European Radiology (2018) 28:5111–5120
https://doi.org/10.1007/s00330-018-5516-x

CHEST

# An automated computed tomography score for the cystic fibrosis lung

Guillaume Chassagnon[1,2] · Clémence Martin[3] · Pierre-Régis Burgel[3] · Dominique Hubert[3] · Isabelle Fajac[4] · Nikos Paragios[2] · Evangelia I. Zacharaki[2] · Paul Legmann[1] · Joel Coste[5] · Marie-Pierre Revel[1]

Received: 8 September 2017 / Revised: 16 April 2018 / Accepted: 26 April 2018 / Published online: 4 June 2018


## Abstract

**Objectives** To develop an automated density-based computed tomography (CT) score evaluating high-attenuating lung structural abnormalities in patients with cystic fibrosis (CF).

**Methods** Seventy adult CF patients were evaluated**.** The development cohort comprised 17 patients treated with ivacaftor, with 45 pre-therapeutic and follow-up chest CT scans. Another cohort of 53 patients not treated with ivacaftor was used for validation. CT-density scores were calculated using fixed and adapted thresholds based on histogram characteristics, such as the mode and standard deviation. Visual CF-CT score was also calculated. Correlations between the CT scores and forced expiratory volume in 1 s ($FEV_1$% pred), and between their changes over time were assessed.

**Results** On cross-sectional evaluation, the correlation coefficients between $FEV_1$%pred and the automated scores were slightly lower to that of the visual score in the development and validation cohorts ($R$ = up to -0.68 and -0.61, versus $R$ = -0.72 and $R$ = -0.64, respectively). Conversely, the correlation to $FEV_1$%pred tended to be higher for automated scores ($R$ = up to -0.61) than for visual score ($R$ = -0.49) on longitudinal follow-up. Automated scores based on Mode + 3 SD and Mode +300 HU showed the highest cross-sectional ($R$ = -0.59 to -0.68) and longitudinal ($R$ = -0.51 to -0.61) correlation coefficients to $FEV_1$%pred.

**Conclusions** The developed CT-density score reliably quantifies high-attenuating lung structural abnormalities in CF.

**Key Points**

- *Automated CT score shows moderate to good cross-sectional correlations with $FEV_1$%pred*
- *CT score has potential to be integrated into the standard reporting workflow*

**Keywords** Cystic fibrosis · Tomography spiral computed · Scoring methods · Forced expiratory volume · Image processing computer-assisted

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s00330-018-5516-x) contains supplementary material, which is available to authorized users.

✉ Guillaume Chassagnon
gchassagnon@gmail.com

[1] Radiology Department, Groupe Hospitalier Cochin-Hotel Dieu, Université Paris Descartes, 27 Rue du Faubourg Saint-Jacques, 75014 Paris, France

[2] Centre for Visual Computing, Ecole Centrale Paris, Grande Voie des Vignes, 92290 Chatenay Malabry, France

[3] Pulmonary Department and Adult CF Centre, Groupe Hospitalier Cochin-Hotel Dieu, Université Paris Descartes, 27 Rue du Faubourg Saint-Jacques, 75014 Paris, France

[4] Physiology Department, Groupe Hospitalier Cochin-Hotel Dieu, Université Paris Descartes, 27 Rue du Faubourg Saint-Jacques, 75014 Paris, France

[5] Biostatistics and Epidemiology Department, Groupe Hospitalier Cochin-Hotel Dieu, Université Paris Descartes, 27 Rue du Faubourg Saint-Jacques, 75014 Paris, France

**Abbreviations**

CF Cystic fibrosis
$FEV_1$ Forced expiratory volume in 1 s
MLD Mean lung density

## Introduction

Computed tomography (CT) is the "gold standard" method for assessing structural changes in the lungs of patients with cystic fibrosis (CF). Although CT is more sensitive than spirometry for detecting mild disease progression [1], its role in disease monitoring remains controversial, notably because of the cumulative radiation dose. However, very-low-dose CT protocols delivering doses close to those of combined posteroanterior and lateral chest X-rays were recently described [2–4].

Numerous visual CT scoring methods have been proposed to quantify pulmonary structural changes, the most recent being the CF-CT score [5], derived from the Brody-II score [6]. Visual CT scores have higher sensitivity than pulmonary function tests for the detection of early changes in the CF lung [7, 8], and have been shown to correlate with clinical endpoints such as survival [9], quality of life [1, 10] and the rate of exacerbations [8, 11]. However, visual methods have questionable repeatability, require extensive training and are time-consuming [12], preventing their use in clinical practice, as well as for clinical studies. It is noteworthy that none of the recent trials of ivacaftor and/or lumacaftor, the recently released therapies aimed at improving function of the defective cystic fibrosis transmembrane conductance regulator (CFTR) protein, used CT structural changes as an endpoint [13, 14]. Automated methods could overcome the limitations of visual scoring.

Most CF-related lung morphological changes exhibit higher CT attenuation values than normal lung parenchyma. This is especially the case of bronchial wall thickening and mucus plugging, both of which are dramatically improved by ivacaftor [15].

By analogy with automated quantification of CT low-attenuation areas in emphysema [16], we postulated that automated quantification of high-attenuation structures in CF might objectively reflect disease severity and improvement under the newly released therapies.

The purpose of this study was to develop an automated density-based CT scoring method for evaluating high attenuating lung structural abnormalities in patients with cystic fibrosis (CF). To keep the CT radiation dose as low as possible, we focused on a scoring method only requiring inspiratory images.

## Materials and methods

### Patients

This multicentre two-phase retrospective study, based on adult outpatients with CF, was approved by the Paris Ile de France I ethics committee (ref. 13.652). The need for informed consent was waived.

Patient characteristics are presented in Table 1.

The development phase involved CF patients from six French CF centres who had at least one gating (class 3) mutation in the gene encoding cystic fibrosis transmembrane conductance regulator (CFTR) protein and who were treated with ivacaftor. Patients were eligible if they had had at least two volumetric unenhanced chest CT exams, one before starting ivacaftor and at least one during treatment, plus spirometric measurements performed within 1 month before or after each CT exam, performed from November 2010 through September 2015. A total of 45 CT exams (17 at baseline + 28 during follow-up) from 17 patients were included. The improvement of visual score in the patients of the development cohort has been reported in a previous paper [15].

The validation phase involved an independent cohort of CF patients not treated with ivacaftor. This unpublished cohort included all 53 CF outpatients who had unenhanced chest CT and spirometric measurements performed on the same day at our nationally designated adult CF centre in 2013. These combined examinations were performed as part of their routine follow-up.

Exclusion criteria were the unavailability of CT images reconstructed with a soft kernel or a slice thickness of more than 2 mm, or the use of contrast injection. Only CT examinations performed as routine follow-up and outside an exacerbation phase were taken into account.

**Table 1** Patient characteristics

| | Development cohort | Validation cohort |
|---|---|---|
| Number of patients | 17 | 53 |
| Age (years) | 35 [28-43] | 27 [24-33] |
| Sex, *n* (%) | | |
| Male | 14 (82) | 30 (57) |
| Female | 3 (18) | 23 (43) |
| $FEV_1$ % predicted | 38 [33-77] [a] | 52 [38-68] |

Data are medians with interquartile ranges *in brackets*

The development cohort comprised 45 pre-therapeutic and follow-up chest CT scans of 17 adult CF patients treated with ivacaftor. The validation cohort was composed by 53 adult CF patients not treated with ivacaftor

*$FEV_1$* forced expiratory volume in 1 s

[a] $FEV_1$ values before treatment with ivacaftor

## CT examinations

Depending on the centre, inspiratory chest CT examinations of the whole lungs were obtained with eight different 16-128 multislice CT devices from four different vendors (Somatom Sensation 16 and Somatom Definition DS, Siemens Healthcare, Erlangen, Germany; Lightspeed VCT, Optima CT660 and Discovery HD750, GE Healthcare, Milwaukee, WI, USA; Ingenuity CT, Philips Healthcare, Best, The Netherlands; Aquilion, Toshiba Medical System, Otawara, Japan).

Regarding the acquisition protocol, the tube voltage was 80, 100 or 120 kV, depending on body weight, in 3 (3%), 37 (38%) and 58 (59%) cases, respectively. Images were reconstructed with a slice thickness of 0.625-2 mm, with for each case, at least one set of images reconstructed with a standard reconstruction algorithm. This was true for both pre-treatment and follow-up CT scans. Median dose length product was 148 mGy.cm [interquartile range (IQR) = 128-185 mGy.cm]. Iterative reconstructions were used for some CT examinations. All CT acquisitions targeted the whole lung volume. Being not a standard of care in the participating CF centres, spirometric control of the inspiratory CT acquisition was not performed, and additional expiratory images were not acquired.

## Image analysis

Image analysis was performed by two radiologists (G.C. and M.P.R.) with 5 and 16 years of experience in thoracic imaging, respectively.

CT images were scored blindly to clinical information and to the date of CT. Both automated and visual scores were calculated.

## Automated CT scoring

Quantification was performed after whole-lung segmentation, which consisted of separating the lungs from the chest wall and mediastinum. Two commercially available lung-segmentation software programs were systematically used for each CT, in order to later evaluate the lung segmentation software influence on the score results: Myrian XP-lung software version 1.19.1 (Intrasense, Montpellier, France) and Syngo.via Pulmo CT software version VB1B (Siemens Healthcare, Erlangen, Germany). The lung segmentations were completely automated, with no user interference. The central airways were not included, contrary to intrapulmonary bronchi and pulmonary vessels.

Myrian lung segmentation was improved by systematically applying a sequence of basic morphological operators, in automated mode, and was optionally improved by additional manual editing with 3D tools, especially to include peripheral consolidations when they had been excluded from the initial lung segmentation (Fig. E1, Supplementary material).

Three sets of data were thus obtained for the development cohort:

1. Lung segmentation with XP-lung software without manual editing (segmentation 1)
2. Lung segmentation with XP-lung software with manual editing (segmentation 2)
3. Lung segmentation with Pulmo CT software without manual editing (segmentation 3)

Only segmentation 1 was used for the validation cohort.

Structural changes with high attenuation values (e.g. bronchial wall thickening, mucus plugging/ bronchiolar nodules, consolidation, atelectasis) were quantified with a thresholding method. Fourteen threshold values were tested for their correlation with $FEV_1$%pred. Four fixed threshold values were tested [-300, -400, -500 and -600 Hounsfield units (HU)], as well as ten adapted threshold values, taking into account, for each CT examination, individual histogram features, namely mode—corresponding to the most highly represented attenuation value—mean lung density (MLD) and standard deviation (SD), which are known to be influenced by the inspiratory level. Expiration flattens the density distribution curve and also shifts it towards higher density values [17, 18] (*see* Supplementary material, Fig. E2). We hypothesised that adapted thresholds based on Mode or MLD or integrating SD might compensate for the changes of density distribution related to the level of inspiration.

Various combinations were tested, as shown in Table 2.

The automated CT-density score was expressed as the ratio between the high-attenuating (diseased) lung volume and total lung volume. For instance, a CT-density score value of 4 indicated that 4% of the total lung volume had an attenuation value superior or equal to the threshold.

Additional details on image processing are provided in the online Supplementary material.

## Visual scoring

The visual method used the CF-CT score, except that air trapping was not assessed because expiratory images were not available [5]. The visual scores in the two cohorts were calculated by one of the two radiologists (G.C.) who had received 1 week of intensive training in a reference centre (Lung Analysis, Erasmus Medical Centre, Rotterdam, The Netherlands) in order to achieve good interobserver agreement in scoring.

## Scoring repeatability and time required

Twenty-five CT scans from the development cohort were randomly selected to assess repeatability.

Intraobserver repeatability (G.C.) was evaluated for the visual score and the automated score based on segmentation 2,

which included manual editing, by two reading sessions at a 1-month interval.

Interobserver repeatability (M.P.R.) was evaluated for the automated score, based on segmentation 2. Repeatability was not assessed for the automated scores based on segmentation 1 or 3, which did not include manual editing.

The time required to obtain the visual and automated scores was measured in the development cohort.

### Spirometric measurements

The percentage of predicted forced expiratory volume in 1 s (% predicted, $FEV_1$%pred) measured at the time of CT was used as an endpoint to assess the CT score performance for cross-sectional evaluation in the two cohorts. Longitudinal evaluation was also performed in the development cohort. Changes in $FEV_1$%pred ($\Delta FEV_1$%pred) contemporary to each follow-up CT scan were used to assess the performance of the visual and automated scores for patient follow-up. In patients with more than two follow-up CT scans, the comparison was always performed with the closest preceding simultaneous CT and functional examinations.

### Statistical analysis

Spearman's rank correlation coefficient ($R$) was used to judge the correlation between the CT scores and $FEV_1$%pred, and the correlation between changes in the CT scores ($\Delta$scores) and changes in $FEV_1$%pred ($\Delta FEV_1$%pred). Spearman $R$ values were interpreted as follows: <0.4 = absent to weak correlation, 0.40-0.59 = moderate correlation, 0.60-0.79 = good correlation, >0.8 = strong correlation.

The statistical significance of changes in the CT scores and $FEV_1$%pred values between baseline and last follow-up was evaluated with the Wilcoxon test.

The intraclass correlation coefficient (ICC) and Bland-Altman plots were used to assess repeatability. Excellent repeatability was assumed when the ICC was 0.8 or more.

SAS software version 9.4 (SAS Institute, Cary, NC, USA) was used for all analyses.

## Results

### Patients characteristics

A total of 70 patients were evaluated: 17 in the development cohort and 53 in the validation cohort (Table 1).

The development cohort included a median of two scans per patient (IQR = 2-3) with a median interval of 17 months between consecutive scans (IQR = 12.5-21.7). The median $FEV_1$%pred was 38% before treatment (IQR = 33-77) and increased by a median of +3·9% (predicted) between two consecutive CT scans on treatment (IQR = -3 to +8; range = -17 to +28% predicted).

In the validation cohort, a single CT scan and the corresponding $FEV_1$%pred value were evaluated for each patient. The median $FEV_1$%pred was 52% (IQR = 38-68).

### Cross-sectional correlation in the development cohort

All CT scans (2-4 per patient) with contemporary $FEV_1$%pred values were analysed for cross-sectional correlations between the CT scores and $FEV_1$%pred in the development cohort (Table 2). Using segmentation 1, the median values of mode and SD were -912 HU (IQR = -899 to -912 HU) and 170 HU (IQR = 150-183 HU), respectively. All automated CT-density scores based on adapted thresholds showed moderate to good correlations with $FEV_1$%pred ($R$ = -0.55 to -0.68, $p < 0.001$), while those based on fixed thresholds tended to show weaker correlations ($R$ = -0.43 to -0.57, $p \leq 0.004$). The highest correlation coefficient values were obtained when using Mode + 3 SD as the threshold ($R$ = -0.61 to -0.68, depending on the segmentation method, $p < 0.001$). This was also true when considering only the initial CT for each patient ($R$ = -0.71 to -0.85, $p \leq 0.005$).

The correlation coefficient value with $FEV_1$%pred was slightly higher for the visual score ($R$ = -0.72, $p < 0.001$).

The correlations between the visual and automated CT scores were good to strong ($R$ = 0.68-0.89, $p < 0.001$) (Table E1, Supplementary material). Median CT-density scores when using Mode + 3 SD or Mode + 300 HU as the threshold were 4.0 (IQR = 3.5-4.4) and 7.5 (IQR = 5.4-9.6) respectively, based on segmentation 1 (Table E2, Supplementary material).

### Cross-sectional correlation in the validation cohort

Correlations in the validation cohort were close to those obtained in the development cohort. CT-density scores based on adapted thresholds (Mode + 300, 400 or 500 HU and Mode + 1.5, 2 or 3 SD) showed good correlations with $FEV_1$%pred ($R$ = -0.60 to -0.61, $p < 0.001$) (Fig. 1).

The correlation between the visual CF-CT score and $FEV_1$%pred was also good ($R$ = -0.64, $p < 0.001$).

### Longitudinal correlations in the development cohort

Longitudinal correlations between the $\Delta$CT-density score and $\Delta FEV_1$%pred based on the 28 follow-up CT scans in the development cohort are summarised in Table 3. The $\Delta$CT-density scores obtained with adapted thresholds tended to better correlate with $\Delta FEV_1$%pred than did scores based on fixed thresholds. The highest correlation coefficient values were obtained with $\Delta$(Mode + 3 SD) ($R$ = -0.55 to -0.61, $p \leq$

**Table 2** Cross-sectional correlations between CT-density scores and $FEV_1$%pred

| | Development cohort | | | | | | Validation cohort | |
|---|---|---|---|---|---|---|---|---|
| | Segmentation 1 (17 patients, *n* = 45 CT) | | Segmentation 2 (17 patients, *n* = 45 CT) | | Segmentation 3 (17 patients, *n* = 38 CT [a]) | | Segmentation 1 (53 patients, *n* = 53 CT) | |
| Threshold | *R* | *p* value | *R* | *p* value | *R* | *p* value | *R* | *p* value |
| Fixed thresholds | | | | | | | | |
| (-) 300 HU | -0.54 | <0.001 | -0.48 | <0.001 | -0.46 | 0.004 | -0.51 | <0.001 |
| (-) 400 HU | -0.57 | <0.001 | -0.52 | <0.001 | -0.50 | 0.002 | -0.53 | <0.001 |
| (-) 500 HU | -0.57 | <0.001 | -0.53 | <0.001 | -0.51 | 0.001 | -0.54 | <0.001 |
| (-) 600 HU | -0.56 | <0.001 | -0.53 | <0.001 | -0.51 | 0.001 | -0.43 | 0.001 |
| Adapted thresholds | | | | | | | | |
| MLD +2.5 SD | -0.61 | <0.001 | -0.58 | <0.001 | -0.56 | <0.001 | -0.56 | <0.001 |
| MLD +2 SD | -0.64 | <0.001 | -0.61 | <0.001 | -0.60 | <0.001 | -0.59 | <0.001 |
| MLD +1.5 SD | -0.57 | <0.001 | -0.64 | <0.001 | -0.57 | <0.001 | -0.59 | <0.001 |
| Mode +500 HU | -0.63 | <0.001 | -0.59 | <0.001 | -0.55 | <0.001 | -0.61 | <0.001 |
| Mode +400 HU | -0.64 | <0.001 | -0.61 | <0.001 | -0.58 | <0.001 | -0.60 | <0.001 |
| Mode +300 HU | -0.64 | <0.001 | -0.62 | <0.001 | -0.59 | <0.001 | -0.60 | <0.001 |
| Mode +3 SD | -0.68 | <0.001 | -0.64 | <0.001 | -0.61 | <0.001 | -0.61 | <0.001 |
| Mode + 2 SD | -0.66 | <0.001 | -0.65 | <0.001 | -0.64 | <0.001 | -0.61 | <0.001 |
| Mode + 1.5 SD | -0.65 | <0.001 | -0.64 | <0.001 | -0.64 | <0.001 | -0.60 | <0.001 |
| Mode + 1 SD | -0.67 | <0.001 | -0.68 | <0.001 | -0.67 | <0.001 | -0.56 | <0.001 |

The development cohort comprised 45 pre-therapeutic and follow-up chest CT scans of 17 adult CF patients treated with ivacaftor. The validation cohort was composed by 53 adult CF patients not treated with ivacaftor

For comparison, the correlation between the visual CF-CT score and $FEV_1$%pred was R = -0.72 ($p < 0.001$) in the development cohort and R = -0.64 ($p < 0.001$) in the validation cohort

*CT* computed tomography, *HU* Hounsfield units, *MLD* mean lung density, *SD* standard deviation

[a] Lung segmentation using Pulmo-CT software (segmentation 3) failed in 7 of the 45 CT examinations, in 6 patients (*see* Supplementary material)

0.008) and Δ(Mode + 300 HU) ($R$ = -0.51 to -0.60, $p \leq 0.008$) depending on the segmentation method.

By contrast, the Δvisual CF-CT score showed only a moderate correlation with Δ$FEV_1$%pred ($R$ = -0.49, $p$ = 0.008).

FEV1%pred and visual CF-CT score showed discordant evolution in 32% of cases overall, with either improvement of FEV1%pred but increase of CF-CT score or worsening of FEV1%pred but decrease of CF-CT score. Changes in the CT-density score and $FEV_1$%pred, with automated scores based on Mode + 3 SD or Mode + 300 HU, were discordant in 24% (5/21) to 32% (9/28) of cases, depending on the segmentation method used (Fig. 2).

$FEV_1$%pred values improved significantly between the pretherapeutic examination and last follow-up on ivacaftor [+6.3% predicted; 95% confidence interval (95% CI), 0-14.5; $p$ = 0.045]. Significant improvements were also noted in the visual CF-CT score ($p$ = 0.016) and in the CT-density scores based on Mode + 3 SD and Mode + 300 HU when calculated from segmentations 1 and 2 ($p < 0.05$). The CT-density score calculated from segmentation 3 also improved, but not significantly ($p > 0.05$) (Table E3, Supplementary material).

## Repeatability and time required

The intraobserver repeatability of both the visual CF-CT score and the automated CT-density score based on segmentation 2 was excellent (ICC > 0.8). However, intraobserver repeatability was higher with the automated score (ICCs ≥ 0.947), regardless of the threshold.

The interobserver repeatability of the automated score was also excellent (ICC, 0.947-0.997) (Table E4, Supplementary material)

The average time required to obtain the automated scores was respectively 2.0 ± 0.5 and 0.8 ± 0.2 min when based on segmentation 1 and 3 (no manual editing). It was 6.6 ± 2.4 min based on segmentation 2 with manual editing. The visual CF-CT score took an average of 17.8 ± 7.8 min.

## Discussion

We report a good cross-sectional correlation between a new automated density-based CT score for high-attenuating lung structural abnormalities and $FEV_1$%pred in adults with CF.

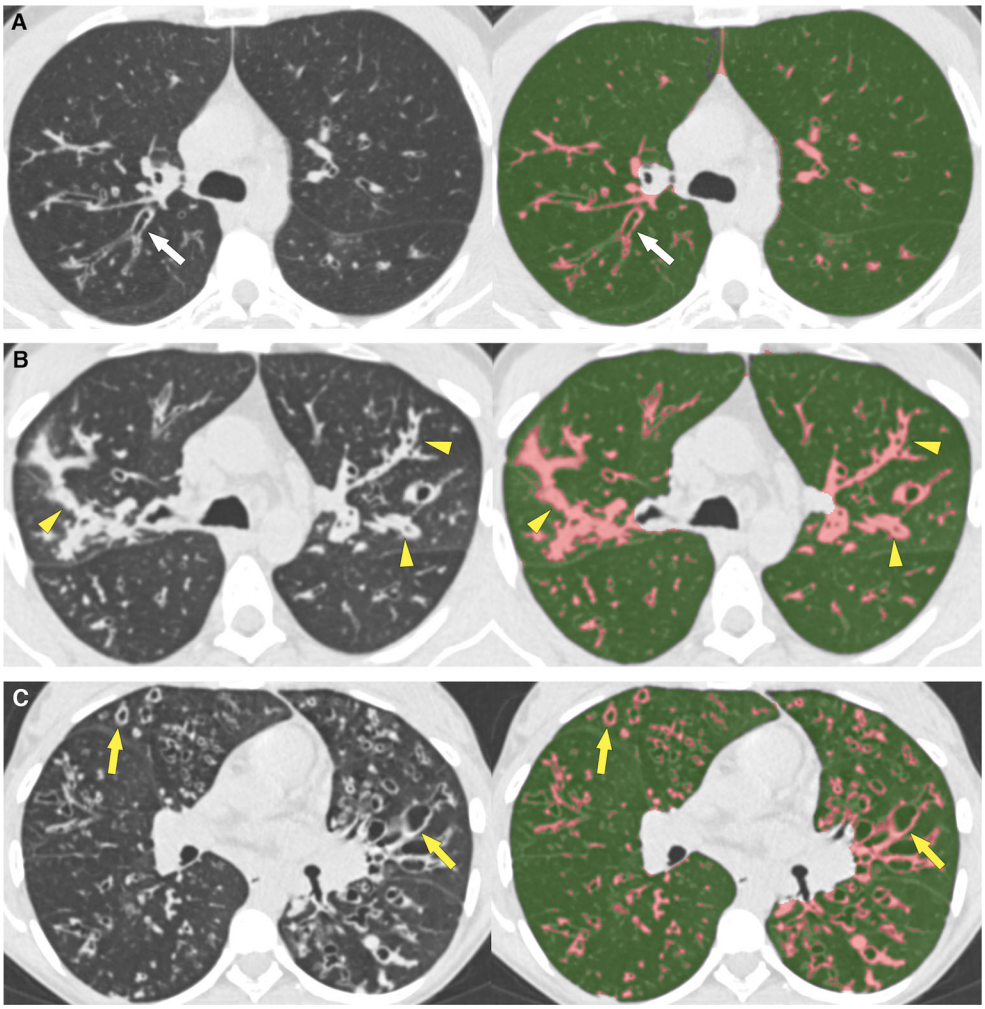


**Fig. 1** Automated CT scoring in the validation cohort of patients with various disease severities. **a** Axial CT image in a patient with mild lung disease ($FEV_1$%pred =77%). Bronchiectasis and bronchial wall thickening are seen in the posterior segment of the right upper lobe (*white arrow*). These lesions are included in areas of high attenuation (*pink areas*). Scoring with Mode + 300 HU yielded a CT-density score of 4.4. This means that 4.4% of the total lung volume had an attenuation value superior or equal to mode (-899 HU) + 300 HU. **b** Axial CT image in a patient with moderate disease ($FEV_1$%pred =56%) shows bilateral mucus plugging (*yellow arrowheads*). The CT-density score was 9.8. **c** Axial CT image in a patient with severe disease ($FEV_1$%pred =31%) shows diffuse bronchiectasis and bronchial wall thickening (*yellow arrows*). The CT-density score was 14.5

The automated score tended to better correlate with changes in $FEV_1$%pred among patients treated with ivacaftor than did the visual score. The developed score was validated in a larger independent cohort of unselected adult CF patients, with similar results for two different commercially available lung segmentation software.

Although the CT attenuation of pulmonary structural abnormalities associated with CF differs from that of normal lung, few attempts have been made to quantify CF-related pulmonary lesions in terms of CT density distribution. Quantification of low-attenuation areas in order to assess air trapping showed a good correlation with residual volume (RV) and maximum mid-expiratory flow [17, 19–21]. However, most of these correlations were weak, and the need for both expiratory and inspiratory images raise concerns as to the radiation dose.

Our aim was to develop a score only requiring inspiratory images, to keep the CT radiation dose as low as possible. Indeed, the report by O'Connell et al. [22] highlighted the increasing exposure to ionising radiation to patients with CF, being mainly attributable to CT scanning. This is the reason why we exclude routinely performing expiratory CT in our centre.

Even though Palumbo et al. [23] described a flattening of the lung parenchyma CT density distribution in CF patients with severe lung impairment, we are first to propose an automated CT scoring method based on quantification of high-attenuating lung structures.

Quantitative evaluation of airway disease in CF has been previously performed with other approaches, focusing on the analysis of airway size and geometry [24–26]. Wielpütz et al. [24], performing automated airway analysis, reported high negative correlations between enlarged airway dimensions and $FEV_1$ in adults.

Quantifying high attenuation lung structures, instead of irreversible bronchial lumen dilatation allows monitoring changes under the newly developed targeted therapies and this explains why the score correlated well with $FEV_1$ on longitudinal follow-up. Normal high-attenuating lung structures such as pulmonary vessels are also included. However, differences in pulmonary vessel volume among patients had probably little influence on score variations compared to those due to the bronchial disease.

Compared to fixed thresholds, adapted thresholds taking into account CT acquisition-dependent variations in lung density

**Table 3** Correlations between longitudinal changes in CT-density scores (Δ scores) and in $FEV_1$%pred (Δ $FEV_1$%pred) in the development cohort

| | Segmentation 1 ($n$ = 28) | | Segmentation 2 ($n$ = 28) | | Segmentation 3 ($n$ = 21)[a] | |
|---|---|---|---|---|---|---|
| Thresholds | $R$ | $p$ value | $R$ | $p$ value | $R$ | $p$ value |
| Fixed thresholds | | | | | | |
| Δ (-) 300 HU | -0.40 | 0.035 | -0.40 | 0.035 | -0.37 | 0.098 |
| Δ (-) 400 HU | -0.39 | 0.039 | -0.49 | 0.035 | -0.38 | 0.094 |
| Δ (-) 500 HU | -0.37 | 0.050 | -0.44 | 0.009 | -0.39 | 0.081 |
| Δ (-) 600 HU | -0.39 | 0.043 | -0.44 | 0.021 | -0.36 | 0.104 |
| Adapted thresholds | | | | | | |
| Δ(MLD + 2.5 SD) | -0.52 | 0.005 | -0.53 | 0.004 | -0.51 | 0.017 |
| Δ(MLD + 2 SD) | -0.49 | 0.008 | -0.52 | 0.005 | -0.50 | 0.022 |
| ΔMLD + 1.5 SD | -0.46 | 0.014 | -0.51 | 0.005 | -0.50 | 0.022 |
| Δ(Mode + 500 HU) | -0.50 | 0.006 | -0.56 | 0.002 | -0.46 | 0.038 |
| Δ(Mode + 400 HU) | -0.53 | 0.004 | -0.57 | 0.002 | -0.49 | 0.023 |
| Δ(Mode + 300 HU) | -0.51 | 0.006 | -0.60 | <0.001 | -0.56 | 0.008 |
| Δ(Mode + 3 SD) | -0.55 | 0.002 | -0.61 | <0.001 | -0.56 | 0.008 |
| Δ(Mode + 2 SD) | -0.45 | 0.016 | -0.58 | 0.002 | -0.54 | 0.012 |
| Δ(Mode + 1.5 SD) | -0.33 | 0.090 | -0.46 | 0.014 | -0.51 | 0.018 |
| Δ(Mode + 1 SD) | -0.26 | 0.187 | -0.31 | 0.112 | -0.49 | 0.024 |

*CT* computed tomography, *HU* Hounsfield unit, *MLD* mean lung density, *SD* standard deviation

[a] Because 7 segmentations failed with Pulmo-CT software, only 21 of 28 interscan changes could be evaluated with segmentation 3

distribution improved the correlation with $FEV_1$%pred. Indeed, lung density is known to be influenced by the level of inspiration, the scanning parameters, the quality of CT calibration and even the CT device manufacturer [27–30]. Whereas expiration flattens the density distribution curve with a shift towards higher density values [17, 18], most other parameters mainly shift the density distribution towards higher or lower values [30, 31]. Various correction methods have been proposed to quantify emphysema and air trapping, but none had previously been proposed for quantifying high-attenuating structures [17, 30–32]. We suspected that adapting thresholds based on Mode or MLD might partially correct the shift in density distribution, whereas the use of SD might partially correct the flattening of the density distribution due to a low level of inspiration. The benefit of adapting the threshold was supported by the stronger $FEV_1$%pred correlations obtained with adapted thresholds. Mode + 3 SD or Mode + 300 HU offered the best compromise, optimising the correlations in both the cross-sectional and longitudinal analyses.

Adapted thresholds based on Mode also allow taking into account the attenuation variations due to various tube voltage setting, since Mode represents the most frequent attenuation value observed in the lung histogram. De Lavernhe et al. [33] studied the correlation between lung function and other histogram characteristics and found a moderate cross-sectional correlation of Log-iKurtosis with $FEV_1$%pred in CF patients. This approach is different from ours, which is based on the quantification of high-attenuating structures.

The cross-sectional correlation of our automated score with $FEV_1$%pred was slightly weaker than that of the visual CF-CT score in both the development and validation cohorts in our study, still remaining in the upper range of correlation values previously reported for visual scores (-0.33 to -0.78) [6, 34–36]. In contrast, as already mentioned, longitudinal correlation with $FEV_1$%pred was slightly better for the automated score. This may be because subtle changes are more difficult to assess visually than to detect by objective measurements of attenuation, or because visual score also takes into account irreversible changes such as bronchial dilatation.

The rates of discordance with changes in $FEV_1$%pred were similar with the visual and automated CT scores, and in the range of those previously reported (31%, vs 24-32% in our series) [37].

Median CT-density scores calculated with lung segmentation 1 or 2 improved significantly in the CF patients treated with ivacaftor. An improvement was also noted for scores obtained with segmentation 3, even though statistical significance was reached for fewer thresholds. The similar results obtained with and without manual editing of small segmentation errors (segmentation 1 and 2) demonstrate the process can be fully automated.

Automated CT scoring has the potential to overcome the main limitations of visual scoring methods, one of which is the training required to attain and maintain adequate repeatability. Calder et al. [12] suggested that central reading by highly trained scorers might be an option for CT research studies [20].

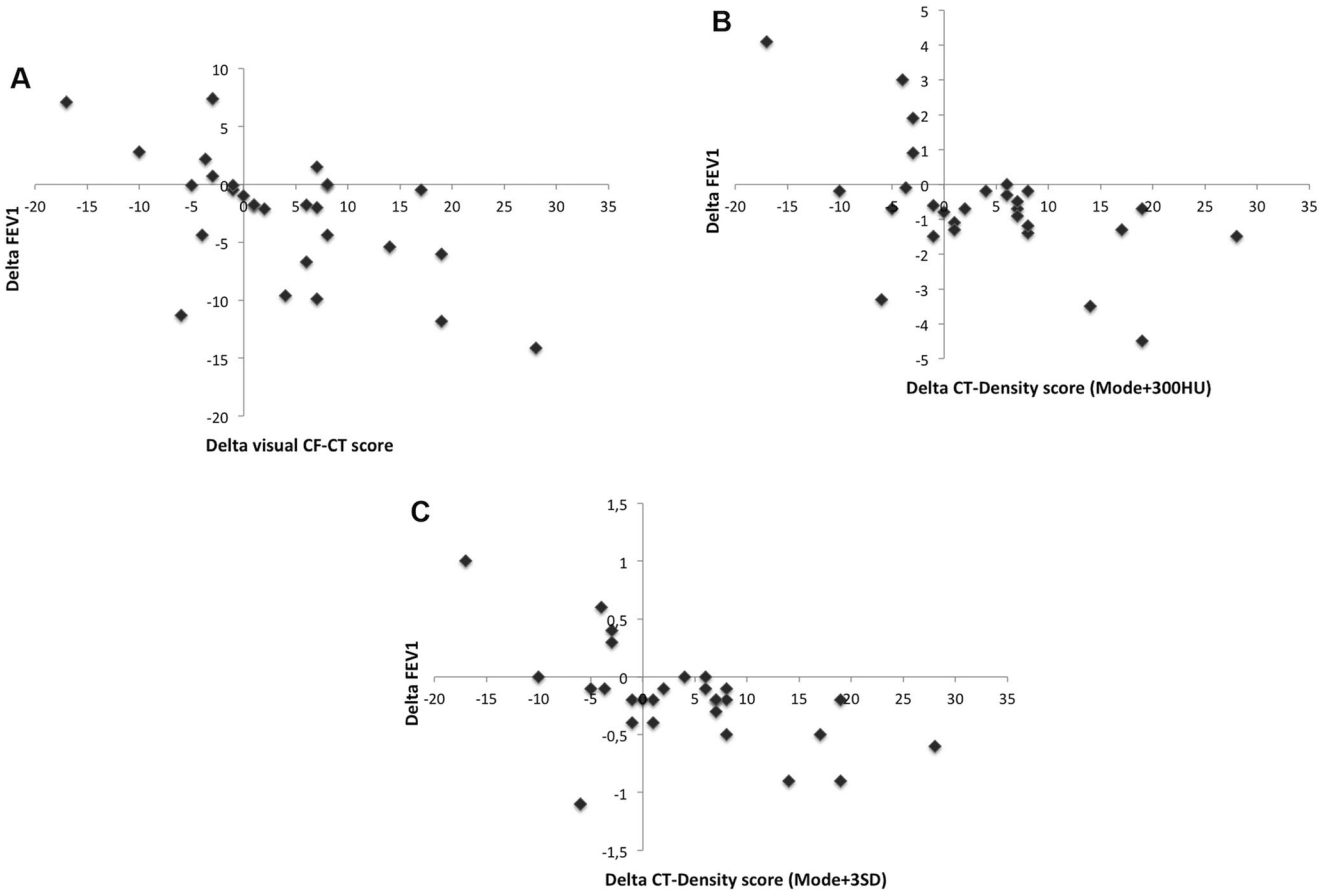


**Fig. 2** Development cohort: longitudinal changes in $FEV_1$ % predicted versus changes in **a** the visual CF-CT score, and in the CT-density scores based on **b** Mode + 300 HU and **c** Mode + 3 SD with segmentation 2. The changes are concordant in the left upper quadrant (improvement in both $FEV_1$%pred and the CT score) and in the right lower quadrant (worsening of both $FEV_1$%pred and the CT score)

Automated methods do not suffer from repeatability issues and, even in case of manual editing, repeatability was excellent regardless of the threshold (ICC > 0.90). However, we only evaluated repeatability based on the same set of scans; thus, only the influence of variation in the manual edited segmentations was evaluated, no other potential variation factors, which would require repeating scans within a short time frame.

Lastly, the time required for automated scoring was far shorter than for visual scoring and is compatible with daily practice.

Our study has several limitations. First, owing to the retrospective and multicentre design, the scanning techniques were not standardised. Contrary to visual scoring, density-based automated scoring is highly dependent on the scanning technique. The CT examinations used in our development cohort often had different slice thicknesses and/or tube voltages. This heterogeneity may have been detrimental for our scoring method. However, this corresponds to routine practice and has not prevented from obtaining good correlations to the pulmonary function. The use of standardised scanning protocols and calibrated breath-holds, as previously suggested [38, 39], would probably even improve the performance of our automated CT scoring method. Another limitation is that the number of CT scans per patient in the development cohort was uneven. However, our results were confirmed in the validation cohort, with only one CT scan and one $FEV_1$%pred value per patient. A third limitation is that $FEV_1$%pred was the only CF outcome measure. However $FEV_1$%pred measurement is the only surrogate for mortality to be considered as primary endpoint by the European Medicines Agency for clinical trials in CF [40]. Correlations with other clinical endpoints such as quality of life, the exacerbation rate, and survival are important for the validation of chest CT as a surrogate outcome and should be assessed in further studies [39].

In conclusion, our results demonstrate that CT density-based automated scoring of lung structural abnormalities is feasible in CF patients. The use of adapted thresholds such as Mode + 3 SD or Mode + 300 HU to quantify high-attenuating CF-related lesions yielded a good correlation with $FEV_1$%pred, in both cross-sectional and longitudinal analyses. This scoring method, validated in a second, independent cohort, is much less time-consuming than visual scoring and could prove suitable both in daily practice and as an objective endpoint for clinical trials.

**Funding** This study has received funding by the patient association "Vaincre la Mucoviscidose".

## Compliance with ethical standards

**Guarantor** The scientific guarantor of this publication is Marie-Pierre Revel

**Conflict of interest** The authors of this manuscript declare relationships with the following companies:
• Pierre-Régis Burgel, Dominique Hubert and Isabelle Fajac have received personal fees from Vertex Pharmaceuticals, outside the submitted work.
• Guillaume Chassagnon and Marie-Pierre Revel have made a patent application for the image analysis method presented in this article.

The other authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

**Statistics and biometry** Joel Coste has significant statistical expertise.

**Informed consent** The need for informed consent was waived, in accordance with French rules for retrospective studies

**Ethical approval** The study was approved by the Paris Ile de France I ethics committee (ref 13.652).

**Study subjects or cohorts overlap** The 17 study subjects of the development cohort have been previously reported in Chassagnon et al. [15].

**Methodology**
• retrospective
• observational
• multicentre study

## References

1. Tepper LA, Caudri D, Utens EMWJ et al (2014) Tracking CF disease progression with CT and respiratory symptoms in a cohort of children aged 6-19 years. Pediatr Pulmonol 49:1182–1189. https://doi.org/10.1002/ppul.22991
2. O'Connor OJ, Vandeleur M, McGarrigle AM et al (2010) Development of low-dose protocols for thin-section CT assessment of cystic fibrosis in pediatric patients. Radiology 257:820–829. https://doi.org/10.1148/radiol.10100278
3. Miéville FA, Berteloot L, Grandjean A et al (2013) Model-based iterative reconstruction in pediatric chest CT: assessment of image quality in a prospective study of children with cystic fibrosis. Pediatr Radiol 43:558–567. https://doi.org/10.1007/s00247-012-2554-4
4. Ernst CW, Basten IA, Ilsen B et al (2014) Pulmonary disease in cystic fibrosis: assessment with chest CT at chest radiography dose levels. Radiology 273:597–605. https://doi.org/10.1148/radiol.14132201
5. Wainwright CE, Vidmar S, Armstrong DS et al (2011) Effect of bronchoalveolar lavage-directed therapy on *Pseudomonas aeruginosa* infection and structural lung injury in children with cystic fibrosis: a randomized trial. JAMA 306:163–171. https://doi.org/10.1001/jama.2011.954
6. Brody AS, Klein JS, Molina PL et al (2004) High-resolution computed tomography in young patients with cystic fibrosis: distribution of abnormalities and correlation with pulmonary function tests. J Pediatr 145:32–38. https://doi.org/10.1016/j.jpeds.2004.02.038
7. Cademartiri F, Luccichenti G, Palumbo AA et al (2008) Predictive value of chest CT in patients with cystic fibrosis: a single-center 10-year experience. AJR Am J Roentgenol 190:1475–1480. https://doi.org/10.2214/AJR.07.3000
8. Brody AS, Sucharew H, Campbell JD et al (2005) Computed tomography correlates with pulmonary exacerbations in children with cystic fibrosis. Am J Respir Crit Care Med 172:1128–1132. https://doi.org/10.1164/rccm.200407-989OC
9. Loeve M, Hop WCJ, de Bruijne M et al (2012) Chest computed tomography scores are predictive of survival in patients with cystic fibrosis awaiting lung transplantation. Am J Respir Crit Care Med 185:1096–1103. https://doi.org/10.1164/rccm.201111-2065OC
10. Kilcoyne A, Lavelle LP, McCarthy CJ et al (2016) Chest CT abnormalities and quality of life: relationship in adult cystic fibrosis. Ann Transl Med 4:87. https://doi.org/10.21037/atm.2016.03.08
11. Loeve M, Gerbrands K, Hop WC et al (2011) Bronchiectasis and pulmonary exacerbations in children and young adults with cystic fibrosis. Chest 140:178–185. https://doi.org/10.1378/chest.10-1152
12. Calder AD, Bush A, Brody AS, Owens CM (2014) Scoring of chest CT in children with cystic fibrosis: state of the art. Pediatr Radiol 44:1496–1506. https://doi.org/10.1007/s00247-013-2867-y
13. Ramsey BW, Davies J, McElvaney NG et al (2011) A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. N Engl J Med 365:1663–1672. https://doi.org/10.1056/NEJMoa1105185
14. Wainwright CE, Elborn JS, Ramsey BW et al (2015) Lumacaftor–Ivacaftor in patients with cystic fibrosis homozygous for Phe508del CFTR. N Engl J Med 373:220–231. https://doi.org/10.1056/NEJMoa1409547
15. Chassagnon G, Hubert D, Fajac I et al (2016) Long-term computed tomographic changes in cystic fibrosis patients treated with ivacaftor. Eur Respir J 48:248–252. https://doi.org/10.1183/13993003.01918-2015
16. Gevenois PA, De Vuyst P, de Maertelaer V et al (1996) Comparison of computed density and microscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med 154:187–192. https://doi.org/10.1164/ajrccm.154.1.8680679
17. Goris ML, Zhu HJ, Blankenberg F et al (2003) An automated approach to quantitative air trapping measurements in mild cystic fibrosis. Chest 123:1655–1663
18. Goris ML, Zhu HJ, Robinson TE (2007) A critical discussion of computer analysis in medical imaging. Proc Am Thorac Soc 4:347–349. https://doi.org/10.1513/pats.200701-014HT
19. Bonnel A-S, Song SM-H, Kesavarju K et al (2004) Quantitative air-trapping analysis in children with mild cystic fibrosis lung disease. Pediatr Pulmonol 38:396–405. https://doi.org/10.1002/ppul.20091
20. DeBoer EM, Swiercz W, Heltshe SL et al (2014) Automated CT scan scores of bronchiectasis and air trapping in cystic fibrosis. Chest 145:593–603. https://doi.org/10.1378/chest.13-0588
21. Loeve M, Rosenow T, Gorbunova V et al (2015) Reversibility of trapped air on chest computed tomography in cystic fibrosis patients. Eur J Radiol 84:1184–1190. https://doi.org/10.1016/j.ejrad.2015.02.011
22. O'Connell OJ, McWilliams S, McGarrigle A et al (2012) Radiologic imaging in cystic fibrosis: cumulative effective dose and changing trends over 2 decades. Chest 141:1575–1583. https://doi.org/10.1378/chest.11-1972
23. Palumbo AA, Luccichenti G, Belgrano M et al (2007) Three-dimensional quantitative assessment of lung parenchyma in cystic fibrosis: preliminary results. Radiol Med 112:21–30. https://doi.org/10.1007/s11547-007-0117-9
24. Wielpütz MO, Eichinger M, Weinheimer O et al (2013) Automatic airway analysis on multidetector computed tomography in cystic

fibrosis: correlation with pulmonary function testing. J Thorac Imaging 28:104–113. https://doi.org/10.1097/RTI.0b013e3182765785
25. Montaudon M, Berger P, Cangini-Sacher A et al (2007) Bronchial measurement with three-dimensional quantitative thin-section CT in patients with cystic fibrosis. Radiology 242:573–581. https://doi.org/10.1148/radiol.2422060030
26. de Jong PA, Nakano Y, Hop WC et al (2005) Changes in airway dimensions on computed tomography scans of children with cystic fibrosis. Am J Respir Crit Care Med 172:218–224. https://doi.org/10.1164/rccm.200410-1311OC
27. Robinson PJ, Kreel L (1979) Pulmonary tissue attenuation with computed tomography: comparison of inspiration and expiration scans. J Comput Assist Tomogr 3:740–748
28. Levi C, Gray JE, McCullough EC, Hattery RR (1982) The unreliability of CT numbers as absolute values. AJR Am J Roentgenol 139:443–447. https://doi.org/10.2214/ajr.139.3.443
29. Yuan R, Mayo JR, Hogg JC et al (2007) The effects of radiation dose and CT manufacturer on measurements of lung densitometry. Chest 132:617–623. https://doi.org/10.1378/chest.06-2325
30. Kemerink GJ, Lamers RJ, Thelissen GR, van Engelshoven JM (1995) Scanner conformity in CT densitometry of the lungs. Radiology 197:749–752. https://doi.org/10.1148/radiology.197.3.7480750
31. Perhomaa M, Jauhiainen J, Lähde S et al (2000) CT lung densitometry in assessing intralobular air content. An experimental and clinical study. Acta Radiol 41:242–248
32. Parr DG, Stoel BC, Stolk J et al (2004) Influence of calibration on densitometric studies of emphysema progression using computed tomography. Am J Respir Crit Care Med 170:883–890. https://doi.org/10.1164/rccm.200403-326OC
33. de Lavernhe I, Le Blanche A, Dégrugilliers L et al (2015) CT density distribution analysis in patients with cystic fibrosis: correlation with pulmonary function and radiologic scores. Acad Radiol 22:179–185. https://doi.org/10.1016/j.acra.2014.09.003
34. Shah RM, Sexauer W, Ostrum BJ et al (1997) High-resolution CT in the acute exacerbation of cystic fibrosis: evaluation of acute findings, reversibility of those findings, and clinical correlation. AJR Am J Roentgenol 169:375–380. https://doi.org/10.2214/ajr.169.2.9242738
35. Dorlöchter L, Nes H, Fluge G, Rosendahl K (2003) High resolution CT in cystic fibrosis—the contribution of expiratory scans. Eur J Radiol 47:193–198
36. de Jong PA, Ottink MD, Robben SGF et al (2004) Pulmonary disease assessment in cystic fibrosis: comparison of CT scoring systems and value of bronchial and arterial dimension measurements. Radiology 231:434–439. https://doi.org/10.1148/radiol.2312021393
37. de Jong PA, Lindblad A, Rubin L et al (2006) Progression of lung disease on computed tomography and pulmonary function tests in children and adults with cystic fibrosis. Thorax 61:80–85. https://doi.org/10.1136/thx.2005.045146
38. Mueller KS, Long FR, Flucke RL, Castile RG (2010) Volume-monitored chest CT: a simplified method for obtaining motion-free images near full inspiratory and end expiratory lung volumes. Pediatr Radiol 40:1663–1669. https://doi.org/10.1007/s00247-010-1671-1
39. Loeve M, Krestin GP, Rosenfeld M et al (2013) Chest computed tomography: a validated surrogate endpoint of cystic fibrosis lung disease? Eur Respir J 42:844–857. https://doi.org/10.1183/09031936.00051512
40. Committee for Medicinal Products for Human Use (2009) Guideline on the clinical development of medicinal products for the treatment of cystic fibrosis. European Medicines Agency, London