Respiratory Research

RESEARCH

Open Access

Reliability of crackles in fibrotic interstitial lung disease: a prospective, longitudinal study

Giacomo Sgalla[1,2*], Jacopo Simonetti[1], Arianna Di Bartolomeo[1], Tonia Magrì[1], Bruno Iovene[2], Giuliana Pasciuto[2], Ruben Dell'Ariccia[1], Francesco Varone[2], Alessia Comes[1], Paolo Maria Leone[2], Venere Piluso[1], Alessandro Perrotta[1], Giuseppe Cicchetti[3], Diana Verdirosi[2] and Luca Richeldi[1,2]

Abstract

Background Although crackles on chest auscultation represent a fundamental component of the diagnostic suspect for fibrotic interstitial lung disease (ILD), their reliability has not been properly studied. We assessed the agreement among respiratory physicians on the presence and changes over time of audible crackles collected in a prospective longitudinal cohort of patients with fibrotic ILD.

Methods Lung sounds were digitally recorded at baseline and after 12 months at eight anatomical sites. Nine respiratory physicians blindly assessed randomized couples of recordings obtained from the same anatomical site at different timepoints. The physicians indicated the presence of crackles in individual recordings and which recording from each couple eventually had more intense crackles. Fleiss' kappa coefficient was used to measure inter- and intra-rater agreement.

Results Fifty-two patients, mostly with a diagnosis of IPF ($n=40$, 76.9%) were prospectively enrolled between October 2019 and May 2021. The final acoustic dataset included 702 single recordings, corresponding to 351 couples of recordings from baseline and 12-months timepoints. Kappa coefficient was 0.57 (95% CI 0.55–0.58) for the presence of crackles and 0.42 (95% CI 0.41–0.43) for acoustic change. Intra-rater agreement, measured for three respiratory physicians on three repeated assessments, ranged from good to excellent for the presence of crackles ($\kappa=0.87$, $\kappa=0.86$, $\kappa=0.79$), and from moderate to good for acoustic change ($\kappa=0.75$, $\kappa=0.76$, $\kappa=0.57$).

Conclusions Agreement between respiratory physicians for the presence of crackles and acoustic change was acceptable, suggesting that crackles represent a reliable acoustic finding in patients with fibrotic ILD. Their role as a lung-derived indicator of disease progression merits further studies.

Keywords Crackles, Interstitial lung disease, Lung sounds, Pulmonary fibrosis

*Correspondence:
Giacomo Sgalla
giacomo.sgalla@policlinicogemelli.it
[1]Università Cattolica del Sacro Cuore, Rome, Italy
[2]Dipartimento di Neuroscienze, Organi di Senso e Torace, Unità Operativa Complessa di Pneumologia, Fondazione Policlinico Universitario A. Gemelli IRCCS, Largo Agostino Gemelli 8, Rome 00168, Italy
[3]Dipartimento di Diagnostica per immagini e Radioterapia Oncologica, Centro Avanzato di Radiodiagnostica, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/.

Background

Accurate identification of disease progression in patients with fibrotic interstitial lung disease (ILD) is pivotal to optimize management, including the appropriate timing to start antifibrotic treatment, introduce supportive care strategies and guide the prompt referral for lung transplantation. Idiopathic pulmonary fibrosis (IPF) represents the archetype of progressive fibrotic lung disease, however other forms of ILD can share similar behaviour and prognosis [1, 2]. The recently emerged concept of progressive pulmonary fibrosis (PPF) corroborated the importance of a timely identification of disease progression, as antifibrotic therapies slow disease progression in these patients [3–5]. Chest auscultation represents a valuable point-of-care, low-cost, safe opportunity to prompt a timely diagnostic work up in patients with fibrotic ILD. The role of lung sounds as a screening and monitoring tool in ILD has been advocated [6–8]. International consensus guidelines recommend that IPF should be suspected in all patients with bibasilar inspiratory crackles [2, 9], which have been described as brief, discontinuous pathological lung sounds, explosive and transient in character, named after their similarity to the sound generated by Velcro™ strips separating [10, 11]. The presence of crackles, subjectively assessed by respiratory physicians, was found to be independently associated with distinct radiological features in the lung parenchyma, including honeycombing and reticular opacities [12]. Such finding suggests that crackles may be produced by advanced fibrosis as well as by less severe interstitial changes, thus representing a potential tool for the early detection of fibrotic ILD. Over the last decades, computerized methods of analysis of lung sounds allowed quantitative characterization of crackles in fibrotic ILD, that were shown to have distinctive features as compared to other conditions, such as chronic heart failure and pneumonia [13, 14]. In a prospective pilot study, quantitative features of lung sounds recorded from IPF patients demonstrated to change over time and correlate with the extent of fibrosis on CT scans [15]. While the incorporation of such methods in clinical practice requires further validation, chest auscultation performed by physicians remains limited by its subjectivity, despite the efforts made towards a standardization of the nomenclature of adventitious lung sounds [16]. With regards to fibrotic ILD, whether the presence or acoustic changes of crackles could be reliably appreciated over time is unknown. In this study, we aimed to evaluate the agreement on the presence and changes of audible crackles between physicians. New insights into the reliability of crackles' assessment could help clarify the potential role of lung sounds as an indicator of disease severity and progression.

Methods

Study design, population, and data collection

Consecutive patients referred to the ILD clinic at Fondazione Policlinico Universitario "A. Gemelli" IRCCS in Rome, Italy were invited to join a prospective longitudinal cohort. Eligibility criteria for this study were a multidisciplinary diagnosis of fibrotic ILD according to available international guidelines [17, 18], age between 40 and 90 years and the availability of at least 12-month follow up data. This study obtained ethics approval from the Fondazione Policlinico Universitario "A. Gemelli" IRCCS review board, Prot. 21,807/18, ID: 2104. Written informed consent was obtained from those patients willing to take part to the study before proceeding to data collection.

At baseline, demographics, smoking, exposures, family history for ILD and comorbidities were collected. Pulmonary function tests, including spirometry and diffusion capacity for carbon monoxide (DLco), were performed [19, 20] and exercise tolerance was measured using the 6-minute walk test (6MWT) [21]. The Gender/Age/Physiology (GAP) stage [22] was calculated for each patient. A high-resolution computed tomography (HRCT) of the chest was performed at enrolment (if not already performed within 12 months before). Lung function tests were repeated every 6 months. A thoracic radiologist (GC) with 7 years' experience in ILD independently reviewed anonymized HRCT scans and assessed radiological patterns according to the current classification of radiological usual interstitial pneumonia (UIP; probable UIP; indeterminate for UIP; alternative diagnosis) [2].

At baseline and after 12 months, lung sounds were digitally recorded sequentially at eight anatomical sites identified according to the guidelines for Computerized Respiratory Sounds Analysis (CORSA) [16]: upper sites were located at the first or second intercostal spaces, 2 cm from the paravertebral line; middle sites were located at the fourth or fifth intercostal space, 2 cm from the paravertebral line; lower sites were located 7 cm below the scapular angle and 5 cm from the paravertebral line; lateral sites were located at the fourth or fifth intercostal space on the mid-axillary line. Lung sounds were recorded for approximately ten seconds at each site or a time sufficient to record a minimum of two full breathing cycles, using a hand-held, commercially available electronic stethoscope (Littmann 3200™, 3 M, USA). Lung sound recording was not performed if patients were experiencing, at the time of the clinical appointment, events such as acute exacerbation of chronic obstructive pulmonary disease or acute heart failure, which could influence the auscultation findings related to the underlying fibrotic disease. The audio files were transferred to the Littmann StethAssist™ software (3 M, USA) via

Bluetooth, and exported in the wav format (sampled at 4 kHz with a resolution of 16-bit).

Assessment of crackles

The digital recordings obtained from the study participants at baseline and after 12 months were anonymized and randomized by assigning a random number to each audio file of the dataset. After receiving proper training by an investigator with expertise in digital auscultation (GS), an independent investigator (ADB) blindly listened to the lung sound dataset to assess the audio quality of each recording. Recordings that were deemed not suitable for assessment due to poor sound quality were excluded from further analyses. Criteria for exclusion included: short time of recording (less than 10 s or less than 3 complete breathing cycles), shallow breathing by the patient and the presence of acoustic artifacts to an extent, which could interfere with the proper assessment of lung sounds. For each patient, electronic folders were created that included two recordings obtained from the same anatomical site at baseline and at 12 months. A random sequence of the folders was then generated to avoid that files from the same patient were played consecutively. Nine physicians were invited to perform a subjective assessment of lung sounds: three were respiratory physicians with specific expertise in ILD ranging from 5 to 20 years; three were respiratory physicians with experience ranging from 3 to 15 years, but no specific expertise in ILD; three were fellow residents in respiratory medicine. The group of physicians attended consecutive listening sessions, held in a quiet room, where an independent investigator (JS) played the anonymized recordings from each folder. The sound files were played via personal computer using an open-source audio player (Audacity software, GNU General Public License) and were listened simultaneously by all physicians through over-ear headphones with active noise-cancelling feature (Sennheiser Momentum 4, Sennheiser, DE), connected to the source via Bluetooth technology. The randomization of the recordings ensured that baseline and 12-month recording could be played in either order: therefore, the recording time point could not be identified by the physicians. The listening sessions did not exceed the 2-hour length to avoid excessive tiredness from listening. A flow-chart describing how lung sounds assessment was performed by physicians is shown in Fig. 1, while a template of the assessment sheet provided to the physicians is reported in table S1 in the online data supplement. The two audio files from each folder were played twice. The physicians, blinded to clinical data and the timepoint at

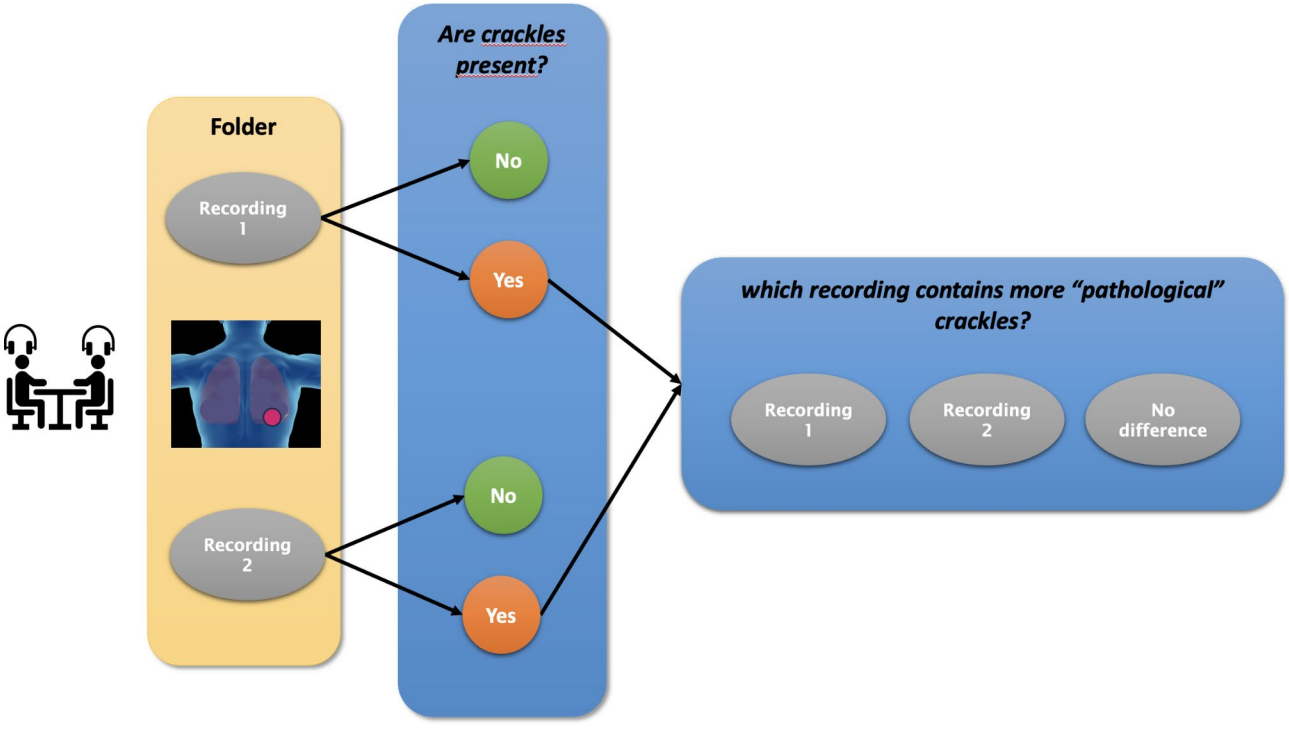


Fig. 1 Flow-chart of how lung sounds were assessed. Each electronic folder per patient contained two anonymized recordings obtained from the same recording site (right lower region, in the example shown in the figure) at baseline and 12 months. The two recordings were played consecutively. Each physician was asked to indicate whether crackles were absent or present in each recording. If crackles were indicated as present in both recordings, the physician was asked to indicate which recording had the more intense crackles, or if there was no difference in crackles' intensity between the two recordings

which randomized lung sounds were recorded, assessed the presence of crackles in the two recordings. If crackles were assessed to be present in both recordings of a couple, the physicians were also asked to indicate which of the two recordings contained the more "pathological" crackles, considering the frequency of the crackles, their qualitative characteristics (e.g., finer vs. coarser) and their distribution across the respiratory phases (e.g. inspiratory crackles vs. crackles present during expiration as well). The adjudication of what is acoustically more "pathological" was left to the subjective interpretation of the physicians, as there is no consensus on the nomenclature and the assessment of lung sounds. The physicians were also advised not to consider the potential presence of further adventitious lung sounds (e.g., rhonchi or squawks). After the group listening sessions were completed, three physicians (a respiratory physician with expertise in ILD, a respiratory physician without expertise in ILD, and a fellow in respiratory medicine) performed three repeated assessments of a subset of recordings for intra-rater agreement analysis. The dataset was randomized at each session, and a minimum one-week time interval was maintained between sessions.

Statistical analysis

Continuous and categorical data were summarized using means and standard deviations or counts and percentages respectively to describe the study population's characteristics.

Fleiss' kappa coefficient (κ) was used to evaluate levels of inter- and intra-rater agreement for the presence of crackles and for the longitudinal acoustic changes related to crackles characteristics. Inter-rater agreement was calculated for all nine physicians and for sub-groups of physicians with similar level of experience. To calculate agreement on acoustic change in crackles between study timepoints, physicians' assessments were converted into four levels: (1) crackles indicated as absent in both recordings; (2) crackles indicated as present in both recordings, with acoustic equivalence; (3) acoustic change occurring in the first recording; (4) acoustic change occurring in the second recording. An acoustic change event was met when a physician indicated either (1) absence of crackles in one recordings and presence of crackles in the other recording, or (2) presence of crackles in both recordings with more pathological crackles in one of the two recordings. Kappa coefficients were categorized as poor ($0<\kappa\leq0.20$), fair ($0.20<\kappa\leq0.40$), moderate ($0.40<\kappa\leq0.60$), good ($0.60<\kappa\leq0.80$), and excellent ($0.80<\kappa\leq1.00$) [23].

The assessments made by the nine physicians participating in the study were used to explore the relationship of longitudinal acoustic changes with disease progression. For each patient, the individual scores made by physicians for acoustic progression events, defined as either appearance of crackles in the 12-month recording or more pathological crackles in the 12-month recording, were summed to obtain overall acoustic progression scores (APS). Since some recordings were excluded after sound quality check, APS were weighed on the number of available recording sites for each patient. An example of how scores were calculated is described in table S2 in the online data supplement. Kaplan-Meyer curves and Cox proportional hazards regression were used to determine the association of 12-month acoustic progression with progression-free survival, defined as the occurrence of categorical decline in absolute % predicted forced vital capacity (FVC)≥10% from baseline or death from any cause.

Statistical significance was set at $p<0.05$. All statistical analyses were performed using SPSS version 28 (IBM, USA).

Results

Characteristics of study population and dataset

One hundred and eighteen patients were prospectively enrolled in the ILD longitudinal cohort between October 2018 and May 2022 and screened for eligibility (Fig. 2). Of these, 54 patients with at least 12 months of available follow up data were included in the study. The COVID-19 pandemic was mainly responsible for the low proportion of included patients, since several 12-month follow up visits were scheduled during the lockdown, when outpatients' clinical appointments were suspended. The initial dataset included 860 recordings, as 4 recordings were not retrieved. The 4 recordings left unpaired were discarded to consolidate a dataset including paired recordings only. After the sound quality check, 86 recordings (10% of the initial dataset) were excluded: of these, 27 (31.4%, of which 3 paired recordings) due to insufficient recording time, while 59 (68.6%, of which 6 paired recordings) due to insufficient sound quality. The 68 recordings left unpaired after quality check were also excluded. As such, 702 recordings were included in the final acoustic dataset. Since all the recordings obtained from 2 patients were excluded after the sound quality check, 52 patients formed the final study population.

Baseline characteristics of the study population are described in Table 1. Mean age was 73.8 years (SD 7.7) and there was a slight male predominance ($n=28$, 53.8%). The most frequent diagnosis was IPF ($n=40$, 76.9%), followed by autoimmune ILD ($n=6$, 11.5%), fibrotic hypersensitivity pneumonitis ($n=3$, 5.8%), idiopathic non-specific interstitial pneumonia, pneumoconiosis and unclassifiable ILD ($n=1$, 1.9%). Mean baseline predicted FVC was 85.7% (SD 20.6). Most patients had either a UIP ($n=25$, 53.2%) or a probable UIP ($n=12$, 25.5%) pattern at

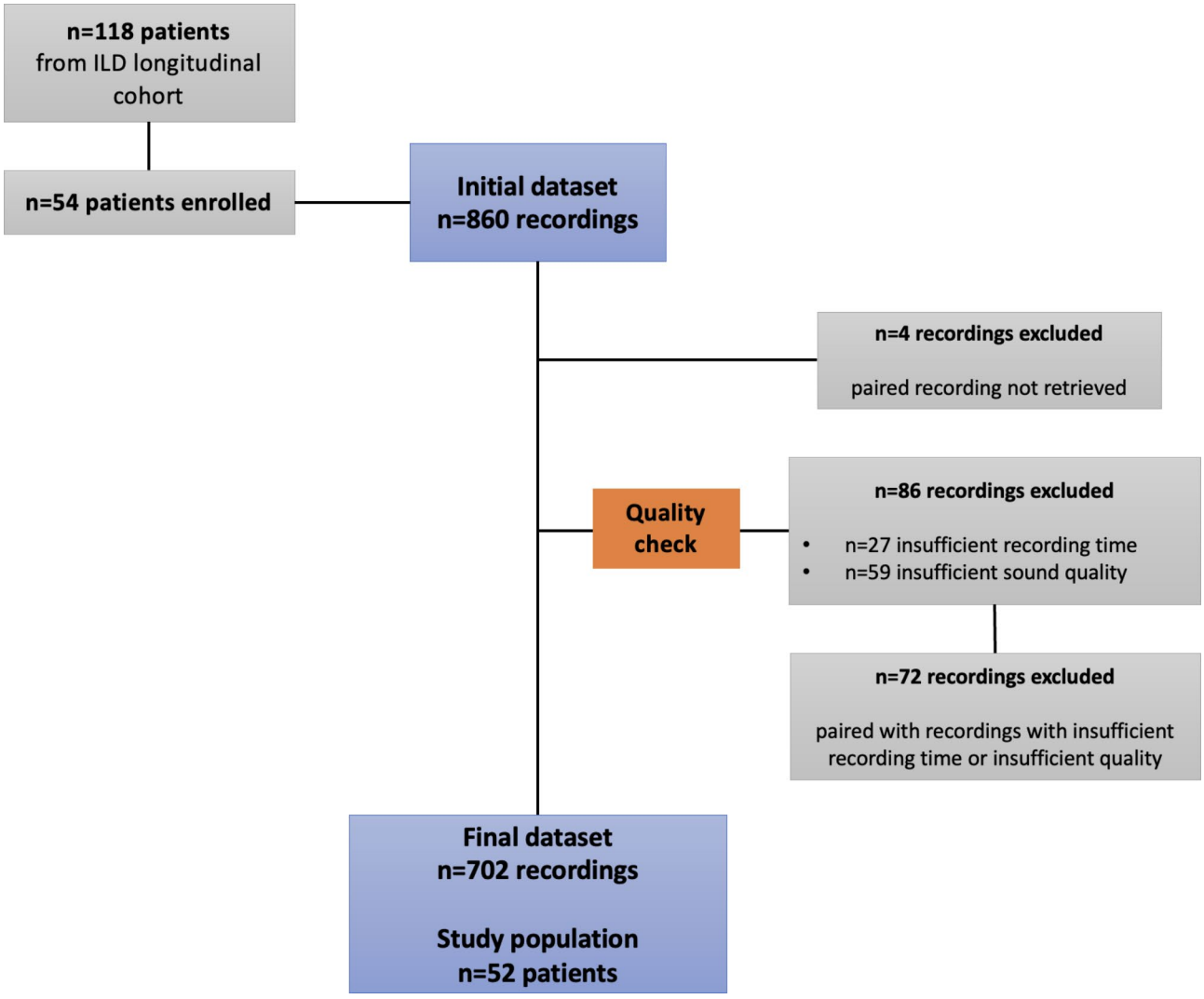


Fig. 2 Flowchart of data collection

baseline HRCT. Mean time from ILD diagnosis was 12.8 months (SD 15.9).

Identification of crackles

The frequencies of audible crackles in the acoustic dataset for all physicians are reported in Table 2. On average, crackles were identified in slightly more than half of the recordings (56.9%). The lowest rate of crackles was reported by physician 2.2 (n=300, 42.7%), while the highest rate was reported by physician 2.1 (n=561, 79.9%) (Table 2). After stratification by recording site, the frequency of crackles was found to be higher in recordings obtained from lower and lateral chest regions (right lower region=78.8%; left lower region=76.4%; right lateral region=70.9%; left lateral region=79.4%) as compared to medium and upper regions (right upper region=18.6%; left upper region=18.5%; right medium region=55.8%; left medium region=47.1%) (Fig. 3 and table S3 in the online data supplement). Subpopulations of patients with IPF (n=40) and non-IPF ILD (n=12) had a similar distribution of crackles across chest regions as compared to the overall population, although the average frequencies of crackles in recordings obtained from lower and lateral chest sites were higher in IPF patients (table S5 and S6 in the online data supplement). Patients with UIP or probable UIP pattern at HRCT (n=37) had similar distribution of crackles to patients with IPF, and higher frequencies of crackles in recordings obtained from lower and lateral chest sites as compared to patients with pattern indeterminate or suggestive of alternative diagnosis (n=10) (table S9 and S10 in the online data supplement). Inter-rater agreement levels among respiratory physicians for presence of crackles are reported in Table 3. Overall agreement among all physicians was moderate with a kappa coefficient of 0.57 (95% CI 0.55 to 0.58). Agreement levels were similar between pre-specified

Table 1 Baseline characteristics of the study population. Data are expressed as counts (%) or mean with standard deviation. BMI = body mass index; ILD = interstitial lung diseases; IPF = idiopathic pulmonary fibrosis; fHP = fibrotic hypersensitivity pneumonia; CTD-ILD = connective tissue disease related ILD; iNSIP = idiopathic non-specific interstitial pneumonia; UIP = usual interstitial pattern; COPD = chronic obstructive pulmonary disease; OSAS = obstructive sleep apnea syndrome; GERD = gastroesophageal reflux disease; GAP = gender age physiology; FVC = forced vital capacity; TLC = total lung capacity; DLco = diffusion lung capacity for carbon monoxide

	Patients (n = 52)	
Age, years	52	73.8 (7.7)
Sex	52	
Male		28 (53.8)
Female		24 (46.2)
Smoking history	52	
Current		5 (9.6)
Former		25 (48.1)
Never smoker		22 (42.3)
BMI	48	28 (4.8)
ILD diagnosis	52	
IPF		40 (76.9)
fHP		3 (5.8)
CTD-ILD		6 (11.5)
iNSIP		1 (1.9)
Pneumoconiosis		1 (1.9)
Unclassifiable ILD		1 (1.9)
Time from ILD diagnosis	50	12.8 (15.9)
Available recording sites	52	6.8 (1.7)
HRCT pattern	47	
UIP		25 (53.2)
Probable UIP		12 (25.5)
Indeterminate for UIP		2 (4.3)
Alternative diagnosis		8 (17)
ILD treatment		
Pirfenidone	52	17 (32.7)
Nintedanib		23 (44.2)
Steroids		4 (7.7)
Immunosuppressive		1 (1.9)
No treatment		7 (13.5)
Comorbidities		
COPD	52	8 (15.4)
Emphysema		9 (17.3)
OSAS		9 (17.3)
Chronic heart disease		18 (34.6)
Pulmonary hypertension		11 (21.2)
GERD		21 (40.4)
Anxiety/Depression		7 (13.5)
Diabetes		12 (23.1)
History of cancer (active or past)		10 (19.2)
Pulmonary function		
FVC, % pred	51	85.7 (20.6)
FEV_1/FVC	42	83.6 (9.6)
TLC, % pred	50	67.4 (14.5)
DL_{co}, % pred	49	66.6 (105.2)
GAP ILD stage		
I	49	12 (24.5)
II		31 (63.3)
III		6 (12.2)

subgroups of ILD respiratory physicians (k = 0.62, 95% CI 0.58 to 0.66) and non-ILD respiratory physicans (k = 0.54, 95% CI 0.52 to 0.56). The level of agreement was also similar in a subgroup of non-ILD physicians including only the respiratory medicine fellows (k = 0.66, 95% CI 0.62 to 0.7). Intra-rater agreement was then calculated for 3 physicians who performed 3 repeated assessments of 100 couples of recordings randomly extracted from the whole dataset. The level of agreement was excellent for the respiratory physician with and without ILD expertise (k = 0.87, 95% CI 0.79 to 0.95 and k = 0.86, 95% CI 0.78 to 0.94 respectively) and good for the respiratory medicine fellows (k = 0.79, 95%CI 0.71 to 0.87) (Table 4).

Longitudinal change in crackles

On average, acoustic change was indicated in most (54.3%) recording couples, with lowest rate reported by physician 1.3 (n = 157, 44.7%) and highest rate reported by physician 1.1 (n = 230, 65.5%) (Table 2). Acoustic change events were more frequently determined by a qualitative change in crackles (32.5%) as compared to the appearance of crackles at one timepoint (21.9%). Conversely, when crackles were assessed as present in both recordings, acoustic equivalence was indicated in less than one third (26.9%) of recording couples, with a large range of variability, with the lowest rate (1.9%) reported by physician 2.2 and the highest (43.6%) reported by physician 1.3. Just as for the presence of crackles, acoustic change events were also more frequent in couples of recordings obtained from lower and lateral chest regions (right lower region = 69.0%; left lower region = 68.1%; right lateral region = 67.7%; left lateral region = 69.9%) as compared to medium and upper regions (right upper region = 21.4%; left upper region = 21.9%; right medium region = 52.9%; left medium region = 52.3%) (Fig. 3 and table S4 in the online data supplement). Similar distributions of acoustic change events across chest regions were found in subpopulations of patients with IPF/UIP and non-IPF ILD/non-UIP, although patients with IPF/UIP had more acoustic change events in lower and lateral chest sites as compared to patients with non-IPF ILD/non-UIP (table S7, S8, S11, S12 in the online data supplement). Inter-rater agreement on acoustic change among all physicians was moderate and was lower when compared to the agreement for the presence of crackles (k = 0.42, 95%CI 0.41 to 0.43) (Table 3). Agreement levels for acoustic change were similar among ILD physicians (k = 0.47, 95% CI 0.43 to 0.5) and non-ILD physicians (k = 0.41, 95% CI 0.39 to 0.43). The level of agreement was also similar in a subgroup of non-ILD physicians including only the respiratory medicine fellows (k = 0.48, 95% CI 0.45 to 0.52). Intra-rater agreement for acoustic change crackles was good for the physicians with and without expertise in ILD (k = 0.75, 95% CI 0.69 to 0.82

Table 2 Crackles and acoustic change events indicated by physicians. Data are expressed as counts (%). Acoustic change was met when the physician indicated either (1) absence of crackles at one time point and presence of crackles at the other timepoint or (2) presence of crackles in both recordings, with more intense crackles audible at either of the two timepoints

Group	Physician	Presence of crackles	Acoustic change					
			Yes			No		
			Crackles in both recordings	Crackles in one recording	All	Crackles in both recordings	No crackles	All
ILD Physicians	1.1	408 (58.1)	122 (34.8)	108 (30.8)	230 (65.6)	28 (8)	93 (26.4)	121 (34.4)
	1.2	395 (56.3)	113 (32.2)	73 (20.8)	186 (53)	48 (13.7)	117 (33.4)	165 (47)
	1.3	394 (56.1)	93 (26.5)	64 (18.2)	157 (44.7)	72 (20.5)	122 (34.8)	194 (55.3)
	Mean, %	56.8	31.2	23.3	54.4	14.1	31.5	45.6
Non-ILD Physicians	2.1	561 (79.9)	151 (43)	29 (8.3)	180 (51.3)	115 (32.8)	56 (16)	171 (48.7)
	2.2	300 (42.7)	102 (29.1)	90 (25.6)	192 (54.7)	3 (0.9)	156 (44.4)	159 (45.3)
	2.3	376 (53.6)	120 (34.2)	94 (26.8)	214 (61)	21 (6)	116 (33)	137 (39)
	2.4	398 (56.7)	98 (27.9)	62 (17.7)	160 (45.6)	70 (19.9)	121 (34.4)	192 (54.4)
	2.5	380 (54.1)	133 (37.9)	88 (25.1)	221 (63)	13 (3.7)	117 (33.4)	130 (37)
	2.6	382 (54.4)	93 (26.5)	82 (23.4)	175 (49.9)	57 (16.2)	119 (34)	176 (50.1)
	Mean, %	56.9	33.1	21.2	54.2	13.2	32.5	45.7
Overall mean, %		56.9	32.5	21.9	54.3	13.5	32.2	45.7

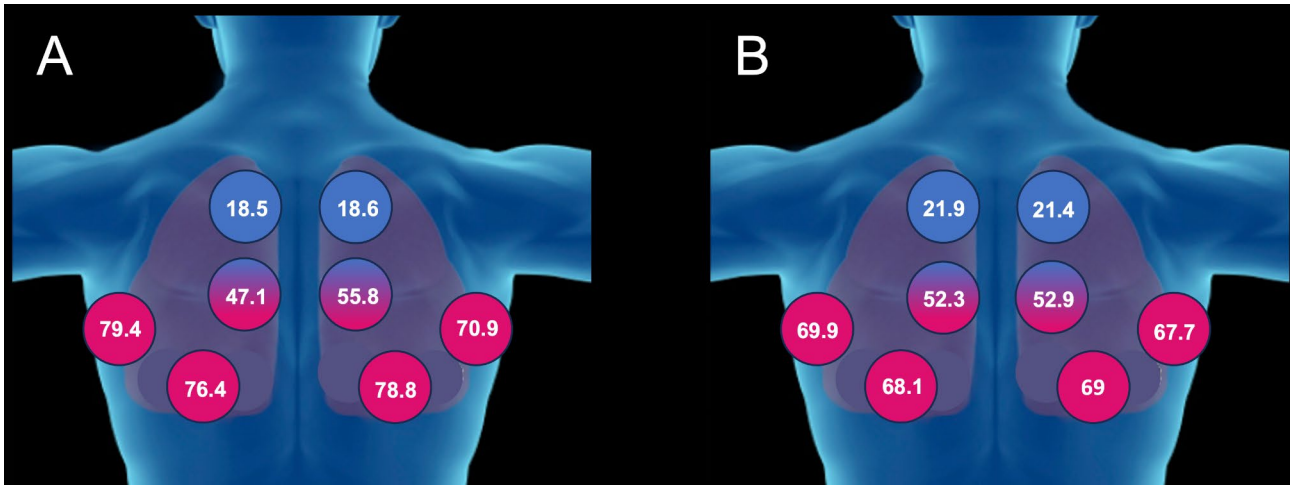


Fig. 3 Crackles in single recordings (panel A) and acoustic change events (panel B) reported by respiratory physicians. Data are mean percentages among all nine physicians

Table 3 Inter-rater agreement on presence of crackles and acoustic change for all respiratory physicians and for subgroups of physicians with different level of experience. Data are expressed as Fleiss' kappa values (95% CI). All calculated kappa values were significant with $p < 0.001$

	All physicians ($n=9$)	ILD physicians ($n=3$)	Non-ILD physicians ($n=6$)
Presence of crackles ($n=702$)	0.57 (0.55–0.58)	0.62 (0.58–0.66)	0.54 (0.52–0.56)
Acoustic change ($n=351$)	0.42 (0.41–0.43)	0.47 (0.43–0.5)	0.41 (0.39–0.43)

Table 4 Intra-rater agreement on presence of crackles and acoustic change for respiratory physicians with different level of experience. Data are expressed as Fleiss' kappa values (95% CI). All calculated kappa values were significant with $p < 0.001$

	Physician 1.2 (ILD physician)	Physician 2.2 (non-ILD physician)	Physician 2.4 (Respiratory medicine fellow)
Presence of crackles ($n=200$)	0.87 (0.79–0.95)	0.86 (0.78–0.94)	0.79 (0.71–0.87)
Acoustic change ($n=100$)	0.75 (0.69–0.82)	0.76 (0.68–0.842)	0.57 (0.5–0.63)

and k=0.76, 95% CI 0.68 to 0.842 respectively), and was moderate for the respiratory medicine fellow (k=0.57, 95% CI 0.5 to 0.63) (Table 4).

In order to further explore agreement on acoustic change, these events were stratified in those defined by the appearance of new crackles (absence of crackles in one recordings and presence of crackles in the other recording and those defined by the "worsening" of crackles (presence of crackles indicated in both recordings, with more pathological crackles in one of the two recordings). Inter-rater agreement calculated for these subcategories was higher for the appearance of new crackles (k=0.4, 95% CI 0.38 to 0.42) as compared to the identification of worsening crackles (k=0.25, 95% CI 0.23 to 0.26).

Relationships with disease progression

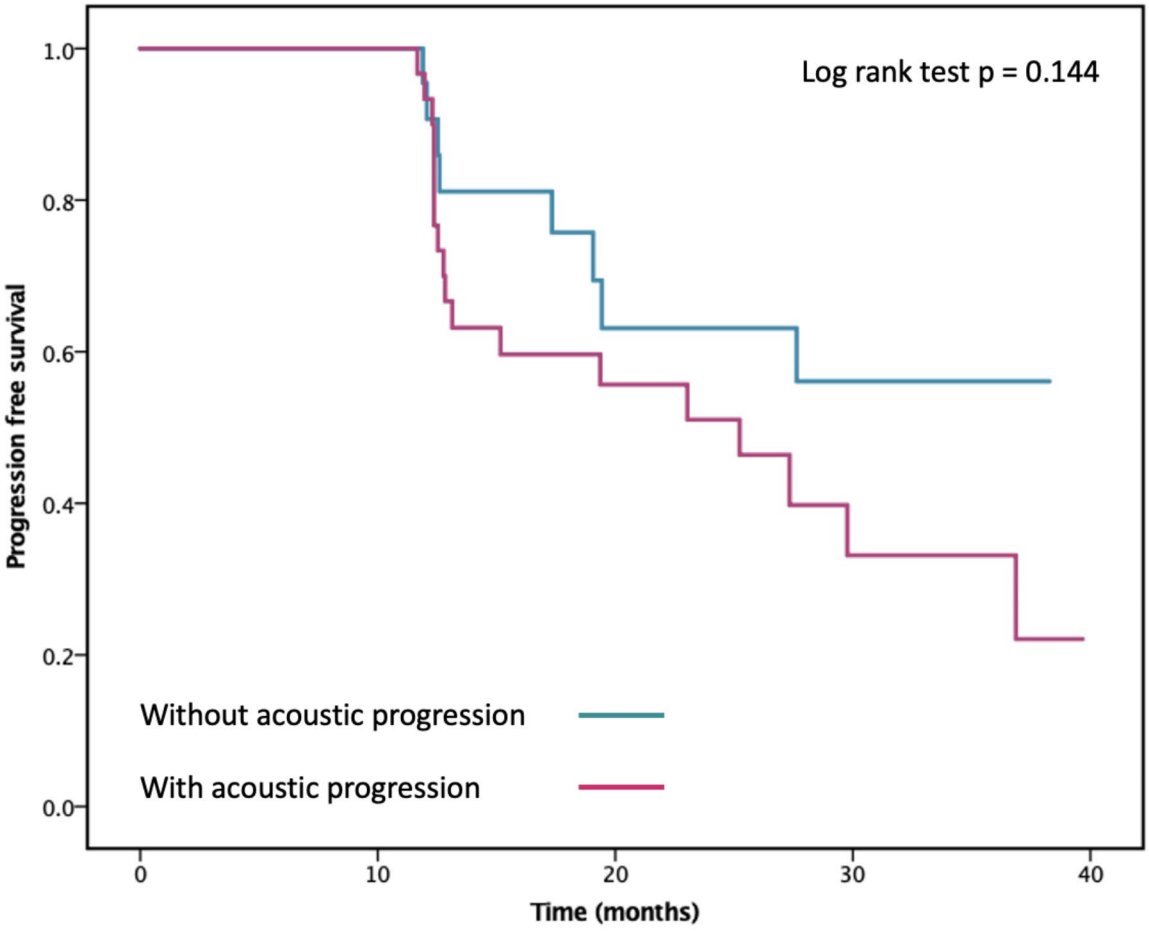
An acoustic progression score (APS) was calculated for each patient included in the study as described in the methods section. The median APS in the study population was 16, with interquartile range (IQR) from 15 to 22 and interval range from 2.6 to 34. The median value of the scores of the study population was then used as cut-off to discriminate patients with (APS≥16) and without (APS<16) significant acoustic progression between baseline and 12 months. Using such criteria, 30 patients (57.7%) with acoustic progression were identified. Patients with significant acoustic progression after 12 months exhibited a trend toward reduced progression-free survival at follow up (log-rank $p=0.144$) (Fig. 4).

Discussion

We studied the inter-rater agreement among respiratory physicians for crackles auscultation in lung sounds digitally recorded in a prospective cohort of patients with fibrotic ILD. The results show that both the presence of crackles and their acoustic changes over time are reliably assessed by respiratory physicians, thus confirming the crucial role of chest auscultation in the management of patients with fibrotic ILD.

In this dataset of digital recordings, crackles were more frequently identified in recordings obtained from anatomical sites corresponding to the lower and the lateral regions of the chest. Such finding reflects the predominant lower distribution of fibrotic changes in many ILD: notably, the distribution of crackles across chest regions was similar regardless ILD diagnosis or the radiological pattern on HRCT, although patients with IPF or with a UIP/probable UIP pattern had more crackles identified in the lower and the lateral regions of the chest as compared to patients with non-IPF ILD or non-UIP pattern. This is consistent with previous findings showing that radiological UIP is the pattern most strongly associated with audible crackles [12]. Acoustic change events were also more frequently identified at anatomical sites corresponding to lower and lateral chest regions, and mostly determined by changes in the qualitative features of crackles, rather than by the appearance of new crackles between study timepoints. We found that the agreement levels for the presence of crackles and their longitudinal change between respiratory physicians are acceptable regardless of their level of expertise in the field of ILD. The high levels of intra-rater agreement suggested that lung sounds are consistently interpreted by the same physician. This is particularly true for more experienced physicians, who showed higher levels of intra-rater agreement as compared to the younger colleague. Overall, the assessment of acoustic changes seems to be less reliable compared to the assessment of the presence of crackles, as demonstrated by the lower levels of inter- and intra-rater agreement. In particular, these appear to be driven by a greater difficulty in identifying qualitative changes in crackles when they are already present, as compared to the appearance of new crackles. Indeed, physicians are not used to compare lung sounds over time, since lung sounds are not usually being recorded during clinical practice. It is also possible that the assessments made by respiratory physicians could be influenced by the presence of sound artifacts. At the sound quality check performed prior to the lung sound assessment, 59 recordings (6.8% of the initial acoustic dataset) were discarded due to insufficient sound quality. Nonetheless, some artifacts due to slight movements of the stethoscope's diaphragm on the skin could still be present in the recordings included in the final dataset. These could mimic or cover crackle sounds, thus influencing the interpretation of acoustic change among physicians in different ways. Previous studies on the quantitative characterization of lung sounds involved devices such as multi-channel analyzers to allow simultaneous recording of multiple sites and the use of pneumotachographs for airflow standardization [14, 24]. Since the scope of the study was to determine the agreement among physicians on the subjective assessment of lung sounds, we preferred the adoption of a commercially available stethoscope over more sophisticated recording tools and tightly standardized conditions. On the other hand, we believe that our findings highlight the need for technological improvements in the field of digital stethoscopes to maximize acoustic reliability.

Worse, although non-statistically significant trends of progression-free survival were shown at follow up by patients with higher APS scores, suggesting that patients showing longitudinal acoustic progression exhibit a more progressive behavior. Indeed, this exploratory analysis was limited by the small population size and the subjective assessment of multiple physicians used for scores' calculation. Further exploration in larger cohorts of patients using quantitative methods for lung sound



	Total N	N of Events	N Censored	% Censored
Without acoustic progression	22	8	14	63.6%
With acoustic progression	30	18	12	40%
Overall	52	26	26	50%

Fig. 4 Kaplan-Meyer curves for progression-free survival of patients with or without significant acoustic progression (APS ≥ 16 and APS < 16, respectively) between baseline and 12 months. Progression-free survival was defined as decline in absolute % predicted FVC ≥ 10% from baseline or death from any cause

analysis is required to investigate the potential role of lung sounds as a prognostic indicator in fibrotic ILD.

Strengths of our study include the prospective design and the standardization of the procedures followed for the assessment of lung sounds. The simultaneous, collective listening sessions and the use of the same headsets ensured to avoid biases resulting from individual listening approaches, such as different times allocated for playback or the use of different models of earphones/headphones. Our study also has limitations. Those related to the use of an electronic stethoscope and unstandardized tidal breathing volumes have been discussed above.

In conclusion, our findings indicate that the presence of crackles is a reliable clinical finding in patients with fibrotic ILD and support the central role of chest auscultation in the diagnostic work up of these patients. The reliability of the assessment of longitudinal changes in crackles is lower, but acceptable. Since acoustic changes could reflect disease progression, further research in the field of automated quantification of lung sounds is

warranted to clarify the role of crackles as lung-derived prognostic indicator, with potential applications in telemedicine-based approaches and as a novel endpoint in pharmacological clinical trials.

Abbreviations

CORSA	Computerized Respiratory Sounds Analysis ILD: interstitial lung disease
GAP	Gender/Age/Physiology
HRCT	high-resolution computed tomography
IPF	idiopathic pulmonary fibrosis
PPF	progressive pulmonary fibrosis

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12931-024-02979-9.

Supplementary Material 1

Supplementary Material 2

Author contributions

Conceptualization of the study: GS, JS, LR. Data collection: GS, JS, ADB, TM, BI, GP, RDA, FV, AC, PML, VP, AP, GC. Statistical Analysis: GS, JS, DV. Writing of study manuscript: GS, JS, GC, ADB. Review of study manuscript: TM, BI, GP, RDA, FV, AC, PML, VP, AP, DV, GC, LR; Supervision: GS, JS, DV, LR. All authors have read and approved the manuscript.

Funding

None.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethical approval

This study obtained ethics approval from the Fondazione Policlinico Universitario "A. Gemelli" IRCCS review board, Prot. 21807/18 (29798/18), ID: 2104.

Consent for publication

Not applicable.

Competing interests

GS reports personal fees from Chiesi Farmaceutici, personal fees from Boehringer Ingelheim, outside the submitted work; JS reports personal fees from Chiesi Farmaceutici, personal fees from Boehringer Ingelheim, outside the submitted work; LR reports personal fees from Biogen, grants and personal fees from Roche, personal fees from ImmuneWorks, grants and personal fees from Boehringer Ingelheim, personal fees from Celgene, personal fees from Nitto, personal fees from FibroGen, personal fees from Promedior, personal fees from Pliant Therapeutics, personal fees from Asahi Kasei, personal fees from Toray, personal fees from BMS, personal fees from RespiVant, personal fees from CSL Behring, outside the submitted work; ADB, TM, BI, GP, RDA, FV, AC, PML, AP, GC, DV have no conflicts of interest to disclose.

Received: 4 August 2024 / Accepted: 16 September 2024
Published online: 28 September 2024

References

1. Wijsenbeek M, Cottin V. Spectrum of Fibrotic Lung diseases. N Engl J Med. 2020;383:958–68.
2. Raghu G, Remy-Jardin M, Richeldi L, Thomson CC, Inoue Y, Johkoh T, Kreuter M, Lynch DA, Maher TM, Martinez FJ, et al. Idiopathic pulmonary fibrosis (an update) and progressive pulmonary fibrosis in adults: an Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. Am J Respir Crit Care Med. 2022;205:e18–47.
3. Flaherty KR, Wells AU, Cottin V, Devaraj A, Walsh SLF, Inoue Y, Richeldi L, Kolb M, Tetzlaff K, Stowasser S, et al. Nintedanib in Progressive Fibrosing interstitial lung diseases. N Engl J Med. 2019;381:1718–27.
4. Behr J, Prasse A, Kreuter M, Johow J, Rabe KF, Bonella F, Bonnet R, Grohe C, Held M, Wilkens H, et al. Pirfenidone in patients with progressive fibrotic interstitial lung diseases other than idiopathic pulmonary fibrosis (RELIEF): a double-blind, randomised, placebo-controlled, phase 2b trial. Lancet Respir Med. 2021;9:476–86.
5. Maher TM, Corte TJ, Fischer A, Kreuter M, Lederer DJ, Molina-Molina M, Axmann J, Kirchgaessler KU, Samara K, Gilberg F, Cottin V. Pirfenidone in patients with unclassifiable progressive fibrosing interstitial lung disease: a double-blind, randomised, placebo-controlled, phase 2 trial. Lancet Respir Med. 2020;8:147–57.
6. Cottin V, Richeldi L. Neglected evidence in idiopathic pulmonary fibrosis and the importance of early diagnosis and treatment. Eur Respir Rev. 2014;23:106–10.
7. Cottin V, Cordier JF. Velcro crackles: the key for early diagnosis of idiopathic pulmonary fibrosis? Eur Respir J. 2012;40:519–21.
8. Richeldi L, Cottin V, Würtemberger G, Kreuter M, Calvello M, Sgalla G. Digital Lung Auscultation: will early diagnosis of Fibrotic interstitial lung disease become a reality? Am J Respir Crit Care Med. 2019;200:261–3.
9. Lederer DJ, Martinez FJ. Idiopathic pulmonary fibrosis. N Engl J Med. 2018;379:797–8.
10. Forgacs P. Crackles and wheezes. Lancet. 1967;2:203–5.
11. Pasterkamp H, Brand PLP, Everard M, Garcia-Marcos L, Melbye H, Priftis KN. Towards the standardisation of lung sound nomenclature. Eur Respir J. 2016;47:724–32.
12. Sgalla G, Walsh SLF, Sverzellati N, Fletcher S, Cerri S, Dimitrov B, Nikolic D, Barney A, Pancaldi F, Larcher L, et al. Velcro-type crackles predict specific radiologic features of fibrotic interstitial lung disease. BMC Pulm Med. 2018;18:103.
13. Flietstra B, Markuzon N, Vyshedskiy A, Murphy R. Automated analysis of Crackles in patients with interstitial pulmonary fibrosis. Pulmonary Med. 2011;2011:590506.
14. Vyshedskiy A, Bezares F, Paciej R, Ebril M, Shane J, Murphy R. Transmission of crackles in patients with interstitial pulmonary fibrosis, congestive heart failure, and pneumonia. Chest. 2005;128:1468–74.
15. Sgalla G, Larici AR, Sverzellati N, Bartholmai B, Walsh SLF, Nikolic D, Barney A, Fletcher S, Jones M, Davies DD, Richeldi L. Quantitative analysis of lung sounds for monitoring idiopathic pulmonary fibrosis: a prospective pilot study. Eur Respir J. 2019;53:1802093
16. Sovijärvi ARA, Vanderschoot J, Earis JE. Computerized respiratory sound analysis (CORSA): recommended standards for terms and techniques : ERS Task Force Report. Copenhagen: Munksgaard; 2000.
17. Travis WD, Costabel U, Hansell DM, King TE Jr., Lynch DA, Nicholson AG, Ryerson CJ, Ryu JH, Selman M, Wells AU, et al. An official American Thoracic Society/European Respiratory Society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. Am J Respir Crit Care Med. 2013;188:733–48.
18. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, Behr J, Cottin V, Danoff SK, Morell F, et al. Diagnosis of idiopathic pulmonary fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. Am J Respir Crit Care Med. 2018;198:e44–68.
19. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CP, Gustafsson P, et al. Standardisation of spirometry. Eur Respir J. 2005;26:319–38.
20. MacIntyre N, Crapo RO, Viegi G, Johnson DC, van der Grinten CPM, Brusasco V, Burgos F, Casaburi R, Coates A, Enright P, et al. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. Eur Respir J. 2005;26:720–35.
21. ATS. Guidelines for the six-minute walk test. Am J Respir Crit Care Med. 2002;166:111–7.
22. Ley B, Ryerson CJ, Vittinghoff E, Ryu JH, Tomassetti S, Lee JS, Poletti V, Buccioli M, Elicker BM, Jones KD, et al. A multidimensional index and staging system for idiopathic pulmonary fibrosis. Ann Intern Med. 2012;156:684–91.
23. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ. 1992;304:1491–4.

24. Ponte DF, Moraes R, Hizume DC, Alencar AM. Characterization of crackles from patients with fibrosis, heart failure and pneumonia. Med Eng Phys. 2013;35:448–56.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.