investigational new drug approval (IND 123577). HP ^{129}Xe MRI was performed in a total of 176 subjects from 02/23/2015 to 08/15/2019 (125 pediatrics and 51 adults), which are summarized in Table 1. Subjects included healthy controls (n=23), patients with cystic fibrosis (CF) (n=37), lymphangioleiomyomatosis (LAM) (n=29), asthma (n=22), systemic juvenile idiopathic arthritis (sJIA) (n=11), interstitial lung disease (ILD) (n=7) (either diagnosed with interstitial pulmonary fibrosis or Hermansky-Pudlak syndrome), or were a bone marrow transplant (BMT) recipient (n=47). Spirometry, percent predicted forced expiratory volume in one second (FEV_1%) and percent predicted forced vital capacity (FVC), were clinically acquired within one year of the scan or were acquired same day on most subjects prior to imaging according to ATS/ERS guidelines.

Due to the variance in lung size from pediatrics to adults, one of two home-built, single-channel, saddle coils tuned to 35.3 MHz (^{129}Xe frequency at 3T), were used to acquire all ^{129}Xe MR images (39). Hyperpolarized ^{129}Xe MRI gas (85% ^{129}Xe-enriched, Linde, Gas, Stewartsville, NJ) was prepared using a commercial polarizer (Polarean 9810 or 9820, Polarean Imaging PLC, Durham, NC) with polarization ranging from approximately 10% to 40%. The range in polarization is due to upgrades in commercial polarizer equipment along with technical advancements achieved over the 4 plus years of the subjects scanned within our study. In accordance with our institutional review board protocols and FDA Investigational New Drug application (IND 123577), the volume of ^{129}Xe accumulated and delivered to a research participant was equivalent to one-sixth of a subject's predicted total lung capacity, calculated using predictive equations, taking into account the subject's sex and height. Dosing volumes in our center typically range from 200mL (small child) to 1000mL (adult).(40) Subjects were coached to exhale to functional residual capacity prior to inhaling the HP ^{129}Xe gas, which was delivered to subjects in a Tedlar bag (Jensen Inert Products, Coral Springs, FL). Heart rate and SpO_2 were monitored throughout imaging protocol by a medical professional. After conventional ^{1}H locator scans, a small (~200 mL) calibration dose of hyperpolarized 129Xe gas was administered via a breath-hold maneuver to optimize the ventilation acquisition flip angle. HP ^{129}Xe ventilation images were acquired in all subjects during a ≤16 second breath-hold using a 2D multi-slice gradient echo sequence on a 3T Philips Achieva scanner (repetition time 8.0ms, echo time 4.0ms, flip angle 10–12°, field of view 300 × 300mm, voxel size ≈ 3 × 3 × 15mm, number of slices 6 to 16). During the same imaging session, structural lung images were acquired via a radial 3D ^{1}H ultrashort echo-time (UTE), stack of stars, MRI sequence. Images were acquired with the following parameters: repetition time 4.7ms, echo time 0.2 ms, field of view 300×300 to 400×400 mm^2, voxel size 1.39×1.39×4.00 mm^3. An echo navigator was positioned on the lung-liver interface to gate the acquisition and capture lung signal at functional residual capacity.

^{129}Xe ventilation images were manually segmented using ^{1}H UTE images as guides, processed using custom software in MATLAB (Mathworks). Additionally, signal to noise ratio (SNR) was quantified in these same images. In short, a background noise mask was generated by dilating the lung mask using a square structural element of 20×20 or 25×25, which removed trachea voxels and partial volume voxels around the boundary of the lung mask, then the complement image set was acquired. The mean signal with the generated lung mask and standard deviation of the signal within the background noise mask was used to determine the SNR slice-by-slice and averaged to obtain a single SNR for each subject.

Due to coil B_1 inhomogeneity, the ventilation images were subject to low frequency imaging artifacts, which can confound ventilation defect analysis. Therefore, to correct for this artifact, prior to VDP analysis, linear or N4ITK bias-field corrections were separately applied to the same set of images, to allow for comparisons between the two bias correction techniques (38). The linear RF bias correction was determined for each subject by dividing the intensity image set by the mean signal intensity value in each dimension. This linear correction reduces the RF bias in the intensity image while preserving subtle and heterogenous ventilation defects. N4ITK bias-field correction in Advanced Normalization Tools (http://picsl.upenn.edu/software/ants/) is a nonparametric nonuniform intensity normalization method, which was introduced by Tustison et al and applied to HP ^{3}He MR images (38). The main advantage of N4ITK bias correction is that it utilizes a B-spline approximation to spatially smooth and the slowly varying intensity bias across the image, which is iteratively applied to correct the image. The parameters of the N4ITK-bias correction were: shrink factor 1, iterations of 100, 50, 50, 50, tolerance of 1×10^{-10}, and spline parameter of 200. This method is used more frequently in the hyperpolarized-gas community, but has the potential of changing the intensity of ventilation defects.

Thresholding VDP Quantification

The ventilation defect percentage (VDP) for all disease groups was thus quantified using the mean optimal threshold for all pulmonary diseases(<60% of the mean whole-lung ^{129}Xe signal) on images with linear bias-correction, which was previously determined for similar groups (17,34–37). For N4ITK bias corrected images, an optimal threshold of <75% of the mean whole-lung ^{129}Xe signal was utilized as it was determined empirically to be the maximum difference between controls and all disease groups combined. Additionally, for N4ITK bias corrected images, VDPs were quantified using a threshold <60% mean whole-lung ^{129}Xe signal, similar to the linear bias correction method for direct comparison. As per our VDP analysis scheme, a median filter (3×3 kernel) was applied to the defect mask array to eliminate ”hot” pixels and small vessels, which can alter the quantification of ventilation defects.

Linear Binning VDP Quantification

For the linear binning method, signal from the airways was retained to equalize and rescale the histograms from 0 to 1 using the 99th percentile of the cumulative distribution for each image set. After histogram rescaling the airway signal was excluded from the quantitative analysis. Twenty-three healthy controls with SNR > 8 and no visible ventilation defects

were utilized to establish an unbiased reference distribution of signal intensity. The mean and standard deviation of the averaged rescaled intensity histogram from these controls was used to define the threshold values of 6 bin-map boundaries. These bins identified ventilation defects, low ventilation, two normal ventilation, and two high ventilation regions. For images with linear bias correction the mean and standard deviation (SD) was determined to be 0.54±0.15 while for the same images with N4ITK bias field correction was calculated to be 0.68±0.14. The boundary between bins 3 and 4 was defined as the mean of this distribution with the width of each bin assigned to1 SD and the bin with the lowest intensity (mean – 2SD) was defined as VDP (30–33).

Statistics

Ventilation defect percentage means and SDs were calculated for each group. Two-sided t-tests were applied to comparisons between controls and the combined disease group data in addition to demographics difference between the individual disease groups and controls. For VDP and spirometry analysis individual groups, a Tukey-Kramer post-hoc analysis was utilized to compare controls and the different pulmonary disease groups, where a P value ≤ 0.05 was considered significant. The associations between diseases and each VDP measure were assessed using univariate logistic regression models. For each model, we plotted the corresponding VDP scores using ROC curves and obtained an optimal cutoff point based on the minimum distance between sensitivity and specificity of the VDP scores in the ROC curve. Here sensitivity is defined as the true positive rate by the proportion of the disease group as being correctly identified as non-healthy, and specificity is defined as the true negative rate by the proportion of healthy controls that are correctly identified as healthy. Sensitivity analyses were conducted by using two additional cut points that were ± 0.1 different from the optimal one. C-statistics were used to evaluate different VDP analyses. Bland-Altman plots were generated to compare threshold mean signal and linear binning quantification schemes.

Results:

Table 1 lists the demographics for each patient group. Table 2 lists spirometry, SNR, and the mean, standard deviation, C-statistic, sensitivity, and specificity of the calculated VDPs utilizing the five different quantification methods in addition to spirometry results. Compared to controls, the percent predicted FEV_1% was significantly lower in BMT and the combined disease group. Percent predicted FVC% was significantly lower in BMT and ILD groups when compared to controls. The ratio of FEV_1/FVC was significantly lower in the LAM subjects and the combined disease group when compared to controls. The SNR varied between the individual groups with the average SNR of all groups combined being 14.1±9.1. When compared to the controls, the SNR of the individual disease groups and the combined disease group were not significantly greater.

In the thresholding quantification scheme, a range of thresholds from 5% to 100% of the mean whole-lung Xe signal, in increments of 5%, were utilized to determine the optimal cut-off in maximizing the average VDP difference between healthy controls and subjects with pulmonary disease. There were disease-specific differences in the optimal threshold

percentages for linear bias corrected images: 60% for BMT, 50% for Asthma, 70% for CF, 70% for LAM, 40% for sJIA, and 70% for ILD. The optimal thresholds for N4ITK bias corrected images were: 80% for BMT, 75% for Asthma, 80% for CF, 80% for LAM, 60% for sJIA, and 80% for ILD. However, for the combined disease group the optimal threshold percentage for linear bias corrected images was determined to be 60% while for NITK bias corrected images it was determined to be 75%.

The VDPs of control subjects were significantly lower compared to BMT, CF, LAM, ILD subjects and the combined disease group for some or all quantification methods ($P<0.05$). However, the mean VDPs of the asthma and sJIA groups were not significantly greater than the controls ($P>0.05$); this is also reflected in the nonsignificant difference in spirometry compared to controls. Furthermore, there were variations in the average VDP measures between the individual disease groups, indicating a wide range of ventilation impairment. In particular, compared to CF subjects, the VDPs of the asthma subjects were significantly lower ($P<0.001$) for each quantification scheme while FEV_1% ($P>0.12$) and FVC% ($P>0.79$) were not. Additionally, for all subjects with spirometry, all VDP quantification schemes displayed moderated and signification correlations to FEV_1% ($r=0.51$ to 0.58, $P<0.0001$) but there was no correlation between VDP and FVC%.

Importantly, no single quantification scheme was better at differentiating all of the individual disease groups from controls. However, Figure 1 illustrates this in part, with ventilation defect maps generated from the quantification schemes and illustrates how well each technique identifies similar ventilation defect regions, though with subtle differences, within the same slice of three different subjects (control, ILD, CF, and LAM). The visible defects in the control images are due to vascular artifacts not removed by the median filter technique. It also shows how VDPs vary for the different quantification schemes and how the two bias correction schemes alter VDPs. Figure 2 displays the threshold VDP quantification method plotted against linear binning VDP method with a strong and significant correlation between both schemes with linear and N4ITK bias-field corrections ($R^2>0.88$,$P<0.001$). Bland-Altman plots, displayed in Figure 3, compare the two quantification schemes with the different bias-field corrections. There is evidence of large differences in the VDPs of several LAM, ILD, and CF subjects as measured via linear binning vs thresholding quantification schemes (−31.9% to 47.8%). The effect of linear and N4ITK bias-field correction on the different VDP quantification schemes is illustrated in Figure 4, which displays a single slice of a severe asthma subject's ^{129}Xe ventilation MR images. This demonstrates how the two bias corrections differ in removing or reducing the differences in regions of minor defects. Additionally, Figure 4 demonstrates how linear binning is more sensitive in identifying subtle ventilation defects compared to thresholding. The effect of the median filter on a control and CF subject ventilation maps, quantified via 60% threshold of the mean signal, are displayed in Figure 5. The red ovals within the Figure 5 highlight the isolated pixels removed by the median filter and how this affects the overall VDP for each subject (Control VDP goes from 1.0% to 3.0% and CF VDP goes from 27.8% to 32.3%). Overall, the statistical modeling of the VDP quantification schemes revealed that N4ITK 60% thresholding had the highest overall predictive efficacy, sensitivity, and specificity at differentiating the combined disease group from healthy controls.

Discussion:

With the HP gas MR imaging community expanding its imaging studies to a wider range of ages and diseases, we sought to understand which quantification scheme provides a robust and accurate method in determining impaired ventilation regions. Here the differences in VDP between healthy controls and diseased patients varied for all of the quantification methods, as anticipated over this range of ages and pulmonary diseases. A recent study found that regional gas distribution and VDP were dependent on a combination of gas mixture composition and or SNR. (41) However, in our study the gas mixtures were consistent for each subject and there was no statistically significant difference in SNR between controls and individual disease groups. The more advanced statistical modeling determined that N4ITK-biased 60% thresholding had the highest sensitivity and specificity of all quantification schemes in separating the combined group of all diseased subjects from healthy controls, but no quantification scheme was universally the most sensitive or specific in separating individual disease groups from controls. Additionally, the sources of ventilation defects caused by either airway obstruction, small airway disease, mucus plugging, air trapping, and or local atelectasis vary between the different disease states studied here. Therefore, a particular VDP quantification method with a high sensitivity and specificity in differentiating a specific disease type from healthy control may be better suited and more applicable than a single universal VDP quantification approach.

For each quantification scheme, there are variations in VDP measurements between the individual disease groups with similar ranges of spirometrically measured lung function. These differences are most notable between the closely aged asthma and CF subjects, whose lung function measurements were not significantly different. Although the asthma subjects studied here ranged from mild to severe, the average VDP was significantly lower than that of the CF subjects for each quantification scheme. Furthermore, no quantification scheme was better able at differentiating either asthma or sJIA subjects from controls, likely due to the majority of these subjects having mild disease as reflected in their spirometry results. Long breath-holds and multibreath ^{129}Xe MR imaging is likely more sensitive to these disease states with delayed ventilation(42–45). Recent studies have demonstrated that subtle defects in asthma may still be better quantified by reader scoring, though we anticipate that artificial-intelligence algorithms will capture these more subtle defects in the future (46).

Not explicitly demonstrated here are the comparisons between different sub-classifications of ventilation between the quantification schemes. The thresholding quantification scheme traditionally identifies ventilation defects, normal ventilation, and hyper-ventilated regions, but no specific “low ventilation” regions like that of linear binning. Because of these differences we have focused on each technique’s ability to identify and label ventilation defect regions over a range of ages and pulmonary diseases. Both linear binning and mean signal thresholding quantification schemes perform similarly at identifying ventilation defects, and both can be adjusted in the future to similar sub-classifications of ventilation. For the thresholding quantification scheme, the mean intensity signal of the whole lung for each individual subject is used to establish ventilation classification cutoffs via an objective, empirically determined threshold. This ventilation defect threshold is determined via the maximum difference between healthy control VDPs and the disease group VDPs, and our

findings revealed that the maximum difference varied between all the diseased groups studied here. As such the VDP threshold cutoff for one disease population may be different from another and may differ between mild and severe obstructive diseases. Future studies identifying specific VDP thresholds or approaches for various diseases with differentiation in disease severity will need to be addressed and agreed upon.

For linear binning, a healthy control cohort was utilized to define the histogram binning thresholds for determining ventilation classification categories. As the HP gas signal intensity distribution is not static with age, there will be a need for guidelines in determining appropriate matched aged controls to accurately quantify VDPs over a range of ages. This is evident in part in the Bland-Altman plots, where compared to the controls the significantly older LAM (49.18±11.62,$P<0.05$) and ILD (48.90±21.83,$P<0.05$) subjects lay outside the 95% limits of agreement. To correct for this bias appropriate control groups will need to be acquired in order to more accurately determine VDPs via linear binning in a specific grouping of ages. Unlike the threshold quantification scheme, linear binning classifies a cluster in-between the ventilation defects and normal ventilation regions as low ventilation region, which can visibly label the subtle ventilation defects observed in Figure 4. Thus, linear binning has greater potential ability to differentiate those subjects with mild disease such as the asthma and sJIA groups from healthy controls. A similar method can be applied to the thresholding technique, however, an expert reader may be needed to intervene in selecting an appropriate threshold as there might not be a universal threshold for subtle ventilation defects labeling.

There is a wide range of image preprocessing steps prior to VDP quantification, one of the most utilized tools is B_1 inhomogeneity correction. Progress has been made in ameliorating B_1 inhomogeneity either by a simple linear bias-field correction scheme (18) or a more complex N4ITK bias-field correction scheme (38). Both bias-field corrections have their advantages and disadvantages, with a linear bias correction retaining subtle ventilation defects while not fully correcting for all the B_1 inhomogeneities. The N4ITK bias-field correction method removes most of the B_1 inhomogeneities but also tends to overly smooth and homogenize the intensity profile, potentially removing subtle ventilation defects. This can be problematic for subtle ventilation defects in subjects with early lung and small airway disease, thereby resulting in an underestimation of true VDP. N4ITK has not been thoroughly validated in functional lung imaging and that based on flip angle maps derived from 3D-radial acquisitions of ventilation image it over-corrects the bias field thereby removing physiological gradients (47). Further, “static” ventilation imaging really is slightly dynamic in that slow-filling regions increase with long breath-holds. In our study, breath-hold time was near constant for all subjects.

Small blood vessels observed as dark areas within the ventilation images and can be incorrectly labeled as defects and as such various techniques have been developed to eliminate this issue. A vesselness filter has been developed and utilized by different groups though with variations in the technique (23,30,33,48). Unlike He ventilation images with higher SNR and spatial resolution, Xe ventilation images require a co-registered proton image set of the thoracic cavity to correctly identity and label vasculature to be removed from the ventilation image (30,33). In our study, an inspiratory level matched thoracic

proton image set was not acquired, therefore, we utilized a median filter after manual lung and large vessel segmentation prior to VDP quantification. The original intention of the filter was to remove defect artifacts caused by vessels, pixel intensity hot spots, and single-voxel partial volumes at the lung mask boundaries. The filter reduces the number of isolated defect voxels thereby a providing a more reliable contrast between controls and subjects with pulmonary disease (18). However, the application of the median filter also tends to remove subtle ventilation defects leading to a potential artificial decrease in VDP in subjects with mild to moderate respiratory disease. Overall, an agreed upon approach for removing small blood vessels from ventilation images will reduce bias between sites and between studies.

In conclusion, we have demonstrated similarities and differences between linear binning and thresholding quantification schemes in calculating VDPs over a wide range of ages and pulmonary disease. With caveats discussed above, linear binning with N4ITK bias correction provides superior group differentiation in most of the disease groups studied here with the ability to quantify subtle low ventilation regions. We propose that most future clinical trials will benefit from N4ITK linear binning analysis for ventilation, though individual patient groups, such as mild asthma, may benefit from different methods. As such additional thresholds or other scoring may be needed to address specific ventilation classifications. Greater consensus on the optimum of method and image processing tools for individual pulmonary diseases and their severity will benefit future clinical trials and the eventual transition of hyperpolarized-gas MRI into routine clinical practice.

Acknowledgments

Funding: Part of this study was funded by R01 HL134801 (Z.I.C.) and R01 HL131012 (J.C.W.).

Abbreviations:

HP	hyperpolarized
Xe	xenon
MRI	magnetic resonance imaging
VDP	ventilation defect percentage
CF	cystic fibrosis
LAM	lymphangioleiomyomatosis
sJIA	systemic juvenile idiopathic arthritis
ILD	interstitial lung disease
BMT	bone marrow transplant recipient

References:

1. Translational applications of hyperpolarized 3He and 129Xe. Walkup LL, Woods JC. NMR Biomed. 2014 Dec;27(12):1429–38. doi: 10.1002/nbm.3151. [PubMed: 24953709]

2. Hyperpolarized and inert gas MRI: the future. Couch MJ, Blasiak B, Tomanek B, Ouriadov AV, Fox MS, Dowhos KM, Albert MS. Mol Imaging Biol. 2015 Apr;17(2):149–62. [PubMed: 25228404]

3. Cystic fibrosis: combined hyperpolarized 3He-enhanced and conventional proton MR imaging in the lung--preliminary observations. Donnelly LF, MacFall JR, McAdams HP, Majure JM, Smith J, Frush DP, Bogonad P, Charles HC, Ravin CE. Radiology. 1999 Sep;212(3):885–9. [PubMed: 10478261]

4. Imaging the lungs in asthmatic patients by using hyperpolarized helium-3 magnetic resonance: assessment of response to methacholine and exercise challenge. Samee S, Altes T, Powers P, de Lange EE, Knight-Scott J, Rakes G, Mugler JP 3rd, Ciambotti JM, Alford BA, Brookeman JR, Platts-Mills TA. J Allergy Clin Immunol. 2003 Jun;111(6):1205–11. [PubMed: 12789218]

5. Changes in regional airflow obstruction over time in the lungs of patients with asthma: evaluation with 3He MR imaging. de Lange EE, Altes TA, Patrie JT, Battiston JJ, sJuersivich AP, Mugler JP 3rd, Platts-Mills TA. Radiology. 2009 Feb;250(2):567–75.) [PubMed: 19188325]

6. Clinical correlates of lung ventilation defects in asthmatic children. Altes TA, Mugler JP 3rd, Ruppert K, Tustison NJ, Gersbach J, Szentpetery S, Meyer CH, de Lange EE, Teague WG. J Allergy Clin Immunol. 2016 Mar;137(3):789–96.e7. doi: 10.1016/j.jaci.2015.08.045. Epub 2015 Oct 29. [PubMed: 26521043]

7. Evaluation of structure-function relationships in asthma using multidetector CT and hyperpolarized He-3 MRI. Fain SB, Gonzalez-Fernandez G, Peterson ET, Evans MD, Sorkness RL, Jarjour NN, Busse WW, Kuhlman JE. Acad Radiol. 2008 Jun;15(6):753–62 [PubMed: 18486011]

8. Lung MRI as a Potential Complementary Diagnostic Tool for Early COPD.Polverino F, Hysinger EB, Gupta N, Willmering M, Olin T, Abman SH, Woods JC. Am J Med. 2020 Jun;133(6):757–760. doi: 10.1016/j.amjmed.2019.12.009. Epub 2020 Jan 16. [PubMed: 31954683]

9. Direct comparison of ^{129} Xe diffusion measurements with quantitative histology in human lungs. Thomen RP, Quirk JD, Roach D, Egan-Rojas T, Ruppert K, Yusen RD, Altes TA, Yablonskiy DA, Woods JC. Magn Reson Med. 2017 Jan;77(1):265–272. [PubMed: 26778748]

10. Imaging alveolar-capillary gas transfer using hyperpolarized 129Xe MRI. Driehuys B, Cofer GP, Pollaro J, Mackel JB, Hedlund LW, Johnson GA. Proc Natl Acad Sci U S A. 2006 Nov 28;103(48):18278–83. [PubMed: 17101964]

11. Hyperpolarized Xe MR imaging of alveolar gas uptake in humans. Cleveland ZI, Cofer GP, Metz G, Beaver D, Nouls J, Kaushik SS, Kraft M, Wolber J, Kelly KT, McAdams HP, Driehuys B. PLoS One. 2010 Aug 16;5(8):e12192. [PubMed: 20808950]

12. Combined helium-3/proton magnetic resonance imaging measurement of ventilated lung volumes in smokers compared to never-smokers. Woodhouse N, Wild JM, Paley MN, Fichele S, Said Z, Swift AJ, van Beek EJ.J Magn Reson Imaging. 2005 Apr;21(4):365–9. [PubMed: 15779032]

13. Hyperpolarized 3He magnetic resonance imaging of chronic obstructive pulmonary disease: reproducibility at 3.0 tesla. Mathew L, Evans A, Ouriadov A, Etemad-Rezai R, Fogel R, Santyr G, McCormack DG, Parraga G. Acad Radiol. 2008 Oct;15(10):1298–311. [PubMed: 18790402]

14. Hyperpolarized 3He magnetic resonance imaging of ventilation defects in healthy elderly volunteers: initial findings at 3.0 Tesla. Parraga G, Mathew L, Etemad-Rezai R, McCormack DG, Santyr GE. Acad Radiol. 2008 Jun;15(6):776–85. [PubMed: 18486013]

15. Quantitative analysis of hyperpolarized 129Xe ventilation imaging in healthy volunteers and subjects with chronic obstructive pulmonary disease. Virgincar RS, Cleveland ZI, Kaushik SS, Freeman MS, Nouls J, Cofer GP, Martinez-Jimenez S, He M, Kraft M, Wolber J, McAdams HP, Driehuys B. NMR Biomed. 2013 Apr;26(4):424–35. [PubMed: 23065808]

16. Detection of longitudinal lung structural and functional changes after diagnosis of radiation-induced lung injury using hyperpolarized 3He magnetic resonance imaging. Mathew L, Gaede S, Wheatley A, Etemad-Rezai R, Rodrigues GB, Parraga G. Med Phys. 2010 Jan;37(1):22–31. [PubMed: 20175462]

17. Regional ventilation changes in severe asthma after bronchial thermoplasty with (3)He MR imaging and CT. Thomen RP, Sheshadri A, Quirk JD, Kozlowski J, Ellison HD, Szczesniak RD, Castro M, Woods JC. Radiology. 2015 Jan;274(1):250–9. [PubMed: 25144646]

18. Probing Changes in Lung Physiology in COPD Using CT, Perfusion MRI, and Hyperpolarized Xenon-129 MRI. Qing K, Tustison NJ, Mugler JP 3rd, Mata JF, Lin Z, Zhao L, Wang D, Feng X,

Shin JY, Callahan SJ, Bergman MP, Ruppert K, Altes TA, Cassani JM, Shim YM. Acad Radiol. 2019 Mar;26(3):326–334. [PubMed: 30087065]

19. Mapping of regional lung microstructural parameters using hyperpolarized (129) Xe dissolved-phase MRI in healthy volunteers and patients with chronic obstructive pulmonary disease.Kern AL, Gutberlet M, Voskrebenzev A, Klimeš F, Rotärmel A, Wacker F, Hohlfeld JM, Vogel-Claussen J. Magn Reson Med. 2019 Apr;81(4):2360–2373. [PubMed: 30362620]
20. Hyperpolarized Gas Magnetic Resonance Imaging of Pediatric Cystic Fibrosis Lung Disease. Santyr G, Kanhere N, Morgado F, Rayment JH, Ratjen F, Couch MJ. Acad Radiol. 2019 Mar;26(3):344–354 [PubMed: 30087066]
21. Assessment of the influence of lung inflation state on the quantitative parameters derived from hyperpolarized gas lung ventilation MRI in healthy volunteers. Hughes PJC, Smith L, Chan HF, Tahir BA, Norquay G, Collier GJ, Biancardi A, Marshall H, Wild JM. J Appl Physiol (1985). 2019 Jan 1;126(1):183–192 [PubMed: 30412033]
22. Hyperpolarized 3He magnetic resonance functional imaging semiautomated segmentation. Kirby M, Heydarian M, Svenningsen S, Wheatley A, McCormack DG, Etemad-Rezai R, Parraga G. Acad Radiol. 2012 Feb;19(2):141–52. [PubMed: 22104288]
23. Ventilation-based segmentation of the lungs using hyperpolarized (3)He MRI. sTustison NJ, Avants BB, Flors L, Altes TA, de Lange EE, Mugler JP 3rd, Gee JC. J Magn Reson Imaging. 2011 Oct;34(4):831–41 [PubMed: 21837781]
24. Hyperpolarized 3He and 129Xe MR imaging in healthy volunteers and patients with chronic obstructive pulmonary disease. Kirby M, Svenningsen S, Owrangi A, Wheatley A, Farag A, Ouriadov A, Santyr GE, Etemad-Rezai R, Coxson HO, McCormack DG, Parraga G. Radiology. 2012 Nov;265(2):600–10 [PubMed: 22952383]
25. Regional pulmonary response to a methacholine challenge using hyperpolarized (3)He magnetic resonance imaging. Costella S, Kirby M, Maksym GN, McCormack DG, Paterson NA, Parraga G. Respirology. 2012 Nov;17(8):1237–46. [PubMed: 22889229]
26. Hyperpolarized 3He magnetic resonance functional imaging semiautomated segmentation. Kirby M, Heydarian M, Svenningsen S, Wheatley A, McCormack DG, Etemad-Rezai R, Parraga G. Acad Radiol. 2012 Feb;19(2):141–52. [PubMed: 22104288]
27. Redistribution of inhaled hyperpolarized 3He gas during breath-hold differs by asthma severity. Hahn AD, Cadman RV, Sorkness RL, Jarjour NN, Nagle SK, Fain SB.J Appl Physiol (1985). 2016 Mar 1;120(5):526–36. doi: [PubMed: 26635346]
28. Clinical correlates of lung ventilation defects in asthmatic children. Altes TA, Mugler JP 3rd, Ruppert K, Tustison NJ, Gersbach J, Szentpetery S, Meyer CH, de Lange EE, Teague WG. J Allergy Clin Immunol. 2016 Mar;137(3):789–96.e7. [PubMed: 26521043]
29. Comparison of Functional Free-Breathing Pulmonary 1H and Hyperpolarized ^{129}Xe Magnetic Resonance Imaging in Pediatric Cystic Fibrosis. Couch MJ, Munidasa S, Rayment JH, Voskrebenzev A, Seethamraju RT, Vogel-Claussen J, Ratjen F, Santyr G.Acad Radiol. 2020 Jun 10:S1076–6332(20)30284–1. doi: 10.1016/j.acra.2020.05.008.
30. Extending semiautomatic ventilation defect analysis for hyperpolarized (129)Xe ventilation MRI. He M, Kaushik SS, Robertson SH, Freeman MS, Virgincar RS, McAdams HP, Driehuys B. Acad Radiol. 2014 Dec;21(12):1530–41. [PubMed: 25262951]
31. Using Hyperpolarized $_{129}$Xe MRI to Quantify the Pulmonary Ventilation Distribution. He M, Driehuys B, Que LG, Huang YT. Acad Radiol. 2016 Dec;23(12):1521–1531. [PubMed: 27617823]
32. Hyperpolarized 129Xenon Magnetic Resonance Imaging to Quantify Regional Ventilation Differences in Mild to Moderate Asthma: A Prospective Comparison Between Semiautomated Ventilation Defect Percentage Calculation and Pulmonary Function Tests. Ebner L, He M, Virgincar RS, Heacock T, Kaushik SS, Freemann MS, McAdams HP, Kraft M, Driehuys B. Invest Radiol. 2017 Feb;52(2):120–127. [PubMed: 27662575]
33. A Comparison of Two Hyperpolarized $_{129}$Xe MRI Ventilation Quantification Pipelines: The Effect of Signal to Noise Ratio. He M, Zha W, Tan F, Rankine L, Fain S, Driehuys B. Acad Radiol. 2019 Jul;26(7):949–959. [PubMed: 30269957]

34. Hyperpolarized ^{129}Xe for investigation of mild cystic fibrosis lung disease in pediatric patients. Thomen RP, Walkup LL, Roach DJ, Cleveland ZI, Clancy JP, Woods JC. J Cyst Fibros. 2017 Mar;16(2):275–282. [PubMed: 27477942]
35. Cyst Ventilation Heterogeneity and Alveolar Airspace Dilation as Early Disease Markers in Lymphangioleiomyomatosis. Walkup LL, Roach DJ, Hall CS, Gupta N, Thomen RP, Cleveland ZI, McCormack FX, Woods JC. Ann Am Thorac Soc. 2019 Aug;16(8):1008–1016. [PubMed: 31038987]
36. Xenon-129 MRI detects ventilation deficits in paediatric stem cell transplant patients unable to perform spirometry.Walkup LL, Myers K, El-Bietar J, Nelson A, Willmering MM, Grimley M, Davies SM, Towe C, Woods JC.Eur Respir J. 2019 May 2;53(5).
37. A two-center analysis of hyperpolarized $_{129}$Xe lung MRI in stable pediatric cystic fibrosis: Potential as a biomarker for multi-site trials. Couch MJ, Thomen R, Kanhere N, Hu R, Ratjen F, Woods J, Santyr G. J Cyst Fibros. 2019 Sep;18(5):728–733. [PubMed: 30922812]
38. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 2010 Jun;29(6):1310–20. [PubMed: 20378467]
39. Loew W, Thomen R, Pratt R, Cleveland ZI, Dumoulin C, Woods JC, Giaquinto R. A Volume Saddle Coil for Hyperpolarized 129Xe Lung Imaging. Proc Intl Soc Mag Reson Med. 2015;23:1507.
40. Stocks J and Quanjer PH, Reference values for residual volume, functional residual capacity and total lung capacity: ATS Workshop on Lung Volume Measurements Official Statement of the European Respiratory Society. European Respiratory Journal, 1995. 8: p. 492–506.
41. Reproducibility of Hyperpolarized ^{129}Xe MRI Ventilation Defect Percent in Severe Asthma to Evaluate Clinical Trial Feasibility. Svenningsen S, McIntosh M, Ouriadov A, Matheson AM, Konyer NB, Eddy RL, McCormack DG, Noseworthy MD, Nair P, Parraga G. Acad Radiol. 2020 May 14:S1076–6332(20)30237–3. doi: 10.1016/j.acra.2020.04.025.
42. (3)He pO2 mapping is limited by delayed-ventilation and diffusion in chronic obstructive pulmonary disease. Marshall H, Parra-Robles J, Deppe MH, Lipson DA, Lawson R, Wild JM.Magn Reason Med. 2014 Mar;71(3):1172–8. doi: 10.1002/mrm.24779.
43. A Comparison of Hyperpolarized Gas Imaging of Aeration and Fractional Ventilation. Hamedani H, Bermudez F, Baron R, Kadlecek S, Ruppert K, Xin Y, Siddiqui S, Pourfathi M, Amzajerdian F, Loza L, Achekzai T, Duncan I, Cereda M, Sertic F, Rizi RR 10.1164/ajrccmconference.2019.199.1_MeetingAbstracts.A5772
44. Quantification of regional fractional ventilation in human subjects by measurement of hyperpolarized 3He washout with 2D and 3D MRI. Horn FC, Deppe MH, Marshall H, Parra-Robles J, Wild JM.J Appl Physiol (1985). 2014 Jan 15;116(2):129–39. doi: 10.1152/japplphysiol.00378.2013 [PubMed: 24311749]
45. Regional Fractional Ventilation by Using Multibreath Wash-in (3)He MR Imaging. Hamedani H, Clapp JT, Kadlecek SJ, Emami K, Ishii M, Gefter WB, Xin Y, Cereda M, Shaghaghi H, Siddiqui S, Rossman MD, Rizi RR.Radiology. 2016 Jun;279(3):917–24. doi: 10.1148/radiol.2015150495. [PubMed: 26785042]
46. (129)Xe MRI as a measure of clinical disease severity for pediatric asthma. Lin NY, Roach DJ, Willmering MM, Walkup LL, Hossain MM, Desirazu P, Cleveland ZI, Guilbert TW, Woods JC. J Allergy Clin Immunol. 2020 Nov 20:S0091–6749(20)31626–2. doi: 10.1016/j.jaci.2020.11.010. Online ahead of print.
47. Bias Field Correction in Hyperpolarized 129Xe Gas Ventilation MRI. ISMRM& SMRT 08–14 August 2020. Lu Junlan, Wang Ziyi, Nouls John C., Gulati Kush, Bier Elianna, Mummy David, and Driehuys Bastiaan.
48. Semiautomated Ventilation Defect Quantification in Exercise-induced Bronchoconstriction Using Hyperpolarized Helium-3 Magnetic Resonance Imaging: A Repeatability Study. Zha W, Niles DJ, Kruger SJ, Dardzinski BJ, Cadman RV, Mummy DG, Nagle SK, Fain SB. Acad Radiol. 2016 Sep;23(9):1104–14. doi: 10.1016/j.acra.2016.04.005. [PubMed: 27263987]

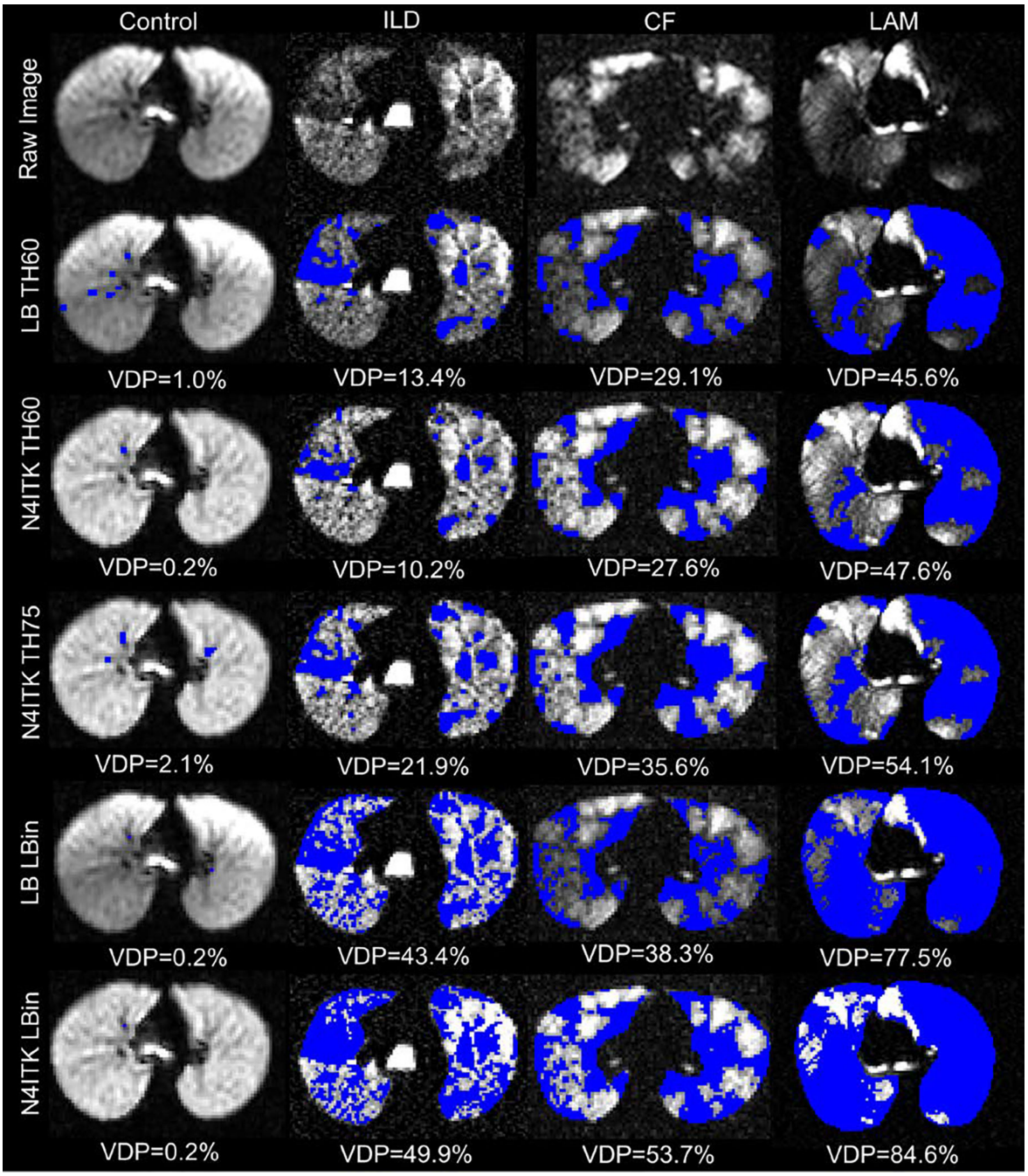

Figure 1.

Ventilation defects (blue) quantified in a control, ILD, CF, and LAM subjects via the different quantification schemes where LB = linear bias corrected, N4ITK = N4ITK-bias field corrected, TH60 is thresholding at 60% mean lung signal, TH75 is thresholding at 75% mean lung signal, and Lbin = Linear Binning.

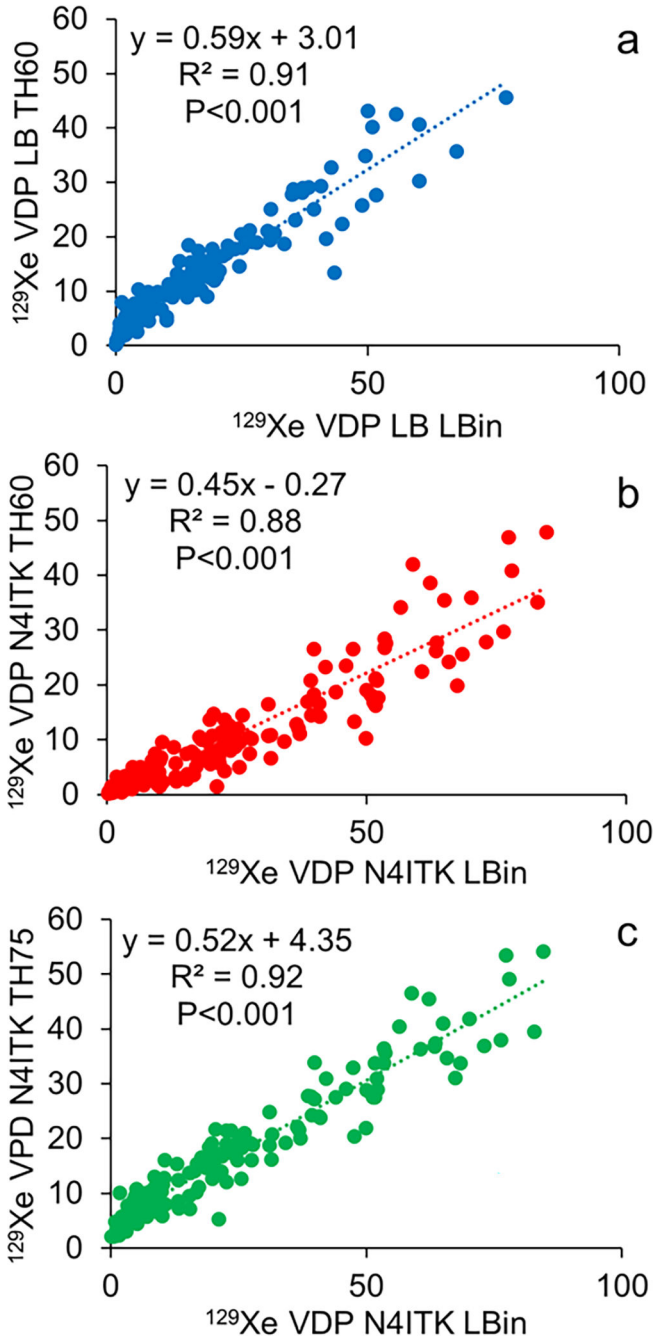

Figure 2.

a) ^{129}Xe ventilation defect percentage (VDP) quantified via linear bias corrected 60% threshold mean signal plotted against linear binning. b) ^{129}Xe VDP quantified via N4ITK-bias corrected 60% threshold mean signal plotted against linear binning. c) ^{129}Xe VDP quantified via N4ITK-bias corrected 75% threshold mean signal plotted against linear binning. LB = linear bias corrected, N4ITK = N4ITK-bias field corrected, TH60 is thresholding at 60% mean lung signal, TH75 is thresholding at 75% mean lung signal, and Lbin = Linear Binning.

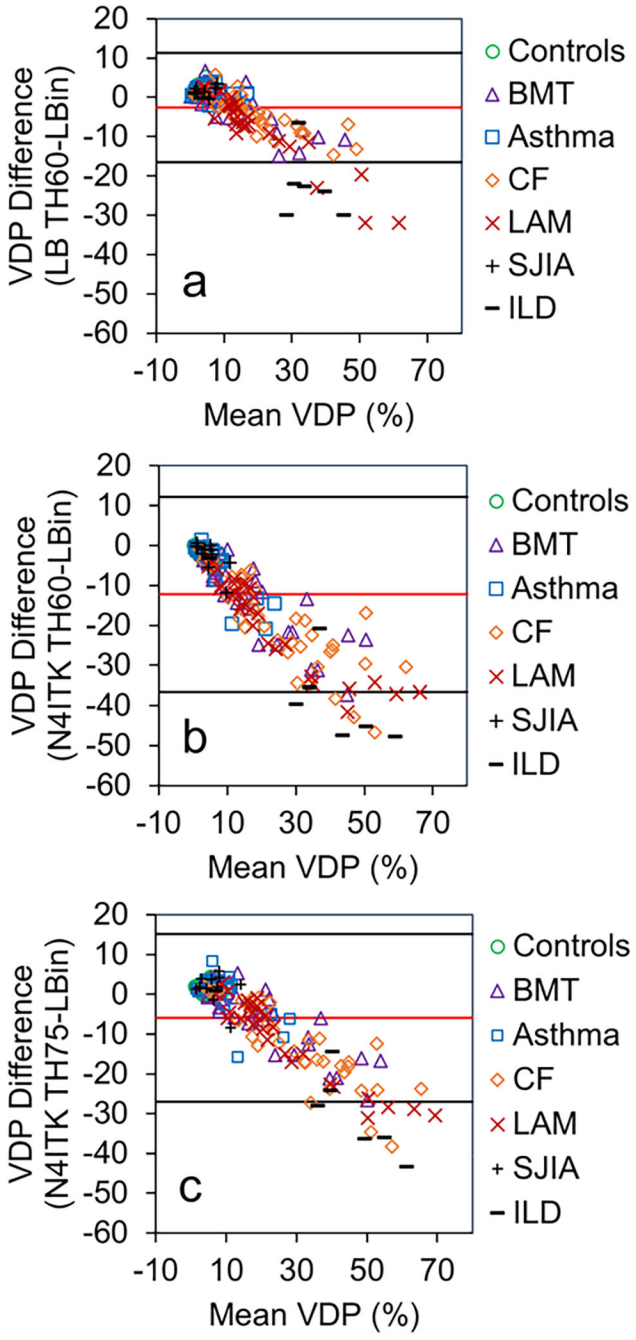

Figure 3.

Bland-Altman plots comparing thresholding mean signal and linear binning ventilation defect quantification schemes. a) ^{129}Xe ventilation defect percentage (VDP) quantified via linear bias corrected 60% threshold mean signal and linear binning. b) ^{129}Xe VDP quantified via N4ITK-bias corrected 60% threshold mean signal plotted against linear binning. c) ^{129}Xe VDP quantified via N4ITK-bias corrected 75% threshold mean signal plotted against linear binning. LB = linear bias corrected, N4ITK = N4ITK-bias field corrected, TH60 is thresholding at 60% mean lung signal, TH75 is thresholding at 75% mean lung signal, and Lbin = Linear Binning.

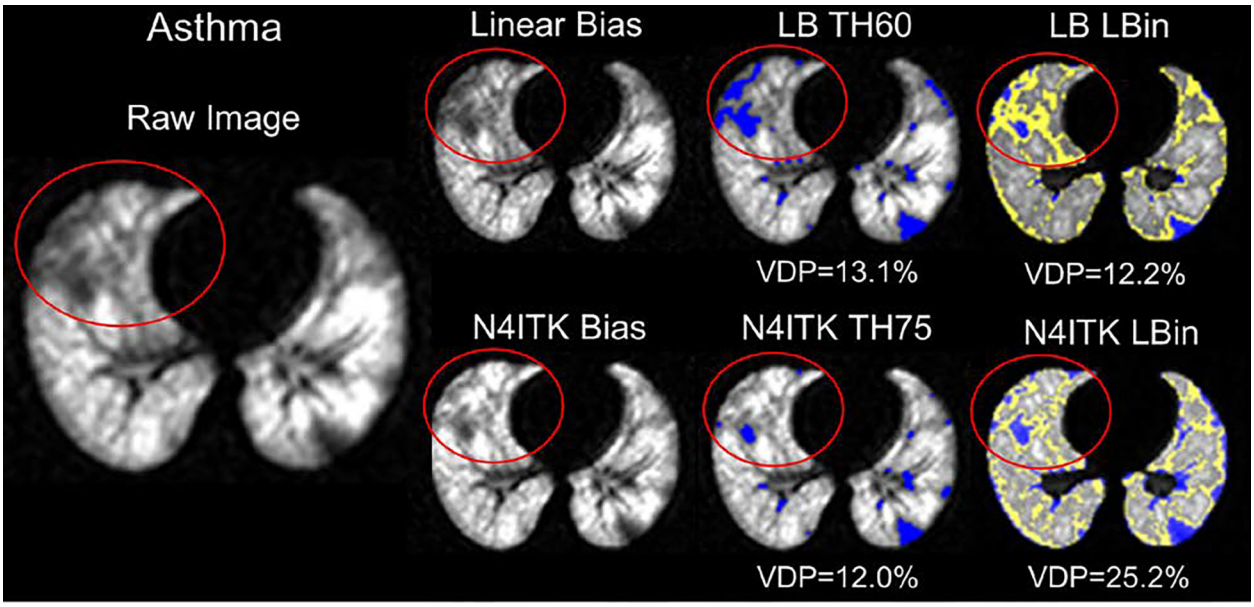

Figure 4.

Blue identifies ventilation defects determined via the different quantification schemes where LB = linear bias corrected, TH60 is thresholding at 60% mean lung signal, TH75 is thresholding at 75% mean lung signal, N4ITK = N4ITK-bias field corrected, and Lbin = Linear Binning. Yellow identifies low ventilation regions identified via linear binning.

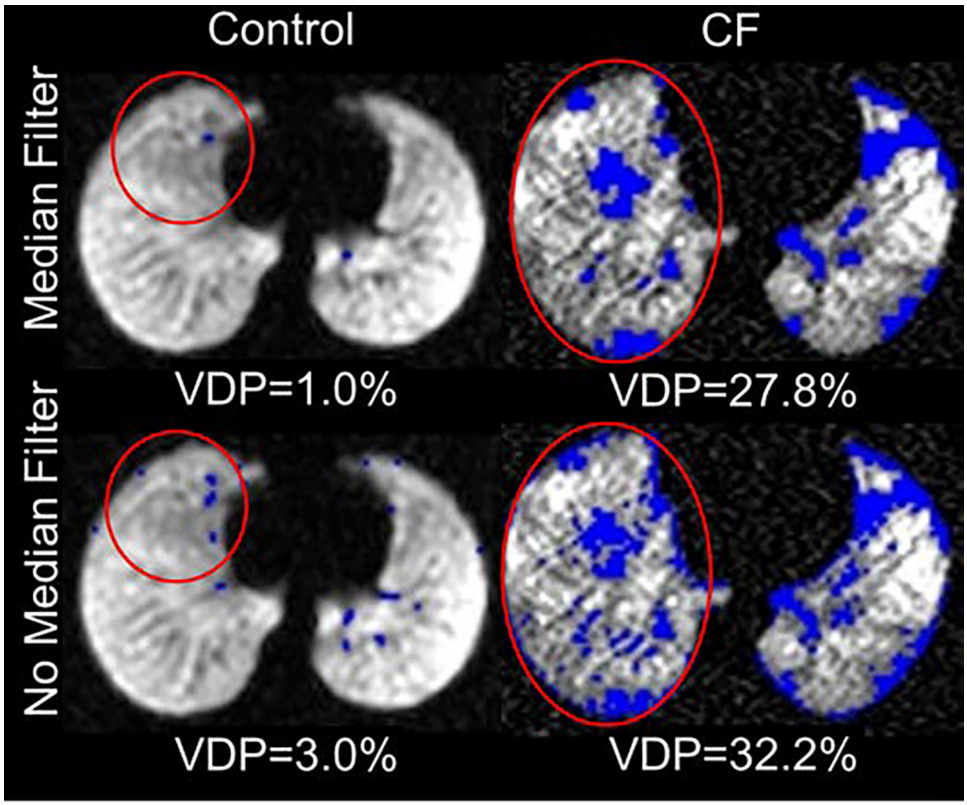

Figure 5.
Median effect on ventilation defects (blue) quantified in a control and a CF subject via 60% thresholding mean signal. Red circles highlight pixels affected by median filter.

Table 1.

Subject demographics and spirometry results.

Subject Group	# Subjects	Age (Years)	Sex
Controls	23	12.10±4.12	13M/10F
BMT (Ttest vs Controls)	47	13.53±4.36 (P=0.874)	23M/24F
Asthma (Ttest vs Controls)	22	12.82±2.99 (P=0.284)	10M/12F
CF (Ttest vs Controls)	37	17.01±8.74 (P=0.084)	20M/17F
LAM (Ttest vs Controls)	29	49.18±11.62 (P<0.001)	29F
sJIA (Ttest vs Controls)	11	10.04±4.33 (P<0.05)	2M/9F
ILD (Ttest vs Controls)	7	48.90±21.83 (P<0.01)	6M/1F
All Disease Groups (Ttest vs Controls)	153	22.32±17.22 (P<0.0001)	61M/92F

Table 2.

Spirometry and VDP Analysis.

Subject Group	FEV1% Predicted	FVC% Predict	FEV_1/FVC	SNR	Linear-Bias Linear Binning VDP	N4ITK-Bias Linear Binning VDP	Linear-Bias VDP 60% Threshold	N4ITK-Bias VDP 60% Threshold	N4ITK-Bias VDP 75% Threshold
Controls	101.50±18.90 (n=18)	103.78±16.58 (n=18)	85.22±7.28 (n=18)	16.2±6	2.60±2.05	3.80±2.27	4.28±2.14	1.58±1.00	5.54±2.34
BMT Cut-point, C Stat, Sens., Spec. (Tukey-Kramer vs Controls)	**79.13±21.86 (n=38,P=0.0071)**	**86.76±18.30 (n=37,P=0.0145)**	80.17±13.96 (n=37,P=0.809)	15.2±8.3 (P=0.99)	10.41±12.72 > 2.3, 0.623, 63.8%, 60.9% (P=0.192)	**17.36±17.50 >4.8, 0.742, 74.5%, 73.9% (P=0.0337)**	9.32±8.90 > 4.2, 0.602, 59.6%, 60.9% (P=0.201)	8.01±9.09 >2.0, 0.720, 74.5%, 69.6% (P=0.0746)	14.18±10.54 >7.1, 0.753, 72.3% 78.3% (P=0.011)
Asthma Cut-point, C Stat, Sens., Spec. (Tukey-Kramer vs Controls)	96.43±18.61 (n=21,P=0.990)	107.48±15.27 (n=21,P=0.994)	78.96±8.92 (n=21,P=0.723)	21.3±14.8 (P=0.36)	4.35±4.64 > 2.0, 0.601, 63.6%, 56.5% (P=0.992)	9.37±9.31 >4.6, 0.689, 68.2%, 69.6% (P=0.929)	6.04±4.30 >4.2, 0.600, 59.1%, 60.9% (P=0.991)	4.42±4.11 >2.0, 0.689, 68.2%, 69.6% (P=0.936)	9.74±5.85 >6.6, 0.689, 68.2%, 69.6% (P=0.774)
CF Cut-point, C Stat, Sens., Spec. (Tukey-Kramer vs Controls)	87.21±23.92 (n=34,P=0.264)	98.12±16.65 (n=33,P=0.926)	77.61±11.50 (n=33,P=0.276)	10.1±6.7 (P=0.09)	**20.94±13.54 > 6.8, 0.911, 86.5%, 95.7% (P<0.0001)**	**35.68±20.85 >7.0, 0.916, 91.6%, 91.3% (P<0.0001)**	**16.91±9.84 >8.1, 0.911, 86.5%, 95.7% (P<0.0001)**	**15.97±11.40 >3.5%, 0.924, 89.2%, 95.7%> (P<0.0001)**	**23.40±11.96 10.1, 0.911, 86.5%, 95.7% (P<0.0001)**
LAM Cut-point, C Stat, Sens., Spec. (Tukey-Kramer vs Controls)	84.17±23.83 (n=27,P=0.125)	95.31±18.43 (n=27,P=0.680)	**70.18±16.78 (n=27,P=0.003)**	10.5±5.6 (P=0.18)	**22.80±19.30 >6.8, 0.909, 86.2%, 95.7% (P<0.0001)**	**31.84±21.99 >7.3, 0.939, 96.6%, 91.3% (P<0.0001)**	**15.44±10.66 >8.1, 0.892, 82.8%, 95.74% (P<0.0001)**	**13.05±11.63 >3.5, 0.927, 89.7%, 95.7% (P=0.0002)**	**20.83±12.09 >10.1, 0.927, 63.6%,95.7% (P<0.0001)**
sJIA Cut-point,C Stat, Sens., Spec. (Tukey-Kramer vs Controls)	101.00±14.56 (n=7,P=1.0)	105.06±17.78 (n=7,P=1.0)	86.97±5.19 (n=7,P=0.999)	19.1±6.0 (P=0.96)	2.69±2.36 No optimal points found (P=1)	5.79±4.71 >3.5, 0.646, 72.7%, 56.5% (P=0.999)	4.10±2.78 No optimal points found (P=1.00)	3.25±2.32 >1.4, 0.670, 81.8%, 52.2% (P=0.998)	7.31±3.77 >5.5, 0.623, 63.6%, 60.9% (P=0.999)
ILD Cut-	80.71±19.90 (n=7,P=0.322)	**79.65±17.50 (n=7,P=0.045)**	79.82±15.40 (n=7,P=0.963)	6.7±2.9 (P=0.12)	**40.21±17.74 Over**	**54.10±24.86 Over**	**20.83±9.51 Over**	**19.84±10.98 Over**	**28.16±10.88**

Subject Group	FEV1% Predicted	FVC% Predict	FEV_1/FVC	SNR	Linear-Bias Linear Binning VDP	N4ITK-Bias Linear Binning VDP	Linear-Bias VDP 60% Threshold	N4ITK-Bias VDP 60% Threshold	N4ITK-Bias VDP 75% Threshold
point, C Stat, Sens., Spec. (Tukey-Kramer vs Controls)					**Fitting (P<0.0001)**	**Fitting (P<0.0001)**	**Fitting (P=0.0001)**	**Fitting (P<0.001)**	**Over Fitting (P<0.0001)**
All Disease Groups Cut-point, C Stat, Sens., Spec. (T-test vs Controls)	**86.13±22.50 (n=134,P=0.0042)**	**95.24±18.86 (n=131,P=0.056)**	**77.54±13.59 (n=131,P=0.0007)**	13.8±9.4 (P=0.12)	**15.24±16.13 >2.9,0.720, 74.5%, 69.9% (P<0.001)**	**24.23±21.86 >4.9, 0.797, 81.0%, 78.3% (P<0.0001)**	**11.99±9.90 >4.8, 0.701, 70.6%, 69.6% (P<0.0001)**	**10.57±10.57 >2.3, 0.817, 76.5%, 87.0% (P<0.0001)**	**17.18±11.84 >7.1, 0.790, 79.7%,78.3% (P<0.0001)**

VDP Mean ± Standard Deviation, VDP Cut-point C-statistics, Sensitivity, Specificity

*Bold indicates statistically significant, Yellow = most sensitive method