Title: Interobserver variability in visual evaluation of thoracic CT scans and comparison with automatic computer measurements of CT lung density

Dr. Mathilde M.W. 14514 Wille mathilde.winkler@gmail.com MD 1, Dr. Laura H. 14515 Thomsen laurahohwu@gmail.com MD 1, Mr. Jens 14516 Petersen jipetersen78@gmail.com 2, Dr. Saher B. 14528 Shaker saher@dadlnet.dk MD 1, Prof. Dr Asger 14517 Dirksen adi@dadlnet.dk MD 1 and Dr. Jesper H. 14518 Pedersen jesper.holst.pedersen@rh.regionh.dk MD 3. 1 Department of Respiratory Medicine, Gentofte University Hospital, Hellerup, Denmark; 2 Department of Computer Science (DIKU), University of Copenhagen, Denmark and 3 Department of Thoracic Surgery, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark.

Body: Introduction – Emphysema is defined by pathology, but is most precisely evaluated in vivo by computed tomography (CT). Aims – were to determine the reproducibility of visual evaluation of emphysema, i.e. the observer variability, and furthermore to compare the visual evaluations to automatic CT lung density measurements, i.e. densitometry. Methods – In a pilot study 60 CT scans were selected from a sample of 3980 CT scans from The Danish Lung Cancer Screening Trial (DLCST). The amount of emphysema in these scans was scored independently by two observers, who were blinded regarding clinical information. The lung was segmented automatically by in-house developed computer software, and the percentage of pixels below -950 HU was used as a surrogate marker for emphysema. The observer variability, as well as the correlation with the lung density measurements, was analysed using Spearman’s rank correlation. Results – Spearman’s correlation coefficient between the two observers was r = 0.85, p < 0.001. However, the combined observations for both observers had a correlation with the CT lung density measurements of r = 0.25, p = 0.05. Conclusions – We found a high degree of interobserver consistency in emphysema grading. However, the agreement with the CT lung density measurement was poor, indicating that the two types of evaluation represent different aspects of emphysema. Most likely, they should be seen as complementary rather than competitive evaluations. Future comparison with physiological tests might elucidate the reason for differences and demonstrate the usefulness of these evaluations.