Localization and segmentation of optimal slices for chest fat quantification in CT via deep learning

Yi, Jizheng, Udupa, Jayaram K., Tong, Yubing, Anderson, Michaela R., Lederer, David, et al.

Jizheng Yi, Jayaram K. Udupa, Yubing Tong, Michaela R. Anderson, David Lederer, Jason Christie, Drew A. Torigian, "Localization and segmentation of optimal slices for chest fat quantification in CT via deep learning," Proc. SPIE 11317, Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, 113170K (28 February 2020); doi: 10.1117/12.2550377

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

Localization and Segmentation of Optimal Slices for Chest Fat Quantification in CT via Deep Learning

Jizheng Yi[1, 2], Jayaram K. Udupa[2, *], Yubing Tong[2], Michaela R. Anderson[3], David Lederer[3], Jason Christie[4], Drew A. Torigian[2]

[1]College of Computer and Information Engineering, Central South University of Forestry and Technology, Changsha, 410004, China; [2]Medical Image Processing Group, School of Medicine, University of Pennsylvania, Philadelphia, 19104, USA; [3]Department of Medicine, Columbia University Medical Center, New York, NY; [4]Department of Medicine, University of Pennsylvania, Philadelphia, PA.

ABSTRACT

Accurate measurement of subcutaneous adipose tissue (SAT) and visceral adipose tissue (VAT) in the thorax is important for understanding the impact of body composition upon various clinical disorders. The aim of this paper is to explore a practical system for the automatic localization of the axial slices through the thorax at the T7 and T8 vertebral levels in computed tomography (CT), and automatic segmentation of VAT in T7 slice and SAT at T8 slice via deep learning (DL). The methodology mainly consists of two models: the localization model based on AlexNet and the segmentation model based on UNet. For the first one, two slices (T7 and T8) at the middle of the seventh and eighth thoracic vertebrae, respectively, from the full or partial body scan of each patient are automatically detected. For the second one, all the CT images and the associated adipose tissue ground truth segmentations are used for training, where just T7 and T8 slices are tested by the two-label Unet. The datasets from four universities (Penn, Duke, Columbia, and Iowa) are used for training and validation of the models. In the experiments, relevant statistical parameters including Mean Distance (MD), Standard Deviation (SD), True Positive Rate (TPR), and True Negative Rate (TNR) indicate that the proposed algorithm has high reliability and may be useful for fully automated body composition analysis with high accuracy.

Keywords: visceral adipose tissue (VAT), subcutaneous adipose tissue (SAT), thorax, deep learning (DL), localization and segmentation, computed tomography (CT).

1. INTRODUCTION

Many studies have demonstrated that the measurement and analysis of body composition, particularly fat, is important for improved assessment of patients with cancer and other diseases [1]. Usually, human body composition analysis based on CT or magnetic resonance imaging (MRI) includes two main steps – localization of the body region and segmentation of the tissues in the localized region. Specifically, for adipose tissues, there are many ways for both steps [2]. Tong et al. [1] explored the AAR approach for the segmentation of subcutaneous adipose tissue (SAT) and visceral adipose tissue (VAT) components of thoracic fat. An end-to-end segmentation model of SAT and VAT is still scarce, especially for thoracic fat quantification in CT, which is much more challenging than fat quantification in the abdomen on which most papers have focused. This paper proposes a new deep learning (DL)-based strategy for thoracic fat quantification. DL has shown promising performances in computer-aided detection (CAD), localization [3], segmentation [4], classification [5], etc.

Our prior works have shown that the total body region adipose tissue volumes are most correlated with adipose tissue areas derived from axial (cross-sectional) CT images taken at specific vertebral levels – for the abdomen: T12-L1 for SAT and L3-L4 for VAT [6]; and for the thorax: T8 for SAT and T7 for VAT [7]. Based on thoracic SAT and VAT estimation at T8 and T7 levels, a large clinical study has recently demonstrated the utility of these measurements in studying pulmonary graft dysfunction in lung transplant surgery [8, 9]. However, localization of the axial T8 and T7 slices and the segmentation of SAT and VAT components were performed manually [8, 9] in that work. The purpose of this paper is to introduce a system for fully automating both tasks, localizing slice at a specific location and then fat (SAT/VAT) segmentation on the selected slice via DL, so as to support practical production-mode implementation for on-going and future large scale clinical studies.

Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, edited by Andrzej Krol,
Barjor S. Gimi, Proc. of SPIE Vol. 11317, 113170K · © 2020 SPIE · CCC code: 1605-7422/20/$21 · doi: 10.1117/12.2550377

2. MATERIALS AND METHODS

2.1 Image data

This retrospective study was conducted following approval from the Institutional Review Board at the University of Pennsylvania (UPenn) along with a Health Insurance Portability and Accountability Act waiver. Existing unenhanced chest CT image data sets from 457 lung transplant candidates, predominantly with idiopathic pulmonary fibrosis (IPF) or chronic obstructive pulmonary disease (COPD), which were previously acquired as part of a separate prospective study at four lung transplant centers (UPenn, Duke, Columbia, Iowa) were utilized in this study. In the four datasets, the image size is 512×512×(40~700), with a voxel size of 0.7^2-0.97^2×(0.5, 0.6, 1, 1.25, 1.5, 2, 2.5, 3, 3.75, or 5.0) mm^3. The mean age of the patients is 58.0 yrs (±11.7 yrs) with a mean body mass index (BMI) of 26.4 kg/m^2 (±4.3). Chest CT scans had been performed per local clinical protocols during full inspiration.

2.2 General idea of the proposed system

In our approach, we define the thorax body region as extending from an axial level 15 mm superior to the apex of the lungs to a level 5 mm inferior to the base of the lungs. The slice localization model makes use of the AlexNet architecture. Its task is to automatically find axial slices at the mid-T7 (for VAT) and mid-T8 (for SAT) levels in the given full 3D image of the thoracic body region as defined above. We will refer to these slices as T7 and T8 slices. The regions of thoracic VAT and SAT are then delineated in the two localized slices, respectively, through a segmentation model based on UNet.

2.3 Models for localizing T7 and T8 slices

The localization model adopted AlexNet without changing its network structure. As shown in Figure 1, the AlexNet includes 19 layers and mainly contributes to the following aspects: convolution, pooling, and activation. However, in order to make use of the spatial dynamic variations and irregular shapes in contiguous slices, and considering the limitations of tensor dimension and computer hardware, the proposed methodology replaces the original RGB channels with three spatially consecutive slices, and makes the ground truth of the middle slice as the final label for training the network. Similarly, testing data are organized in this way. In the case of a batch size of 6, the optimizer stochastic gradient descent (SGD) in Keras with a learning rate of 1e-04 is used to train and construct the mapping relations between the input three slices and their corresponding label (0 or 1), and another parameter *momentum* for the optimizer is set as 0.9. The loss function, mean squared error (MSE), is defined as follows, where y_{true} is the ground truth of a sample (slice), y_{pred} is its predicted probability, and N is the total number of slices in the input image.

$$\mathrm{MSE} = \left(\sum_N (y_{true} - y_{pred})^2 \right) \Big/ N \quad . \tag{1}$$

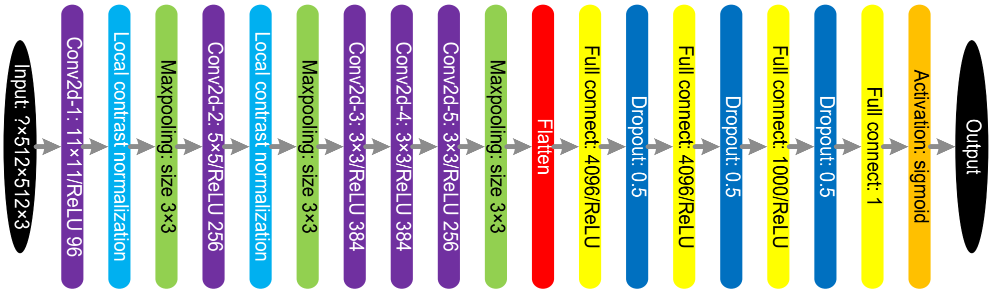


Figure 1. Illustration of the full network architecture for localization. Question mark means the batch size. Five 2D-convolutional layers use filter sizes of 11×11, 5×5, 3×3, 3×3, 3×3, respectively. The filter size for three max-pooling layers is 3×3. It should be noted that the size of a single input sample is 512×512×3 where 3 means three spatially consecutive slices, which is a modification of the traditional AlexNet. The output is a probability associated with one of two classes - T7 slice, not T7 slice, and similarly for the network related to localizing the T8 slice.

2.4 VAT and SAT Segmentation models

As a kind of fully convolutional network (FCN), for the UNet, the input is flexible in terms of size and dimension. Two UNets operate separately for VAT segmentation and SAT segmentation for T7 and T8, respectively. Figure 2 demonstrates the architecture of the proposed segmentation model. Compared with SAT, automated VAT segmentation is much harder because of less clearly definable boundaries and vastly more complex shape. In order to achieve full-automation, the proposed methodology does not crop or rotate the original slice beforehand.

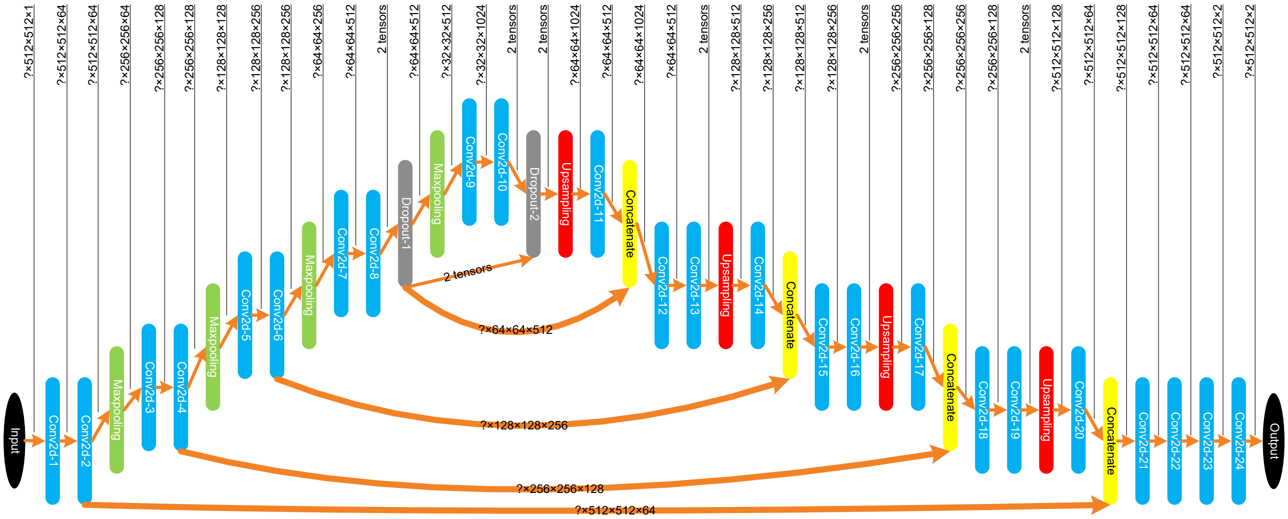


Figure 2. Illustration of the full network architecture for segmentation. Question mark means the batch size. The part of bottleneck is between the down-sampling and up-sampling paths, and is built from simple 2 convolutional layers. All 2D-convolutional layers use a filter size of 3×3, all max-pooling layers and up-sampling filter size is 2×2. The padding for the convolutional operation is the same. It should be noted that the four concatenation layers concatenate with the corresponding uncropped feature map from the down-sampling path.

In the case of a batch size of 2, the optimizer Nadam in Keras with a learning rate of 1e-05 is used to train the input images and their corresponding segmentation binary labels, and the other parameters are set as follows: *beta_1* is 0.9, *beta_2* is 0.999, *schedule_decay* is 0.004. Since there is a binary classification problem in this work, through a *sigmoid* function over the final feature map combined with the loss function named *bce_logdice_loss* (BLL), the weight parameters can be updated.

The *sigmoid* function for the Unet is defined as follows:

$$S(k,p(x,y)) = 1/(1+exp(-a(k,p(x,y)))) \tag{2}$$

where $a(k,p(x,y))$ represents the activation in feature channel k at the voxel position (x, y). The $S(k,p(x,y))$ is between 0 and 1. In the binary classification task, it represents the event probability, i.e., when the output satisfies a certain probability condition, the UNet divides it into positive or negative class. The loss function *bce_logdice_loss* (BLL) is defined as follows.

$$\mathrm{BLL} = \mathrm{BC}(y_{true}, y_{pred}) - log(\mathrm{DC}(y_{true}, y_{pred})) \tag{3}$$

where y_{true} is the ground truth of a sample (here it can be considered as a pixel) and y_{pred} is the predicted probability. The loss functions *binary_crossentropy* (BC) and *dice_coef* (DC) are defined as follows.

$$\mathrm{BC}(y_{true}, y_{pred}) = \sum(y_{true} \cdot log(y_{pred} + o_1) + (1 - y_{true}) log(1 - y_{pred} + o_1)) \tag{4}$$

$$\mathrm{DC}(y_{true}, y_{pred}) = \frac{\kappa + 2\sum y_{true} \times y_{pred}}{\kappa + \sum y_{true} \times y_{true} + \sum y_{pred} \times y_{pred}} \tag{5}$$

where κ is a constant with a value of 1 so that the denominator is not zero, and o_1 is a value approaching zero which makes the logarithmic function tenable.

3. RESULTS

For all data sets and for both slice localization and segmentation tasks, we generated ground truth with expert radiologist guidance.

Several evaluation metrics are utilized in this paper. For the localization task, *Mean Distance* (*MD*), which indicates the average distance between NO_{pred} and NO_{true}, is calculated as

$$MD = \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\left| \mathrm{NO}_{pred} - \mathrm{NO}_{true} \right| \times d \right) \Big/ \mathrm{N}\,, \tag{6}$$

where NO_{pred} is the number of the predicted slice which has the highest score in the tested CT image sequence, NO_{true} is the ground truth slice number of T7 or T8, N is the total number of testing samples, and *d* is the distance (in millimeter (mm)) between adjacent slices which may be different for different patient scans. *Standard Deviation* (*SD*) is calculated for all the distances between NO_{pred} and NO_{true}. In the segmentation experiment, the *Mean* and *SD* for two metrics including *True-Positive Rate* (*TPR*) and *True-Negative Rate* (*TNR*) are presented.

3.1 Evaluation for localization

We selected 357 volumes randomly for training and the remaining 100 for testing. Additionally, the 16-bit image intensities are converted to 12-bit values according to

$$f(x,y) = \begin{cases} 4095 & f(x,y) > 4095 \\ f(x,y) & 0 \le f(x,y) \le 4095 \end{cases}, \tag{7}$$

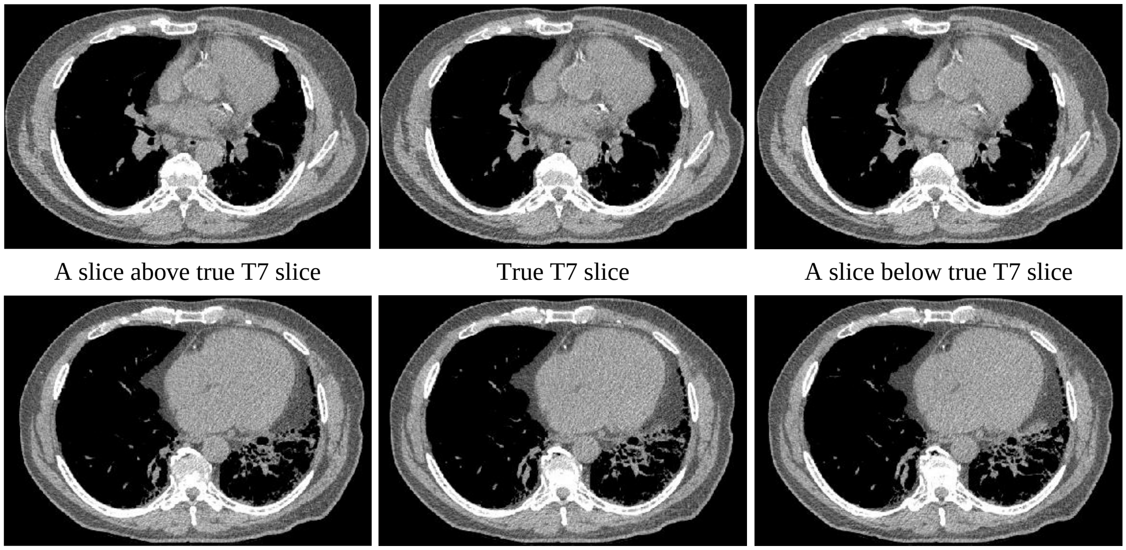
where $f(x,y)$ represents the pixel intensity at position (*x*,*y*). Generally, the inputs of DL networks need to be normalized. The normalization scheme is defined as

$$f(x,y)_{\mathrm{n1}} = \left(f(x,y) - \min\right) / \left(\max - \min\right), \tag{8}$$

where min and max are the minimum and maximum of all voxels in a slice, respectively.

We assess the error in localization via mean distance *MD* in mm, over the tested data sets, of the predicted slice location from the true slice location, separately for T7 slice and T8 slice. As illustrated in Figure 3, the difference between these slices and the nearby slices is very subtle and is quite hard even for experts to localize consistently. In other words, this is a difficult localization problem.

Following the above normalization scheme, the final results for localization are shown in Table 1. As seen in Table 1, for T7 and T8 slices, the *MD*s are 11.2 mm and 10 mm (about 2 slices if the slice interval is 5 mm), and the *SD*s are 12.3 mm and 9.5 mm. Compared with the texture information of slices, localization focuses more on the texture contrast and spatial structure difference of various tissues. The proposed normalization scheme here can effectively improve the contrast and retain the relative texture difference.



A slice above true T7 slice

True T7 slice

A slice below true T7 slice

A slice above true T8 slice

True T8 slice

A slice below true T8 slice

Figure 3. True T7 slice (first row, middle column), true T8 slice (second row, middle column), and subjacent slices (left and right columns), the scanning space (the interval between slices) is 1 mm.

Table 1. Error in localization of T7 and T8 slices expressed by mean distance *MD* in mm and standard deviation *SD*.

Task	Level	Score		Level	Score	
Localization	**T7**	*MD*	**11.2**	**T8**	*MD*	**10**
		SD	**12.3**		*SD*	**9.5**

3.2 Evaluation for segmentation

In order to evaluate the performance of the proposed segmentation model, this study also applies the rule in (7) to convert those slices in the training and testing sets from 16 bits to 12 bits. In addition, another new normalization scheme is used to preprocess training and testing sets, and the segmentation effects among the two proposed normalization schemes are compared.

Assuming that max_min indicates the maximum among all the minimums and min_max indicates the minimum among all the maximums, the distribution range of voxel intensity is compressed according to the following rule:

$$f(x,y)=\begin{cases}\text{min\_max} & f(x,y)>\text{min\_max}\\ f(x,y) & \text{max\_min}\le f(x,y)\le\text{min\_max}\\ \text{max\_min} & f(x,y)>\text{max\_min}\end{cases}. \tag{9}$$

The new normalization scheme is then defined as

$$f(x,y)_{\text{n2}}=\left(f(x,y)-\text{max\_min}\right)/\left(\text{min\_max}-\text{max\_min}\right). \tag{10}$$

Given that the segmentations for SAT and VAT are two distinct tasks, two separate segmentation models are trained. Compared with the continuous distribution and simpler shape of SAT, the occurrence of VAT is much more random and more complex in pattern. Figure 4 presents several examples. One can observe that the segmentation of VAT is very challenging. The evaluation indices of segmentation are shown in Table 2. The reasons that the whole CT image sequence is chosen to train the segmentation model are as follows. Firstly, because of the randomness of VAT distribution, with more samples with different adipose tissue distributions, the training becomes more powerful. Secondly, localization focuses more on the spatial structure differences of various tissues (noting that it does not mean that the texture information is not important), while segmentation focuses more on the specific voxel intensity of adipose tissues. That is to say, even if the slices are different from T7 and T8, they are meaningful for training. As can be seen from Table 2, regardless of SAT or VAT, the new normalization scheme has the better performance with a *mean TPR* of 0.86 and 0.88, while the scheme in (8) has the worse scores with a *mean TPR* of 0.83 and 0.87. The reasons for the distribution of segmentation effects in Table 2 can be summarized as follows: For all slices in the new normalization scheme, the distribution interval of the voxel intensity of adipose tissue does not change after normalization, but for the scheme in (8), the distribution interval is changed, which essentially changes the imaging characteristics of adipose tissue and is disadvantageous to segmentation. These observations emphasize the importance of proper intensity normalization in DL-based image analysis tasks and tailoring this normalization to the task at hand.

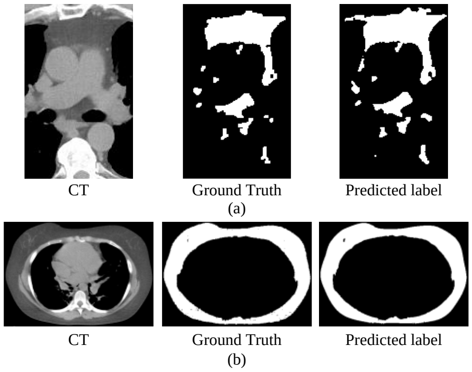


Figure 4. Visual comparison of segmentation results with ground truth segmentations. (a) is one sample of thoracic VAT, and (b) is the result of thoracic SAT from the same sample.

Table 2. Segmentation performance for thoracic adipose tissues at the detected T7 and T8 slices via *True Positive Rate* (*TPR*) and *True Negative Rate* (*TNR*).

Task	Tissue	Score		Scheme in (8)	Scheme in (10)	Tissue	Scheme in (8)	Scheme in (10)
Segmentation	VAT	*TPR*	*Mean*	0.87	**0.88**	SAT	0.83	**0.86**
			SD	0.18	**0.13**		0.11	**0.10**
		TNR	*Mean*	0.996	**0.997**		0.997	**0.996**
			SD	0.003	**0.002**		0.005	**0.004**

4. CONCLUSIONS

In the work described in this paper, a practical system for localization and segmentation of chest CT visceral and subcutaneous adipose tissue components at optimal slice levels via DL has been introduced, which is capable of exhibiting high location and segmentation accuracy. The proposed methodology uses AlexNet and UNet to locate and segment the two key slices at the T7 and T8 levels, respectively. Additionally, this paper innovatively applies three adjacent frames as input data to train the localization model, since their spatial correlation is thought to be helpful for localization. Note that all the experimental results were obtained without pre-specifying perfect experimental materials, but instead are generated from routine clinical practice CT data sets that contain a wide mix of image qualities as encountered in real patient scan data.

REFERENCES

[1] Tong YB, Udupa JK, Wu CY, Pednekar G, Subramanian JR, Lederer DJ, et al., "Fat segmentation on chest CT images via fuzzy models," Proceeding of SPIE, Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling, Vol. 9786, 978609 (2016).

[2] Ida M, Hirata M, Hosoda K, Nakao K, "Abdomen specific bioelectrical impedance analysis (BIA) methods for evaluation of abdominal fat distribution," Nihon Rinsho. 71(2): 262-265 (2013).

[3] Kooi T, Litjens G, van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, et al., "Large scale deep learning for computer aided detection of mammographic lesions," Medical image analysis. 35: 303-12 (2017).

[4] Roth HR, Lu L, Lay N, Harrison AP, Farag A, Sohn A, et al., "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," Medical image analysis. 45: 94-107 (2018).

[5] Xie F, Fan H, Li Y, Jiang Z, Meng R, Bovik A, "Melanoma classification on dermoscopy images using a neural network ensemble model," IEEE Trans Med Imaging. 36(3): 849-58 (2017).

[6] Tong, YB, Udupa JK, Torigian, DA, "Optimization of abdominal fat quantification on CT imaging through use of standardized anatomic space – a novel approach," Medical Physics, 41(6): 063501-1 – 063501-11 (2014).

[7] Tong, YB, Udupa JK, Torigian, DA, Odhner D, Wu CY, Palmer S, Anna Rozenshtein A, Shirk MA, Newell DJ, Porteous M, Diamond JM, Christie JD, Lederer D.J, "Chest fat quantification via CT based on standardized anatomy space in adult lung transplant candidates," PLOS ONE, January 3, 1-16 (2017). DOI:10.1371/journal.pone.0168932.

[8] Anderson MR, Udupa JK, Edwin EA, Diamond JM, Singer JP, Kukreja J, Hays SR, Greenland JR, Ferrante AW, Lippel M, Blue T, McBurnie A, Oyster M, Kalman L, Rushefski M, Wu C, Pednekar G, Liu W, Arcasoy S, Sonett J, D'Ovidio F, Bacchetta M, Newell JD, Torigian D, Cantu E, Farber DL, Giles JT, Tong Y, Palmer S, Ware LB, Hancock WW, Christie JD, Lederer DJ, "Adipose tissue quantification and primary graft dysfunction after lung transplantation: The Lung Transplant Body Composition Study," Journal of Heart and Lung Transplantation, 2019; in press.

[9] Anderson MR, Kolaitis N, Kukreja J, Diamond JM, Palmer S, Arcasoy S, Udupa JK, Christie JD, Lederer DJ, Singer JP, "A non-linear relationship between visceral adipose tissue and frailty in adult lung transplant candidates," American Journal of Transplantation, 2019. DOI: 10.1111/ajt.15525.