



OPEN

# Lung and fissure shape is associated with age in healthy never-smoking adults aged 20–90 years

Mahyar Osanlouy<sup>1</sup>, Alys R. Clark<sup>1</sup>, Haribalan Kumar<sup>1</sup>, Clair King<sup>2</sup>, Margaret L. Wilsher<sup>2,3</sup>, David G. Milne<sup>4</sup>, Ken Whyte<sup>2,3</sup>, Eric A. Hoffman<sup>5</sup> & Merryn H. Tawhai<sup>1</sup>✉

Lung shape could hold prognostic information for age-related diseases that affect lung tissue mechanics. We sought to quantify mean lung shape, its modes of variation, and shape associations with lung size, age, sex, and Body Mass Index (BMI) in healthy subjects across a seven-decade age span. Volumetric computed tomography from 83 subjects (49 M/34 F, BMI  $24.7 \pm 2.7$ ) was used to derive two statistical shape models using a principal component analysis. One model included, and the other controlled for, lung volume. Volume made the strongest contribution to shape when it was included. Shape had a strong relationship with age but not sex when volume was controlled for, and BMI had only a small but significant association with shape. The first principal shape mode was associated with decrease in the antero-posterior dimension from base to apex. In older subjects this was rapid and obvious, whereas younger subjects had relatively more constant dimension. A shift of the fissures of both lungs in the basal direction was apparent for the older subjects, consistent with a change in tissue elasticity with age. This study suggests a quantifiable structure–function relationship for the healthy adult lung that can potentially be exploited as a normative description against which abnormal can be compared.

Advancing age is associated with increasing chest wall stiffness and changes to thorax shape due to calcification of costal cartilages, narrowing of intervertebral spaces, and increased dorsal kyphosis/anteroposterior diameter ('barrel chest')<sup>1</sup>. Given the tight apposition of the lung and chest wall, it is reasonable to expect that the lung changes shape along with the thorax. However, to date adult age-related lung shape has only been directly examined quantitatively using clinical 2D chest X-ray<sup>2</sup> which cannot give a complete shape description.

Quantitative descriptions of normal lung shape, inter-subject variability, and age-related differences are important for several reasons. First, diseases with age-related prevalence that affect the lung tissue can develop with a regional preference; for example, preferentially apical for emphysema as a component of chronic obstructive pulmonary disease (COPD), or subpleural basal for idiopathic pulmonary fibrosis (IPF). In IPF this has been proposed to be related to locally high shear stress<sup>3</sup>, which itself depends upon the degree of lung inflation and change in shape associated with posture and chest wall expansion during breathing. Detailed testing of this hypothesis has been limited by a lack of description of the normal and pathological lung shape. Second, mechanical changes to the lung tissue and chest wall that accompany COPD or IPF affect the functional deformation of the lung tissue during breathing or breath-hold<sup>4</sup>. Lung and lobe shape—as well as the spatial distribution of abnormal tissue—could therefore hold prognostic information for staging or stratification of patients with these conditions. Finally, information on shape would improve automated image processing methods for detection of the pulmonary fissures, where fissure integrity, for example, is an important predictor of outcome for endobronchial valve treatment of severe emphysema<sup>5</sup>.

<sup>1</sup>Auckland Bioengineering Institute, University of Auckland, Auckland 1010, New Zealand. <sup>2</sup>Respiratory Services, Auckland City Hospital, Auckland 1010, New Zealand. <sup>3</sup>Faculty of Medical and Health Sciences, University of Auckland, Auckland 1010, New Zealand. <sup>4</sup>Department of Radiology, Auckland City Hospital, Auckland 1010, New Zealand. <sup>5</sup>Departments of Radiology and Bioengineering, University of Iowa, Iowa 52242, USA. ✉email: m.tawhai@auckland.ac.nz

Anthropometric data (mean $\pm$ SD, N = 83)	
Age (years)	53 $\pm$ 22
Sex (M/F)	34 (41%)/49 (59%)
Height (m)	1.70 $\pm$ 0.12
Weight (kg)	69 $\pm$ 11
BMI (kg/m <sup>2</sup> )	24.7 $\pm$ 2.7
<b>Ethnicity</b>	
Caucasian	78
New Zealand Māori	1
Asian	1
African American	1
Unknown	2

**Table 1.** Summary anthropometric data for a cohort of never-smoking subjects aged 20–93 years, used for lung shape analysis. *SD* standard deviation, *N* sample size.

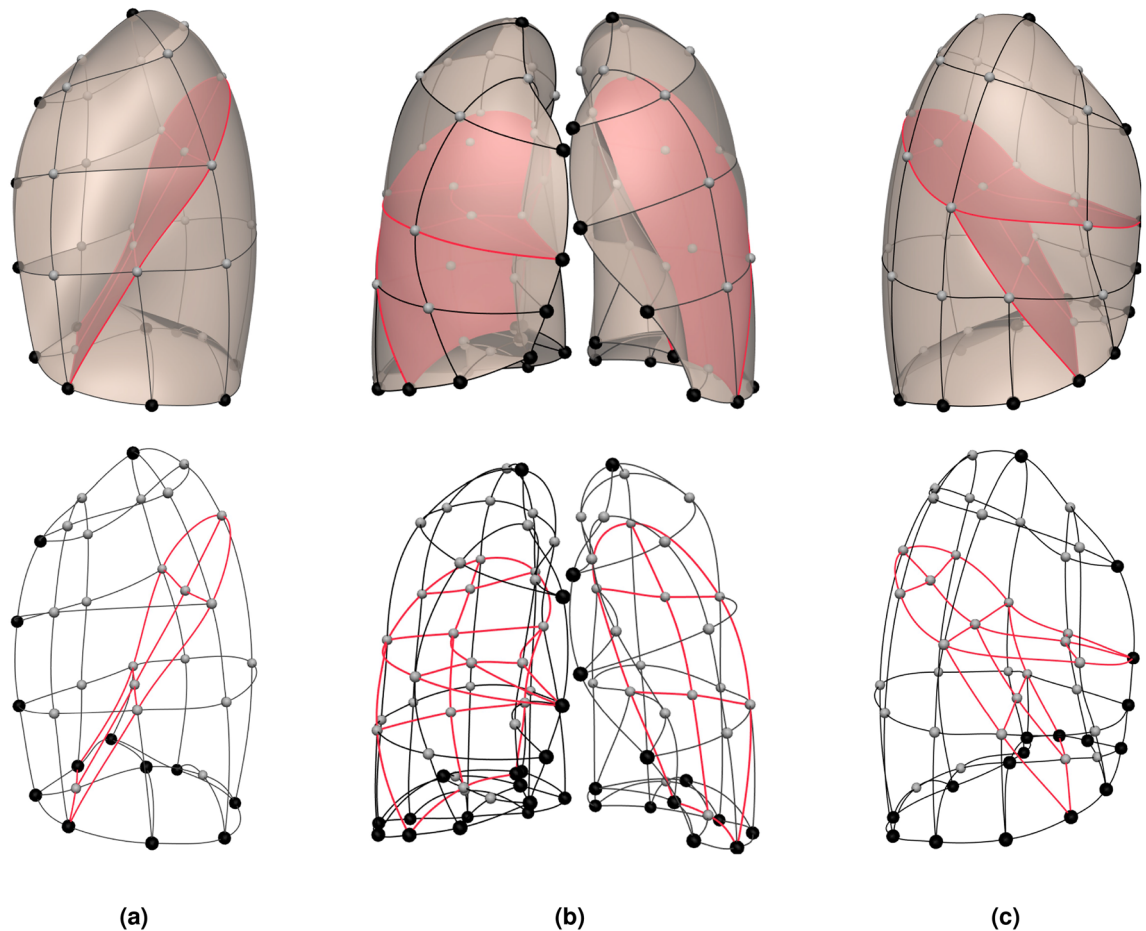
Despite its importance, descriptions of normal lung shape remain largely qualitative except for a relatively small number of studies (e.g. <sup>6–8</sup>). Several studies have suggested sexual dimorphism in ribcage morphology<sup>7,9,10</sup> and to a small extent (7%)—in static supine lung shape<sup>11</sup>. Sex-related shape differences could be important for understanding the greater susceptibility of women to COPD<sup>12</sup>. Other studies of subjects with and without COPD have observed a dependence of diaphragm shape on age (16) and Body Mass Index (BMI)<sup>13,14</sup> and an association between percent emphysema and cross-sectional area of the lung<sup>12</sup>. However, no study has yet considered the impact of healthy adult age on lung—rather than its ‘container’—shape.

Biological shape variations are often complex and highly challenging to interpret. Principal component analysis (PCA) is a mathematical method that has previously been used to describe the shape of the lung<sup>8</sup> and other organs<sup>15–17</sup>. PCA simplifies the complex patterns of shape variation into independent variables (principal components), providing more-readily interpretable metrics of shape than simple linear dimensions. Here, we sought to quantify the average lung shape using a PCA to derive a statistical shape average model and its principal modes of shape variation for a cohort of healthy adults aged 20–93 years of age, and to determine whether lung shape in these subjects is associated with age, sex, or BMI.

## Methods

**Data acquisition and preprocessing.** Imaging and pulmonary function testing data from 83 (34 M/49 F) never-smoking subjects aged 20–93 years, with no history of lung disease or injury to the lung or chest wall were analyzed. Data were retrospectively obtained from two studies with consistent protocols. 47 (19 M/28 F) subjects aged 50–93 years (71.2  $\pm$  10.9 years) were recruited with approval from the Northern A Health and Disability Ethics Committee, and data were acquired at Auckland City Hospital (New Zealand). A further 36 (15/21 M/F) subjects aged 20–49 years (mean 32.0  $\pm$  12 years) were selected retrospectively from the University of Iowa Comprehensive Lung Imaging Center ‘Human Lung Atlas’ database (data acquired under NIH R01-HL-064368, E. A. Hoffman PI, following approval by the University of Iowa Institutional Review Board and Radiation Safety Committees). All the data acquisition and analysis methods were carried out in accordance with relevant guidelines and regulations. An informed consent was obtained from all subjects. Subject data were selected to provide as uniform as possible spread of age by decade and sex, within the constraints of the available data. Data comprised volumetric multi-detector row computed tomography (MDCT) at full inspiration, spirometry, and lung volumes by plethysmography. Exclusion criteria were BMI > 30 kg/m<sup>2</sup>, a history of respiratory or cardiac disease, ever-smoking, FEV1 < 80% predicted and FEV1/FVC < lower limit of normal<sup>18</sup>. Anthropometric data are given in Table 1.

Volumetric imaging had slice spacing 0.5–0.7 mm with reconstruction matrices of 768  $\times$  768 (> 50 years old) or 512  $\times$  512 (< 50 years old). Scan parameters were 120 kV, 100 mAs, and pitch of 1.2. Lung and fissure surfaces were segmented using custom-written software<sup>19</sup> with manual correction of fissures as required. To define the lung shape a high-order finite element (FE) mesh was geometry-fitted to the lung and fissure surfaces<sup>20</sup> using the same mesh topology for all subjects (Fig. 1). The generic FE mesh had 35 nodes and 44 elements covering the left lung and left oblique fissure surfaces, and 50 nodes and 62 elements covering the right lung and right oblique and horizontal fissure surfaces. In order to provide anatomical and mathematical shape correspondence between subjects, anatomical landmarks were placed at the lung apices, dome of the diaphragm, the edge of the lung base, and the edge of anterior segments. Additional pseudo-landmarks were regularly-spaced between these points. The generic mesh was geometrically fitted to each subject’s segmented lung image using a linear least squares optimization<sup>20</sup>. Briefly, the sum of the Euclidean distances between each data point and its projection onto the nearest element was minimized during the fitting process. This distance is a function of the node location and shape parameter, i.e. its derivatives. The points are fitted by minimizing the energy function  $T(u)$  with respect to the global shape parameters:



**Figure 1.** Finite element mesh topology used to construct subject-specific lung shapes. (a) Lateral view of the left lung, (b) front view of both lungs, and (c) lateral view of the right lung. Lobar fissures are also highlighted. The larger black spheres indicate the nodes used as anatomical landmarks, the smaller grey spheres indicate the nodes used as pseudo landmarks, and the lines indicate the interpolation edges that are used as the description of surface curvature—note that the fissure is shown in red colour. This approach to describe the shape ensures that the entire lung surface is taken into account in the statistical modeling calculations.

$$T(u) = \sum \gamma \|z(\xi_1, \xi_2) - z_d\|^2 + \int_{\Omega} g(u(\xi)) d\xi, \quad (1)$$

where  $u$  is a vector of mesh parameters,  $\gamma$  is a weight factor for each point,  $z$  is geometric position which can be given by local element coordinate  $\xi$ ,  $z_d$  is the spatial coordinates of the data point, and  $\Omega$  is the mesh domain.  $g$  is a second order weighted norm which is calculated from the derivatives of the geometry of each element<sup>21</sup>. This term is a smoothness constraint which measures the deformation of the surface, and was included to regularize the problem since data can sometimes be either insufficient or noisy. The root mean squared error for mesh fitting across the cohort was  $5.2 \pm 2.3$  mm.

**Statistical shape analysis: SSA.** We derived two statistical shape models (SSM) for the cohort using a principal component analysis (PCA). An SSM provides a mathematical description of the shape of an object and the ways that its shape varies. It captures the global shape of the object of interest instead of using fixed geometric measurements such as lengths and angles<sup>22</sup>. PCA is a mathematical technique that decomposes the shape of a generic object into its main components such that the weighted sum of these components can retrieve back the object's shape. In our approach, the first model included lung size (using raw unscaled data; the 'size-inclusive' model) and the second model controlled for lung size (using scaled data; the 'size-exclusive' model). The mean shape and principal modes of shape variation were derived for the whole cohort for both the size-inclusive and size-exclusive models. The process of deriving the SSM involves three main steps: data alignment, data assembly, and data decomposition.

*Data alignment.* A key first step in deriving any SSM is to align the object shapes to a chosen reference coordinate. General Procrustes Alignment (GPA)<sup>23</sup> was used to align all meshes to the same axis, hence removing rotation and translation bias; and in the size-exclusive model it was also used to remove scaling bias. This step is

analogous to the rigid registration that removes any rigid body changes to the shape of an object. Without this step, the SSM would include rotation and translation as a component of shape variation, which is not desirable.

For the data alignment, let  $S$  be a vector containing the concatenation of all 3D lung coordinates, i.e.  $S = [x_1, y_1, z_1 \dots]$ , to the reference (mean) surface (i.e.  $1/N \sum_i (x_i^i) \sum_i (y_i^i) \sum_i (z_i^i) \dots$ , where  $i$  iterates over  $N$  cases). We can now define the aligned shape vector  $\bar{S}$  by:

$$\bar{S} = \alpha RS + T, \quad (2)$$

where  $R$  and  $T$  are the unknown rotation matrix and translation vector, respectively, which align shape  $S$ , and  $\alpha \in \mathbb{R}$  is the overall scaling factor.

**Data assembly.** We assemble information about lung shape (here represented as finite element nodes and curvature information at each node) from all of the subjects in the cohort. Each subject's lung shape is considered to be a single observation. Let  $S_L$  be one such observation column containing the shape parameters. In this case:

$$S_L^T = [\bar{u}_1, \bar{u}_2, \bar{u}_3, \dots, \bar{u}_{p-2}, \bar{u}_{p-1}, \bar{u}_p], \quad (3)$$

where  $(^T)$  denotes transpose,  $u$  is the node containing 12 DoFs (four for each direction:  $c, \frac{\partial c}{\partial \xi_1}, \frac{\partial c}{\partial \xi_2}, \frac{\partial^2 c}{\partial \xi_1 \partial \xi_2}$ , where  $c = [x, y, \text{ and } z]$ ). These DoFs together describe the lung shape for one subject in the cohort. For a general surface in 3D space, these DoFs describe both the location and the surface curvature. Furthermore,  $p$  represents the number of nodes for both lungs and the overline ( $\bar{\phantom{x}}$ ) operator symbolizes the GPA to the reference lung model.

A new data matrix  $S$  is built by a concatenation of each lung observation  $S_L$  to construct a lung training set. The full training set can be described as a set of  $N3D$  shapes, where  $N$  is the number of subject lung shapes. Each of these lung shapes is represented by a set of  $n$  landmarks, where  $n$  is the number of nodal shape parameters (i.e. 2700 for each subject, equalling 225 nodes times the number of nodal parameters of coordinates and derivatives for each node). By assembling the shape parameters this way, every lung shape has been converted from a 3D space to a single point in  $3nD$  space. This  $3nD$  space constitutes an 'allowable shape domain'.  $S$  can be seen as a cloud of  $N$  points in the constructed  $3nD$  space that lies within this domain<sup>24</sup>. Once such a space is constructed, any allowable shape will have to be a member of this domain. Conversely, any shape that is extracted from this domain will be an allowable shape.

**Data decomposition.** Decomposition aims to reduce the complexity of the data. Once the data has been assembled, the data matrix  $S$  can be decomposed into modes of shape variation using a number of different techniques. The current study employs a linear decomposition of  $S$  by PCA within the allowable shape domain. One of the benefits of PCA is to find a linear space of eigenvectors  $m_l$  where  $l = 1, \dots, L$ , with  $L$  being much smaller than the number of variables in the original rectangular Cartesian space (in this case  $L \ll 2700 \times 83 = 224,100$ ). In statistical terminology, this technique is referred to as reducing the dimension of the *feature* space.

**Principal component analysis.** PCA was performed by centering the data in  $S$  around the mean, i.e.  $S = S_0 - \bar{S}_0$  where  $S_0$  is the original data matrix and  $\bar{S}_0$  is the mean of  $S_0$ . Next the covariance matrix of  $S$  was built by  $C = SS^T$ . Mean-centered data matrix  $S$  was then factorized using singular value decomposition (SVD) to yield:

$$S = U \Sigma V^T, \quad (4)$$

where  $U$  is an  $m \times m$  unitary matrix of eigenvectors of  $SS^T$ ,  $\Sigma$  is an  $m \times n$  rectangular diagonal matrix with non-negative square roots of the eigenvalues of  $SS^T$ , and  $V$  is an orthonormalized  $n \times n$  unitary matrix containing eigenvectors of  $C = SS^T$ .  $U$  is a rotational representation of the lung matrix  $S$ <sup>25</sup>. Assuming that all dimensionality is preserved, the PCA transformation can be expressed by  $U$  using  $X = U^T S$ . The elements in each column of  $U$ , namely  $u$ , are the new variables transformed into the PCA-space  $X$ . One column (i.e.  $u_l$ ) represents a mode of variation or *principal components*, ordered by the singular values of  $\Sigma$ , namely  $\sigma_l$ . The singular values are proportional to the size of the variance corresponding to each eigenvector. This clearly shows that the variance explained by each component  $u_l$  can be derived as  $\sigma_l \equiv \sqrt{\lambda_l}$ , where  $\lambda_l$  are the eigenvalues of  $C$ . The corresponding eigenvalues of  $C$  show how significant the components  $\sigma$  are: a larger eigenvalue means a more significant eigenvector, hence a more significant mode of variation.

Shape modes are independent descriptors of shape and can be expressed as a percentage of the total shape variation in the population. Modes of variation are defined as perturbations about the mean and can be noted as  $m_l$  for each principal component  $u_l$ . In this case:

$$m_l(w) = \bar{S}_0 + wu_l \equiv M_l, \quad (5)$$

where  $w$  is a weight factor given to each mode of variation and  $l = 1, \dots, L$ . Variations in  $w$  within suitable limits (often between  $\pm 2 - 3\sigma$ ) allows for a direct visualization of the lung shape and shape changes. Each subject's weight values for each mode of variation were calculated by projecting the subject onto the trained population's PCs. The projection is performed by using the dot product of the PCA variance matrix and the subject's shape vector:

$$s = (X_i - s_{mean}) \cdot R, \quad (6)$$

where  $s$  is the array of PC scores,  $X_i$  is the desired subject dataset,  $s_{mean}$  is the mean shape vector, and  $\mathbf{R}$  is the population PCA variance matrix. The scores in the array  $s$  were then converted into relative standard deviation (S.D.) weight values.

**Mode selection and geometrical analysis.** The number of shape modes obtained as a result of a PCA analysis is typically large. It is impractical and also not useful to study all of the shape modes. We employed the cumulative percentage of total variation as the selection criterion. The number of modes included in the analysis is then the smallest value of  $m$  for which this chosen percentage cut-off is exceeded. Modes are successively selected to have the largest possible variance. Let the variance of the  $k$ th mode be denoted as  $l_k$ . Then the sum of the variances of the modes is equal to the sum of the variances of the elements of  $S$  (our data; also note that  $p$  indicates those variables in  $S$ ), which when expressed mathematically:

$$\sum_{k=1}^p l_k = \sum_{j=1}^p s_{jj}, \quad (7)$$

As a result, the definition of ‘percentage of variation’  $v_m$  is:

$$v_m = 100 \times \frac{\sum_{k=1}^m l_k}{\sum_{j=1}^p s_{jj}} = 100 \times \frac{\sum_{k=1}^m l_k}{\sum_{k=1}^p l_k}, \quad (8)$$

A sensible cut-off for  $v_m$  often depends on the nature of a particular data. When  $p$  is very large (like in our case which is 224,100) we end up with an impractically large value of  $m$  for further analysis. For large  $m$ , geometrical interpretation of the modes become extremely difficult or even impossible<sup>26</sup>. It is also important to retain as many  $m$  as feasible to interpret and examine. As a result, a cut-off was chosen empirically by examining all of the modes and retaining those which showed the possibility of interpretation. For the significant modes that were chosen, a careful and detailed approach using 3D graphical visualization techniques was employed to derive information regarding the topological and geometrical changes in lung shape.

**Statistical analysis.** Weightings for the first four shape modes for each model were tested for relationships with age, BMI, sex, and lung volume using ordinary least squares regression and Pearson correlation. Means of the 50 year old groups were compared using one-way ANOVA. For age, we used a standard two-sided independent t-test with unequal population variance (Welch’s  $t$ -test). A confidence level ( $\alpha$ ) of 0.05 was considered statistically significant.

## Results

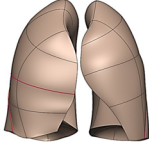
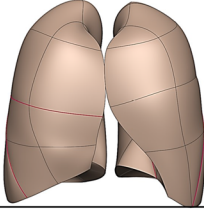
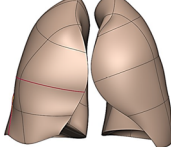
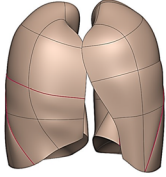
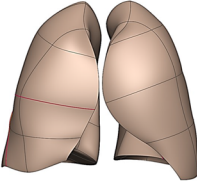
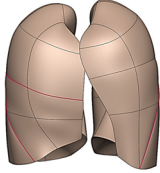
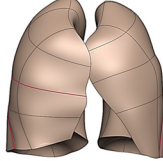
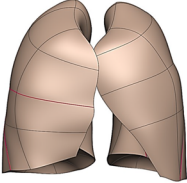
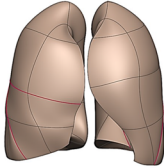
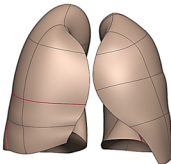
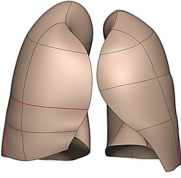
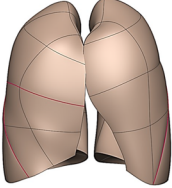
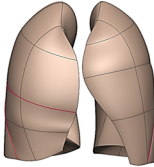
To confirm that the inclusion of data from different centers did not influence the imaged lung volume (and hence potentially the shape), the imaged full inspiratory supine lung volume as a proportion of plethysmographic (upright) assessed total lung capacity (TLC) was compared between subjects from the two centers (the  $> 50$  and  $< 50$  year old groups). There was no significant difference ( $p = 0.15$ ) in the relative lung volume during imaging (ratios were  $0.88 \pm 0.14$  and  $0.93 \pm 0.08$  for the  $> 50$  y.o. and  $< 50$  y.o. group, respectively).

**Qualitative descriptions of modes of lung shape variation.** Two separate shape models were derived to enable evaluation of the contribution of size (volume) to shape: the size-inclusive model and size-exclusive model. Lung shape for  $\pm 2.5$  standard deviations from the mean are illustrated in Fig. 2 for the first four principal modes, for the size-inclusive and -exclusive models. For both models, the first mode explained 11% of the shape variation and the second explained 10%. Modes 3 and 4 explained 9% (6%) and 6% (5%), respectively for the size-inclusive (-exclusive) model; that is, the first four modes explained a total of 36% (32%) of the shape variation in the cohort.

In Fig. 2, negative weightings for mode 1 in the size-inclusive model (tending to be female) corresponded to smaller lungs with thinner anterior edges; positive weightings corresponded to larger lungs (tending to be male) and thicker anterior edges that almost touched each other, with locally dilated regions. In contrast, in the size-exclusive model the first mode corresponded mainly to the antero-posterior diameter of the lung with a mediolateral narrowing towards the apices for negative weightings, and relatively constant mediolateral dimension from base to apex for positive weightings. In addition, this mode in the size-exclusive model showed a large shift in fissure locations in both the right and left lungs with mode weighting: the horizontal fissure descended towards the base as weighting became more negative, and the oblique fissures retracted posteriorly. The second principal mode for the size-inclusive model showed a very similar pattern of variation to the first principal mode for the size-exclusive model, explaining 10% of the cohort’s variation in shape.

For the second shape mode of the size-exclusive model, the shape variability involved an inwards rotation of the anterior aspect of each lung about the cranio-caudal axis, such that the upper anterior region of each lung moved closer to each other for negative weightings. The third mode for the size-inclusive model showed a very similar pattern of variation to this. The size-exclusive mode 3 and size-inclusive mode 4 presented some degree of aspect ratio change (that is, the ratio of the width to the height of lung). The size-exclusive mode 4 presented a more aggressive aspect ratio change with a more inflated and ‘barrel-shaped’ lung for positive weightings, and



Size-inclusive		Modes and contributions (%)	Size-exclusive	
-2.5 $\sigma$	+2.5 $\sigma$		-2.5 $\sigma$	+2.5 $\sigma$
		Mode 1 11%   11%		
			Mode 2 10%   10%	
		Mode 3 9%   6%		
			Mode 4 6%   5%	
		Total 36%   32%		

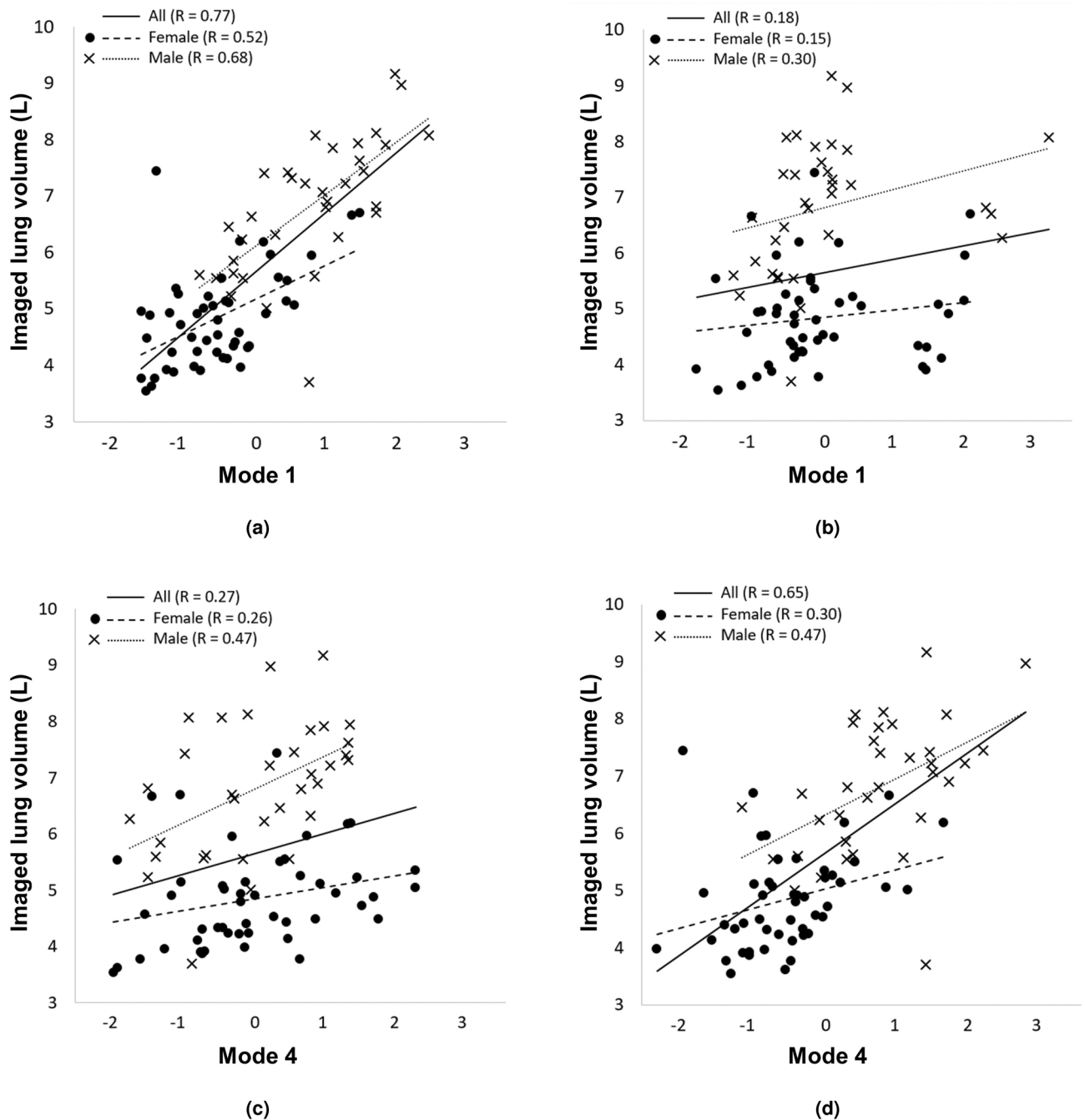
**Figure 2.** Lung shape at  $\pm 2.5$  SD from the mean for each of the first four principal modes, for the size-inclusive (left column) and size-exclusive (right column) statistical shape models. The left main panel illustrates shape modes 1 to 4 for the size-inclusive model, and the right main panel for the size-exclusive model. The middle column shows the relative contribution of each mode as a percentage of the total population shape variance.

a less inflated and elongated lung for negative weightings. This shape mode also showed some rotation about the cranio-caudal axis, with the anterior of the lungs rotating towards each other for positive weightings.

**Associations between lung shape, sex, and size.** All shape modes were examined for associations with age, sex and lung size. Only the modes that have significant relationships with physiological or anthropometric data are illustrated, except for sex differences.

As shown in Fig. 3a, when volume was not controlled for the first principal shape mode was strongly associated with lung size ( $R = 0.77$ ,  $p < 0.001$ ), with the female data tending towards negative, and male data positive, weightings. The relationship for the whole cohort was stronger than for females ( $R = 0.52$ ,  $p < 0.001$ ) or males ( $R = 0.68$ ,  $p < 0.001$ ) considered separately. In contrast, mode 1 was not associated with lung size in the size-exclusive model ( $R = 0.18$ ,  $p = 0.10$ ), but mode 4 was ( $R = 0.65$ ,  $p < 0.001$ ) when considering the entire cohort. While Fig. 2 suggests qualitatively that size could be a factor in modes 2–4 of the size-inclusive model, this is not borne out statistically: size-exclusive mode 4 (Fig. 3d) is the only other shape mode that shows a significant association with lung volume ( $R = 0.65$ ,  $p < 0.001$ ). Similar to the distribution of data for mode 1 of the size-inclusive model (Fig. 3a), mode 4 in Fig. 3d shows females tending to negative weightings and males to positive weightings.

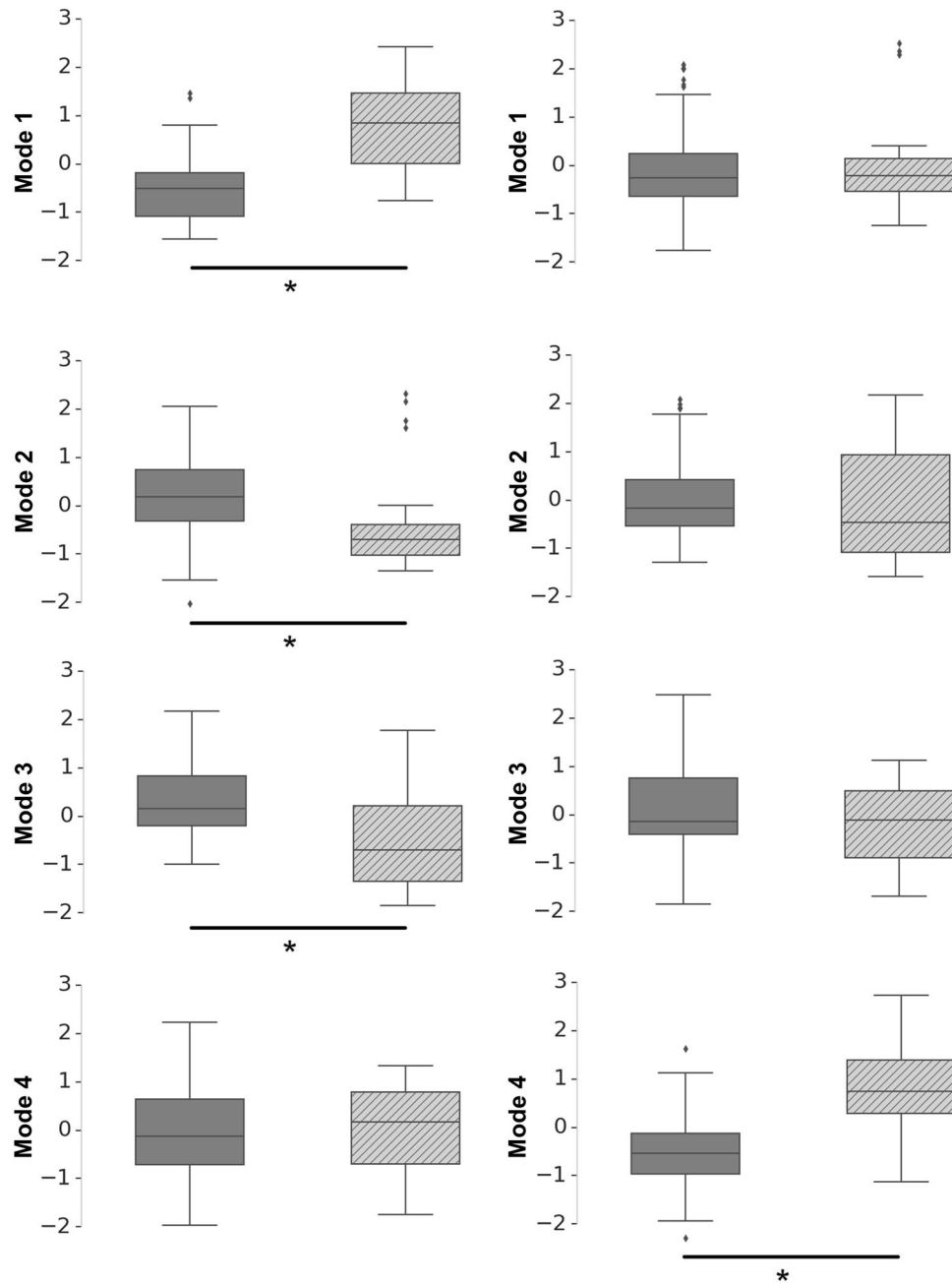
Figure 4 and Table 2 summarize the distributions of shape mode weightings when grouped by sex. The first three shape modes for the size-inclusive model are significantly different ( $p < 0.001$  for mode 1,  $p = 0.002$  for



**Figure 3.** The relationship between total imaged lung volume and principal shape modes 1 and 4 for the (a,c) size-inclusive and (b,d) -exclusive models. Data for females (filled circles) and males (crosses) are indicated separately. Linear regressions to the data are shown for the whole cohort (solid lines) and separately for females (heavy dashed lines) and males (light dotted lines). The size-inclusive mode 1 and size-exclusive mode 4 are strongly associated with volume ( $R = 0.77$ ,  $p < 0.001$ , and  $R = 0.65$ ,  $p < 0.001$ , respectively). Significant relationships are not apparent for the other two model modes ( $R=0.18$ ,  $p=0.10$  for size-exclusive mode 1, and  $R=0.27$ ,  $p=0.08$  for size-inclusive mode 4).

mode 2, and  $p < 0.001$  for mode 3) between males and females, but the fourth mode is not. In contrast, the size-exclusive model modes 1 to 3 have no sex differences, whereas the male/female difference in mode 4 is significant ( $p < 0.001$ ). For the size-inclusive model, males and females tend towards negative and positive weightings for mode 1, respectively.

**Associations between lung shape and age.** Age is strongly associated with shape for both the size-inclusive and -exclusive models in Fig. 5. Strong relationships with age are evident for size-exclusive mode 1 (Fig. 5b,  $R = -0.75$ ,  $p < 0.001$ ) and size-inclusive mode 2 (Fig. 5c,  $R = -0.65$ ,  $p < 0.001$ ). For both modes, the

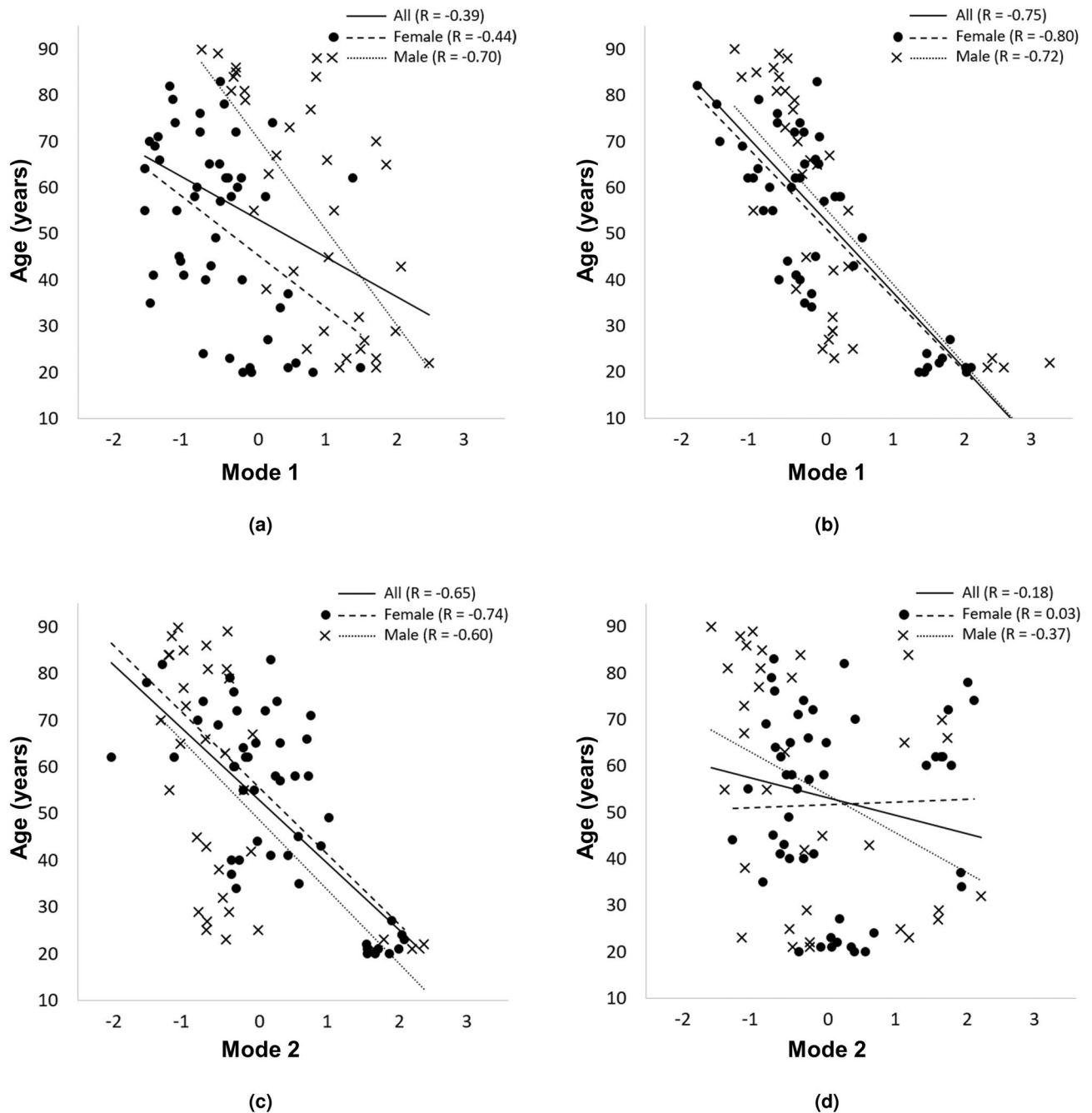


**Figure 4.** Distribution of mode weightings for the size-inclusive (left column) and size-exclusive (right column) models, for females (grey) and males (cross-hatched). The shape modes 1–4 are shown from top to bottom. Statistical significance is indicated using ‘\*’. Modes 1–3 in the size-inclusive model and mode 4 of the size-exclusive model have statistically significant differences between males and females.

	Size-inclusive model			Size-exclusive model		
	Female (mean ± SD)	Male (mean ± SD)	p-value	Female (mean ± SD)	Male (mean ± SD)	p-value
Mode 1	-0.50 ± 0.72	0.77 ± 0.84	< 0.001*	-0.01 ± 0.98	0.04 ± 1.03	0.83
Mode 2	0.28 ± 0.96	-0.40 ± 0.94	0.002*	1.31 ± 0.92	-0.16 ± 1.10	0.20
Mode 3	0.38 ± 0.83	-0.54 ± 0.98	< 0.001*	0.13 ± 1.10	-0.20 ± 0.87	0.14
Mode 4	-0.03 ± 1.06	0.03 ± 0.91	0.80	-0.52 ± 0.75	0.74 ± 0.84	< 0.001*

**Table 2.** Mean (± standard deviation) of the first four principal shape mode weightings for males and females for a size-inclusive and a size-exclusive statistical shape model. Corresponding p-values from an independent t-test for females and males are shown, with p < 0.05 (indicated by \*) considered statistically significant.

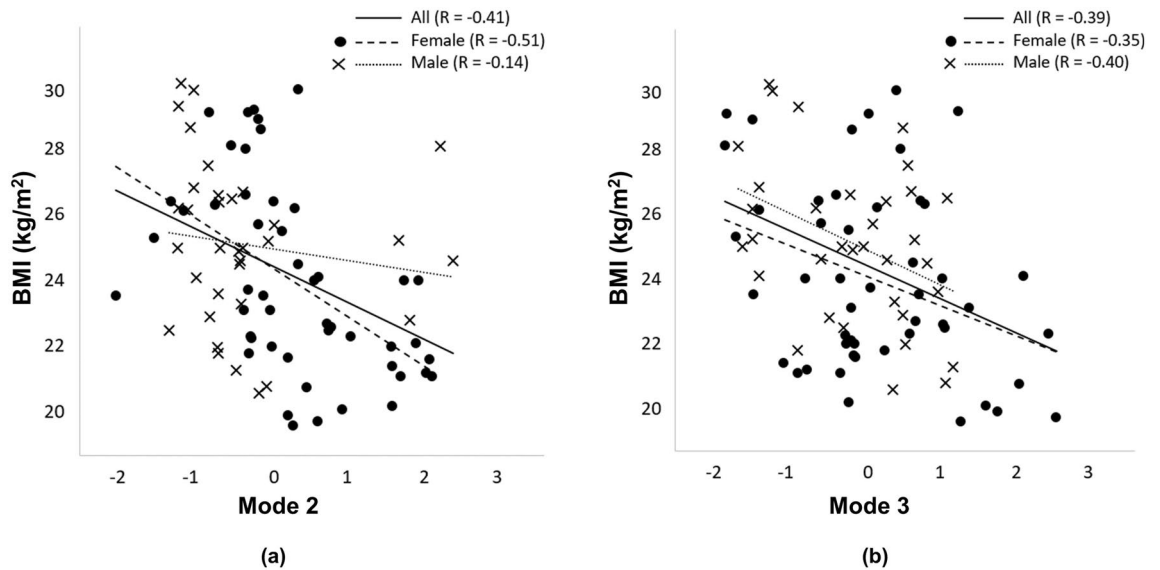




**Figure 5.** The relationship between age and shape modes 1 and 2 for the (a,c) size-inclusive and (b,d) size-exclusive models. Data for females (filled circles) and males (crosses) are indicated separately. Linear regressions to the data are shown for the whole cohort (solid lines) and separately for females (heavy dashed lines) and males (light dotted lines). Qualitatively, size-inclusive mode 1 captures the lung size, size-inclusive mode 2 and size-exclusive mode 1 capture a change in the antero-posterior dimension when moving from base to apex as well as a large shift in fissure location, and size-exclusive mode 2 captures rotation about the medio-lateral axis. Age is strongly associated with mode 1 in the size-exclusive model ( $R = -0.75$ ,  $p < 0.001$ ) and with mode 2 in the size-inclusive model ( $R = -0.65$ ,  $p < 0.001$ ). Age is moderately associated with mode 1 in the size-inclusive model ( $R = 0.39$ ,  $p < 0.001$ ).

relationships for separate sexes are not different from that for the whole cohort. Age and size-inclusive mode 1 shows moderate association for the whole cohort (Fig. 5a,  $R = -0.39$ ,  $p < 0.001$ ), however separation by sex reveals a strong relationship for males with age ( $R = -0.70$ ,  $p < 0.001$ ), and a weaker (than male) but important association for females ( $R = -0.44$ ,  $p = 0.002$ ).

**Associations between lung shape and BMI.** The shape modes that appear to be associated with BMI are shown in Fig. 6. Size-inclusive mode 2 has a significant association with BMI ( $R = -0.41$ ,  $p < 0.001$ ).



**Figure 6.** The relationship between (a) BMI and shape mode 2 for the size-inclusive model and (b) shape mode 3 for the size-exclusive model. Data for females (filled circles) and males (crosses) are indicated separately. Linear regressions to the data are shown for the whole cohort (solid lines) and separately for females (heavy dashed lines) and males (light dotted lines). Both modes for the respective models have moderately strong and significant relationships with BMI when considering the entire cohort. The relationship for males in the size-inclusive model is not significant ( $p = 0.43$ ) when considered separately.

Separation by sex reveals a similar relationship for females ( $R = -0.51$ ,  $p < 0.001$ ) but no relationship for males ( $R = -0.14$ ,  $p = 0.43$ ) with BMI. For size-exclusive mode 3, the individual relationships for females ( $R = -0.35$ ,  $p = 0.02$ ) and males ( $R = -0.40$ ,  $p = 0.02$ ) with BMI are similar to that for the whole cohort ( $R = -0.39$ ,  $p = 0.01$ ).

## Discussion

Older age and chronic disease are both known to affect chest wall shape, lung dimensions, and lung function. Lung shape therefore potentially provides a prognostic marker of lung health or accelerated aging, but this has not previously been explored. To distinguish between normal and abnormal lung shape, a quantitative description of normal shape and its variability is first required. Our study provides a quantitative statistical shape description of the lung and its lobes in a cohort of 83 never-smoking healthy subjects aged 20–93 years. The statistical model shows a strong relationship between shape and age, and an association between shape and BMI. Male and female shape only differed when lung size was not controlled for in the model, which differs from previous studies of sex differences in lung shape<sup>8</sup>.

Rib cage geometry has been studied in detail by a number of authors, using quantitative methods to describe its normal shape and revealing relationships with age, BMI, and sex<sup>1,10</sup>. However, the rib cage only describes one component of lung shape: the diaphragm forms the additional boundary of the lung ‘container’, and the pulmonary fissures provide important information about lobar shape. Some other direct analyses of lung shape have relied upon measurements of linear dimensions or cross-sectional area at select locations<sup>11–13</sup>, but this cannot account for detailed features such as surface curvature or the location of the fissures. We used a PCA to derive statistical shape models for healthy, never-smoking, and radiologically normal subjects. The shape models represent the average shape of this cohort and its principal modes of shape variation ranked in order of the proportion of variation in shape that they explain. Any subject in the study cohort can be modeled as the mean shape plus a weighted sum of principal components, providing a compact quantitative description of individual subject shape. A further advantage of this approach is that it is straightforward to test whether a subject that was not in the training cohort can be considered to have the same shape as the cohort.

Average lung size is different between males and females, therefore we derived two shape models that allowed us to assess the contribution of lung size (volume) to the quantitative description of shape and therefore the appearance of sex differences. We expect that mode ( $i + 1$ ) in the size-inclusive (not volume controlled) model captures similar shape information to mode ( $i$ ) in the shape-exclusive (volume controlled) model. This is supported by the similarity of shape for modes ( $i + 1$ ) to modes ( $i$ ) (size-inclusive and -exclusive, respectively) in Fig. 2, and the corresponding data in Fig. 4. However, the size-inclusive model—by its nature—includes size variation as a shape feature in each of the first four principal shape modes (Fig. 2).

Our shape models show strong relationships between age and lung shape (summarized in Fig. 5), consistent with relationships between rib morphology and age<sup>10</sup>. The strongest relationship was with the first principal shape mode in the size-exclusive model. There were no sex differences in the relationship for this model/mode. Similar relationships with both age and sex differences were observed for mode 2 of the size-inclusive model, albeit with slightly weaker relationships that presumably reflect some contribution of lung size. Interestingly, when volume

is not controlled for there is a much stronger relationship between the first principal shape mode and age for males compared with the whole cohort or females. The reason for these differences is not clear.

A more 'pyramidal' lung shape was observed with advancing age: that is, a smaller apical dimension compared with the diaphragmatic region in older subjects (Fig. 2). This was the same for both males and females (results not shown). These findings are consistent with previous studies that have found age-related alterations in thoracic shape, often as the consequence of deformity of the vertebral bodies leading to kyphosis<sup>27</sup>. Thoracic kyphosis leads to an increase in the antero-posterior diameter of the chest which, in turn, results in an increased diameter of the lungs and the subsequent 'pyramidal' lung shape. In addition to differences in the exterior shape of the lung, a large displacement of the fissures with age was seen in both lungs (Fig. 2). The horizontal fissure descended towards the lung base in proportion to age, and the oblique fissures retracted posteriorly. The age-associated shape differences that we observed were therefore not just associated with chest wall remodeling, but also with a change in the configuration of the fissures.

Previous studies have reported that normal age-related changes to lung structure become most obvious from the beginning of the third decade of life<sup>28</sup>, and these changes in structure are associated with loss of static elastic recoil pressure of the lung<sup>29,30</sup>. We note that the large majority of the subjects aged greater than 30 years in our study had negative weightings for the size-exclusive mode 1 (i.e. in the direction of 'older' age in Fig. 5c) and almost every subject aged less than 30 years had positive weightings. That is, for the healthy cohort considered here, the average shape (zero SD weighting) was at about 30 years. This compares with an average cohort age of  $53 \pm 22$  years. It is therefore likely that the change in shape—and particularly the fissure displacement—with age that we observe in this cohort reflects lung tissue micro-structural and elasticity changes, and their force balance with the chest wall. For example, the loss of elastic recoil with age leads to airway closure and increase in RV. This implies that at full-inspiration, the dependent airways will be narrower with resultant changes in airways resistance and compliance impacting on lung volume, and this indirectly acting on the position of the fissures.

In our model, males and females differ in their lung shape for the first three principal shape modes only when size (i.e. lung volume) is included in the analysis, as illustrated in Fig. 4. Mode 4 of the size-exclusive model shows apparent sex differences, however this mode is strongly associated with lung volume. In our dataset, full-inspiratory imaged lung volume is significantly larger ( $p < 0.001$ ) in males ( $6.81 \pm 1.16$  L) than females ( $4.84 \pm 0.85$  L). The relationship between lung volume and mode 4 in Fig. 3d shows overlap between male and female lung size, with smaller male lungs tending towards negative weightings along with most female lungs, and larger female lungs tending towards positive weightings along with most of the males. That is, the mode's apparent association with sex is because of the difference in average volume of male and female lungs. Our general lack of sex differences contrasts with studies of rib cage morphology, which describe a rounder (less elliptical) lung cross-section in females<sup>9,12</sup>. It also contrasts with a recent study that used similar methods to quantify shape differences between males and females<sup>8</sup>. This study found a 7% sex difference in shape in healthy subjects (21 females and 19 males) aged  $51.9 \pm 1.2$  years, and described males as having a 'pyramidal' geometry and females having a 'prismatic' shape. Interestingly, this is qualitatively similar to the extremes of the size-exclusive first principal shape mode in our study, which we found was strongly associated with age. A key difference between our study and others is that we explicitly included the pulmonary fissures. As previously explained, we found that the fissure location changes with age and this change is the same for males and females. It is therefore likely that in our cohort and model the fissure location is a more dominant shape feature than the external lung shape (which could have more-apparent sex differences). A further point is that most previous studies do not normalize for lung volume. This is not important when using ratios of linear dimensions (e.g. antero-posterior diameter/lateral diameter) but is important for 3D analyses such as PCA. This is apparent in the size-inclusive model (Fig. 2), where size is the dominant feature of the first shape mode (Fig. 3) and contributes to significant sex differences in Table 2 and Fig. 4. One methodological difference is that our study used far fewer 'landmark' points than other similar studies (e.g. in comparison to the 12 anatomical and 403 'semi-landmarks' of Torres-Tamayo et al.<sup>8</sup>). That is, our approach has far fewer points at which the surface location is sampled. However, our high-order FE mesh description includes surface curvature and smoothness such that additional intermediate points (between landmarks) could be interpolated as a function of the geometric information at the landmark nodes. That is, both approaches contain similar information, but stored in different forms.

BMI is associated with lung shape in Fig. 6, which is consistent with previous studies<sup>13</sup>. However, the relationship with BMI is only apparent in the third principal shape mode (size-exclusive model). That is, the contribution to shape is quantifiable but qualitatively subtle. Mode 2 of the size-inclusive model and mode 3 of the size-exclusive model have similar strength of relationship with BMI when considering the entire cohort. However, separation by sex shows no relationship for males with BMI when size is not controlled for in the model. The impact of BMI on lung and lobe shape appears to be similar between males and females when we control for volume. Our cohort did not include subjects with  $\text{BMI} > 30 \text{ kg/m}^2$ , hence we had a relatively narrow BMI range in the study. It is possible that the inclusion of subjects with higher BMI would reveal a stronger association with lung shape. Conversely, the inclusion of subjects with high BMI could decrease the strength of the relationship due to air trapping (when imaged supine) restricting lung shape.

One limitation of this study is that the cohort was not ethnically uniform. It is possible that genetic differences between different ethnic groups (e.g. European-American or -New Zealander, Asian-American) is associated with lung shape. We repeated our analysis after excluding five subjects who were identified as of non-European ('caucasian') descent. A two-tailed non-parametric t-test between the results of our study and the new modes revealed no significant differences ( $\alpha > 0.05$ ). It is possible that an analysis using a far larger cohort could reveal differences due to ethnicity, however the small ethnic diversity in our study does not appear to have influenced our results.

The first four principal shape modes only explained 32–36% of the shape variability in this cohort. The relevance of other modes was examined (not shown here) but no associations with physiologic or anthropometric

data were found. It is possible that the addition of more subjects, or examination of a separate cohort, could yield different principal shape modes that explain more shape variation in the first few modes. A leave-one-out analysis of the current cohort shows no significant impact on the shape modes.

Our study only quantified shape in the supine lung at full-inspiration. A change to the upright posture is likely to modify lung shape, but we do not have data to determine how variable this is or if there is an association with age, sex, or BMI. It is possible that the stiffer chest wall in older subjects limits the range of 'container' shape change (except at the diaphragm) more than in young subjects with a compliant chest wall, which would imply a stronger relationship between age and upright shape. Differences in tissue elasticity with age might also be more apparent in the upright lung (with gravity acting over a greater height than when supine), hence exaggerating differences in fissure location.

We only examined static lung shape from imaging acquired at full-inspiration because this is the volume at which clinical imaging is most frequently acquired. Other studies have found sex differences in the change in shape of the 'container' of chest wall and diaphragm from end-expiration to full-inspiration<sup>8</sup>. It is not clear how the presence of fissure shape would influence sex differences for a similar analysis using our model.

This study is a first step towards understanding age-related changes of lung shape that presumably impact on lung function, and establishing lung shape as a potential biomarker of healthy aging. The shape of the lung is straightforward to extract from clinical imaging, such as CT or even from non-ionizing sources such as magnetic resonance imaging. Therefore, if shape is shown to be associated with age, it could provide a straightforward means for staging or stratification of lung disease. We have shown very clear changes in lung shape with age that are not simply the result of changes in the rib-cage geometry. These effects of age almost certainly reflect a complex interplay between changes in lung parenchyma and chest wall compliance. The physiological importance of these relative changes remains to be explored.

Received: 27 May 2020; Accepted: 10 September 2020

Published online: 30 September 2020

## References

- Shi, X. *et al.* A statistical human rib cage geometry model accounting for variations by age, sex, stature and body mass index. *J. Biomech.* **47**, 2277–2285 (2014).
- Kovalev, V., Prus, A. & Vankevich, P. Mining lung shape from X-ray images. in *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, 554–568 (Springer, 2009).
- Wells, R. G. Tissue mechanics and fibrosis. *Biochim. et Biophys. Acta Mol. Basis Dis.* **1832**, 884–890 (2013).
- Lim, S. J. *et al.* Altered thoracic cage dimensions in patients with chronic obstructive pulmonary disease. *Tuberc. Respir. Dis.* **81**, 123–131 (2018).
- Sciurba, F. C. *et al.* A randomized study of endobronchial valves for advanced emphysema. *N. Engl. J. Med.* **363**, 1233–1244 (2010).
- Chan, H.-F., Clark, A. R., Hoffman, E. A., Malcolm, D. T. & Tawhai, M. H. Quantifying normal geometric variation in human pulmonary lobar geometry from high resolution computed tomography. *J. Biomech. Eng.* **137**, 0510101 (2015).
- Garcia-Martinez, D. *et al.* Three-dimensional analysis of sexual dimorphism in ribcage kinematics of modern humans. *Am. J. Phys. Anthropol.* **169**, 348–355 (2019).
- Torres-Tamayo, N. *et al.* 3d analysis of sexual dimorphism in size, shape and breathing kinematics of human lungs. *J. Anat.* **232**, 227–237 (2018).
- Bellemare, F., Jeanneret, A. & Couture, J. Sex differences in thoracic dimensions and configuration. *Am. J. Respir. Crit. Care Med.* **168**, 305–312 (2003).
- Weaver, A. A., Schoell, S. L. & Stitzel, J. D. Morphometric analysis of variation in the ribs with age and sex. *J. Anat.* **225**, 246–261 (2014).
- Minaeizaeim, H. *et al.* Do pulmonary cavity shapes influence lung function? *J. Biomech. Eng.* **141**, 111005 (2019).
- Sverzellati, N. *et al.* Computed tomography measurement of rib cage morphometry in emphysema. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0068546> (2013).
- Boriek, A. M. *et al.* Obesity modulates diaphragm curvature in subjects with and without copd. *Am. J. Physiol. Regul. Integr. Compar. Physiol.* **313**, R620–R629 (2017).
- Suwatanapongched, T., Gierada, D. S., Slone, R. M., Pilgram, T. K. & Tuteur, P. G. Variation in diaphragm position and shape in adults with normal pulmonary function. *Chest* **123**, 2019–2027 (2003).
- Lu, Y.-C. & Untaroiu, C. D. A statistical geometrical description of the human liver for probabilistic occupant models. *J. Biomech.* **47**, 3681–3688 (2014).
- Yeung, S. *et al.* Relationship between lower lumbar spine shape and patient bone metabolic activity as characterised by 18f naf bio-markers. *Comput. Biol. Med.* **116**, 103529 (2020).
- Zhang, H. *et al.* Novel indices for left-ventricular dyssynchrony characterization based on highly automated segmentation from real-time 3-d echocardiography. *Ultrasound Med. Biol.* **39**, 72–88 (2013).
- Swanney, M. P. *et al.* Using the lower limit of normal for the fev1/fvc ratio reduces the misclassification of airway obstruction. *Thorax* **63**, 1046–1051 (2008).
- Guo, J., Fuld, M. K., Alford, S. K., Reinhardt, J. M. & Hoffman, E. A. Pulmonary analysis software suite 9.0: Integrating quantitative measures of function with structural analyses. in *First International Workshop on Pulmonary Image Analysis*, 283–292 (Lulu, New York, 2008).
- Bradley, C., Pullan, A. & Hunter, P. Geometric modeling of the human torso using cubic hermite elements. *Ann. Biomed. Eng.* **25**, 96–111 (1997).
- Terzopoulos, D. Regularization of inverse visual problems involving discontinuities. in *IEEE Transactions on pattern analysis and Machine Intelligence*, 413–424 (1986).
- Zheng, G., Li, S. & Szekely, G. *Statistical Shape and Deformation Analysis: Methods, Implementation and Applications* (Academic Press, Cambridge, 2017).
- Dryden, I. & Mardia, K. *Statistical Analysis of Shape* (Wiley, Hoboken, 1998).
- Cootes, T. F., Taylor, C. J., Cooper, D. H. & Graham, J. Active shape models-their training and application. *Comput. Vis. Image Understand.* **61**, 38–59 (1995).
- Golub, G. H. & Reinsch, C. Singular value decomposition and least squares solutions. in *Linear Algebra*, 134–151 (Springer, 1971).
- Jolliffe, I. *Principal Component Analysis* (Springer, New York, 2002).
- Edge, J., Millard, F., Reid, L. & Simon, G. The radiographic appearances of the chest in persons of advanced age. *Br. J. Radiol.* **37**, 769–774 (1964).

28. Ware, J. H. *et al.* Longitudinal and cross-sectional estimates of pulmonary function decline in never-smoking adults. *Am. J. Epidemiol.* **132**, 685–700 (1990).
29. Gibson, G., Pride, N., Ocain, C. & Quagliato, R. Sex and age differences in pulmonary mechanics in normal nonsmoking subjects. *J. Appl. Physiol.* **41**, 20–25 (1976).
30. Skloot, G. S. The effects of aging on lung structure and function. *Clin. Geriatr. Med.* **33**, 447–457 (2017).

## Acknowledgements

This work was supported by the Evelyn May Steer Estate, Medical Technologies Centre of Research Excellence, Green Lane Education and Research Fund, and an Aotearoa Foundation Postdoctoral Fellowship. Dr. Kelly Burrowes provided valuable proof reading and comments. Dr. Duane Malcolm provided valuable advise on PCA techniques and also provided a Python library (morphpic, <https://morphpic.readthedocs.io/en/latest/>) for Finite Element mesh interpolation.

## Author contributions

M.O. was involved in study conception and design, development of code, analysis, interpretation of data, drafting manuscript, revising manuscript, final approval of manuscript. A.R.C. was involved in the study conception and design, interpretation of data, drafting manuscript, revising manuscript, final approval of manuscript. H.K. contributed to study conception and design, drafting and revising the manuscript. C.K. contributed to data acquisition, revising manuscript and final approval of manuscript. M.W., D.M., E.H. and K.W. were involved in data acquisition, manuscript preparation and final approval of manuscript. M.H.T. contributed to the study conception and design, analysis, interpretation of data, drafting manuscript, revising manuscript, final approval of manuscript. All authors reviewed the manuscript.

## Competing interests

EH is a founder and shareholder of VIDA Diagnostics, a company commercializing lung image analysis software developed, in part at the University of Iowa. The other authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to M.H.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020