

Title: Reliability, validity, and minimal detectable change of computerised respiratory sounds in patients with Chronic Obstructive Pulmonary Disease

Short title: Measurement properties of respiratory sounds

Authors: Ana Oliveira¹⁻³, Susan Lage⁴, João Rodrigues⁵, Alda Marques^{2,3}

¹Faculty of Sports, University of Porto, Portugal

²Lab 3R – Respiratory Research and Rehabilitation Laboratory, School of Health Sciences (ESSUA), University of Aveiro, Portugal

³Institute for Research in Biomedicine (iBiMED), University of Aveiro, Portugal

⁴ Rehabilitation Sciences Program, School of Physical Education, Physiotherapy and Occupational Therapy (EEFFTO), Federal University of Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brazil

⁵Institute of Electronics and Informatics Engineering of Aveiro (IEETA), University of Aveiro 3810-193 Aveiro, Portugal

Corresponding author: Alda Marques, PT, MSc, PhD, Senior Lecturer, Lab 3R – Respiratory Research and Rehabilitation Laboratory, School of Health Sciences, University of Aveiro (ESSUA), Agras do Crasto - Campus Universitário de Santiago, Edifício 30, 3810-193 Aveiro, Portugal. Email: amarques@ua.pt; Telephone: 00351234372462.

Author contributions

AO and AM were responsible for the study conception and design. AO performed data collection, data analysis and drafted the manuscript. SL and JR performed data analysis. AO and AM obtained the funding. All authors critically revised the paper for important intellectual content

Declaration of Interest

The authors report no conflicts of interest.

Abstract

Introduction: Computerised respiratory Sounds (CRS) are closely related to the movement of air within the tracheobronchial tree, and are promising outcome measures in patients with chronic obstructive pulmonary disease (COPD). However, CRS measurement properties have been poorly tested.

Objective: To assess the reliability, validity, and the minimal detectable changes (MDC) of CRS in patients with stable COPD.

Methods: Fifty patients (36♂, 67.26±9.31y, FEV₁ 49.52±19.67%predicted) were enrolled. CRS were recorded simultaneously at seven anatomic locations (trachea; right and left anterior, lateral and posterior chest). The number of crackles, wheeze occupation rate (%Wh), median frequency (F50) and maximum intensity (Imax) were processed using validated algorithms. Within-day and between-days reliability, criterion and construct validity, validity to predict exacerbations and MDC were established.

Results: CRS presented moderate-to-excellent within-day reliability ($ICC_{1,3} \geq 0.51$; $p < 0.05$) and moderate-to-good between-days reliability ($ICC_{1,2} \geq 0.47$; $p < 0.05$) for most locations. Low-to-moderate correlations with FEV₁%predicted were found ($-0.53 < r_s < -0.28$; $p < 0.05$), and the inspiratory number of crackles were the best discriminator between mild-to-moderate and severe-to-very severe airflow limitations (area under the curve > 0.78). CRS correlated poorly with patient-reported outcomes ($r_s < 0.48$; $p < 0.05$) and did not predict exacerbations. Inspiratory number of crackles at posterior right chest, inspiratory F50 at trachea and anterior left chest and expiratory Imax at anterior right chest were simultaneously reliable and valid, and their MDC were 2.41, 55.27, 29.55 and 3.98, respectively.

Conclusion: CRS are reliable and valid. Their use, integrated with other clinical and patient-reported measures, may fill the gap of assessing small airways and contribute towards a patient's comprehensive evaluation.

Keywords: Reproducibility of results; Pulmonary Disease, Chronic Obstructive; Respiratory Sounds; Spirometry

Introduction

Chronic Obstructive Pulmonary Disease (COPD) is characterized by persistent respiratory symptoms and airflow limitation due to smaller airway and/or alveolar abnormalities (1). Although diagnosis and monitoring of airflow limitation is usually performed by spirometry (gold standard test of lung function) (1), its usefulness to assess interventions has been questioned, as it mainly assesses large airways (2), changes in response to treatments are small (3,4), and correlates poorly with patient-reported outcomes (5). Thus, international Respiratory Societies have been stressing the need to validate instruments that can express peripheral respiratory function, assess patient's response to interventions and correlate with patient-reported outcomes (6).

Computerised respiratory sounds are a simple, objective, and non-invasive outcome measure that are directly related to the movement of air within the tracheobronchial tree (7). Therefore, changes in airway and/or alveolar mechanisms may be primarily detected by changes in the frequency/intensity of normal respiratory sounds and by the presence of adventitious respiratory sounds (i.e., crackles and wheezes) (7). This theoretical potential of computerised respiratory sounds to be used as an outcome measure has been motivating research of their characteristics and measurement properties (8-10). A recent study in stable patients with COPD has shown that respiratory sounds have adequate within-day reliability (10). However, other measurement properties need to be studied before computerised respiratory sounds utilisation can be recommended for clinical practice (11). Between-days reliability and validity are crucial measurement properties of an outcome measure which, according to the authors' best knowledge, have never been explored in computerised respiratory sounds, hindering the interpretation of its actual usefulness to assess lung function (validity) and its repeatability during prolonged stable phases (reliability).

This study aimed to evaluate the between-days reliability, criterion, construct and predictive validity of computerised respiratory sounds in patients with COPD. The authors hypothesised that computerised respiratory sounds would present i) significant and moderate between-days reliability; ii) significant, negative and low-to-moderate correlations with lung function; iii) significant, negative and moderate

correlations with patient-reported outcome measures and iv) significant ability to predict acute exacerbations of COPD up to 1 year.

Materials and Methods

Study design

A cross-sectional study was conducted. Reliability and validity were explored, described and interpreted following the COnsensus-based Standards for the selection of health status Measurement Instruments (COSMIN) guidelines (11,12).

Sample size

The sample size was determined according to the COSMIN guidelines, which have established that a study with good methodological quality should enrol a minimum of 50 participants (12).

Participants

Outpatients with stable COPD were recruited from a central hospital between January 2016 and 2017. Inclusion criteria were diagnosis of COPD according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria (1) and clinical stability for one month prior to the study (i.e., no hospital admissions, exacerbations as defined by GOLD, or changes in respiratory system medication). Patients were excluded if they had severe co-existing respiratory, neurological, cardiac, musculoskeletal, or psychiatric impairments. Approvals for this study were obtained from the ethics committee of the Central Hospital (13NOV'1514:40065682) and National Data Protection Committee (8828/2016). Eligible patients were identified by clinicians and then contacted by the researchers, who explained the purpose of the study and asked about their willingness to participate. When patients agreed to participate, an appointment with the researchers was scheduled and written informed consent was obtained.

Data collection

Participants were asked to attend to two testing sessions with a five to seven days interval. In the first session, patients completed a questionnaire with sociodemographic (age, gender) and health-related (smoking status, exacerbations in

the previous year, symptoms and impact of the disease) information. Height and weight were recorded to calculate the body mass index (BMI).

Smoking status was evaluated with a 2-question survey on current and previous smoking habits. Cough and wheezing were assessed through a numerical scale (NS) in which the patient reported the severity of the symptoms in the previous 24h. The NS is reliable (Intraclass correlation coefficient, ICC from 0.54 to 0.86) and valid to assess symptoms in patients with respiratory diseases (13,14). Dyspnoea was collected using the modified Medical Research Council (mMRC) dyspnoea scale (15). The patients read the 5-point mMRC scale and pointed the grade (0 to 4) that most closely matched his or her breathlessness. Higher scores represent more breathlessness. The mMRC has shown to be a reliable (ICC=0.82) (16) and valid measure of disability related with dyspnoea (17). Impact of the disease was collected with the COPD Assessment Test (CAT). The CAT is a reliable (Cronbach α =0.88) and valid self-administered 8-question questionnaire, which allows the assessment of the impact of COPD on health status within only a few minutes (18). Higher scores represent higher impact of COPD.

Then, three respiratory sounds recordings were performed with air-coupled electret microphones (C 417 PP, AKG Acoustics GmbH, Vienna, Austria) (19) following the computerised respiratory sound analysis (CORSA) guidelines for short-term acquisitions (20). Finally, lung function was assessed with a portable spirometer (MicroLab 3535, CareFusion, Kent, UK) according to the guidelines (21).

In the second session, only respiratory sounds were recorded. Effort was made to keep all factors associated with the testing sessions consistent, specifically the time of day, location of the sessions, chest locations of the microphones and order of testing.

Additionally, participants were telephoned every three months, up to one year of their initial assessment, to gather information about the occurrence of an exacerbation (1).

Respiratory sound recordings

Recordings were performed simultaneously at seven anatomic locations (trachea and right and left anterior upper, lateral middle, and posterior lower chest) (20). The recording system included eight air-coupled electret microphones, a multi-channel audio interface (AudioBox 1818 VSL, PreSonus, Florida, USA), and a laptop computer running LungSounds@UA interface (22). Seven microphones, mounted in couplers made Teflon (23), were attached on the participant's skin with double-faced adhesive tapes (Double Stick Discs, 3M Littmann, Cheshire, UK), and one microphone was placed closed to the patient to record the background noise. The analog sound signals acquired were amplified and converted to digital by the audio interface with a 24-bit resolution and a sampling rate of 44.1 kHz. Each data acquisition session lasted for 20-s (24) and the recorded data were later converted to .wav format.

Signal processing

Respiratory sound files were processed by automatic algorithms implemented in Matlab R2009a (MathWorks, Natick, Massachusetts). Data were obtained for number of crackles, occupation rate of wheezes (%Wh), median frequency (F50) and maximum intensity (Imax) per respiratory phase (i.e., inspiration and expiration) and per chest location.

Number of crackles per respiratory phase was calculated using equation 1:

$$\text{Number of crackles} = \frac{\text{Sum of crackles per respiratory phase}}{\text{Total number of respiratory phases}} \quad (1)$$

Occupation rate of wheezes (%Wh) was calculated through equations 2 and 3:

$$\text{Rate of each wheeze} = \frac{\text{Duration of wheeze in the respiratory phase}}{\text{Total duration of the respiratory phase}} \times 100 \quad (2)$$

$$\%Wh = \sum (\text{rate of each wheeze in the respiratory phase}) \quad (3)$$

Median frequency (F50) and maximum intensity (Imax) were calculated following the methodology proposed by Pasterkamp, Powell, Sanchez (25) after excluding adventitious respiratory sounds in each file. F50 and Imax were analysed within a frequency band of 300–600 Hz, as this has been indicated as the most representative frequency band for patients with respiratory diseases (26,27). All analyses were checked by two respiratory experts and the average respiratory sound spectra and background noise were plotted to ensure the quality of sound recordings. Background

noise was closely superimposed to respiratory sound intensity at lateral chest; hence, these locations were excluded from further analyses. The average spectra of normal respiratory sounds at trachea, anterior, lateral and posterior chest can be found in the supplementary material and a detailed description of the signal processing is provided elsewhere (28).

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics version 24.0 (IBM Corporation, Armonk, NY, USA) and plots were created using GraphPad Prism version 5.01 (GraphPad Software, Inc., La Jolla, CA, USA). The level of significance was set at 0.05. Descriptive statistics were used to describe the sample. Characteristics were compared between patients at stages I-II and III-IV of airflow limitation (1), using independent t-tests for normally distributed data (age, BMI, and lung function), Mann-Whitney U-tests for ordinal data (mMRC, CAT and NS), and Chi-square tests for categorical data (gender, smoking status, number of exacerbations/year).

Reliability

Within- and between-days reliability were determined. Relative and absolute reliability were calculated with the ICC and the Bland and Altman method, respectively (29). Within-day reliability was computed using the ICC equation (1, k), where k=3 corresponds to the three recordings performed in session 1. The Bland and Altman method assesses the agreement between two sets of measures (30); thus, random numbers were generated in MATLAB to delete one recording. Between-days reliability was computed using the ICC equation (1, k), where k=2 corresponds to the two recordings used (one from session 1 and one from session 2). Bland and Altman plots were also created to analyse the distribution of results from session 1 and 2. ICC was interpreted as excellent (>0.75), moderate-to-good (0.4–0.75), or poor (<0.4) (31).

Validity

Criterion validity was assessed by analysing the degree to which respiratory sounds correlated with lung function (i.e., percent predicted forced expiratory volume in one second, FEV₁%predicted) using Spearman's rank correlation coefficient. The strength of the correlations was interpreted as negligible (i.e., 0-0.30); low (0.31-0.50), moderate (0.51-0.70), high (0.71-0.90) or very high (0.91-1) (32).

Receiver operating characteristic (ROC) analysis was used to assess the ability of respiratory sounds to differentiate between patients' airflow limitation severity. The ROC analysis only allows to plot the performance of a binary classification, thus, patients classified in the GOLD criteria as I and II were labelled as mild-to-moderate airflow limitation and patients classified as III and IV were labelled as severe-to-very severe airflow limitation. The cut-off for each respiratory sound parameter was chosen as the point where the sensitivity and specificity were simultaneously maximized. Area under the curves (AUC) and the 95% confidence interval were determined. AUC was interpreted as: AUC=0.5 no discrimination; $0.7 \leq \text{AUC} < 0.8$ acceptable discrimination; $0.8 \leq \text{AUC} < 0.9$ excellent discrimination and $\text{AUC} \geq 0.9$ outstanding discrimination (33).

Construct validity was assessed by examining the relationship between respiratory sounds, NS, mMRC and CAT using Spearman's rank correlation coefficient.

Predictive validity up to 12 months exacerbations were explored with ROC analysis and the influence of independent predictors (i.e., no. of crackles, %Wh, F50 and I_{max}) on the time until the first exacerbation was analysed by univariate and multivariate Cox regression analyses.

Minimal detectable change

Minimal detectable changes (MDC) were only computed for respiratory sound parameters and locations that have simultaneously shown adequate between-days reliability ($\text{ICC} > 0.75$) and validity (significant correlations with $\text{FEV}_1\%$ predicted).

To determine the MDC, first, the standard error of measurement (SEM) was calculated using the **equation 4**:

$$SEM = SD \sqrt{(1 - ICC_{1,2})} \quad (4)$$

where SD is the standard deviation of the scores obtained from all participants and ICC is the between-days reliability coefficient.

The MDC at the 95% level of confidence (MDC₉₅) was calculated as follows (**equation 5**):

$$MDC_{95} = SEM \times 1.96 \times \sqrt{2} \quad (5)$$

The MDC was also expressed as a percentage (MDC%), defined as (equation 6):

$$MDC_{\%} = (MDC_{95}/mean) \times 100 \quad (6)$$

where “mean” is the mean of the scores obtained in the two testing sessions. A MDC% below 30% was considered acceptable (34).

Results

Participants

Fifty-eight patients were contacted and invited to participate in the study. However, seven refused to participate, as they did not perceive the study as relevant (n=5) or had family constraints to their participation (n=2), and one did not complete the assessment. Therefore, 50 participants (36 males) were enrolled in the study. Participants' characteristics are summarized in Table 1.

Please insert Table 1 here

Reliability

Adventitious respiratory sounds presented excellent within-day reliability ($ICC_{1,3} > 0.75$); except at trachea in both respiratory phases ($0.57 < ICC_{1,3} < 0.74$) and at anterior chest during expiration ($0.65 < ICC_{1,3} < 0.73$) for the number of crackles, and at anterior right chest during both phases ($0.51 < ICC_{1,3} < 0.68$) for %Wh. F50 and I_{max} also presented excellent reliability, except at anterior right chest during inspiration ($0.51 < ICC_{1,3} < 0.73$). Absolute reliability showed no systematic bias for any location and/or respiratory phase according to the Bland and Altman plots. Further information of within-day reliability is on supplementary material.

Table 2 presents the relative between-days reliability. During inspiration, crackles and wheezes showed moderate-to-good or excellent reliability ($0.48 \leq ICC_{1,2} \leq 0.96$; $p < 0.05$) at anterior and posterior chest. F50 and I_{max} showed moderate-to-good or excellent reliability ($ICC_{1,2} > 0.45$ and $ICC_{1,2} \geq 0.60$, respectively; $p < 0.05$), except at posterior left chest ($ICC_{1,2} < 0.41$; $p > 0.05$). During expiration, %Wh and normal respiratory sounds were reliable at trachea and at the anterior chest ($ICC_{1,2} > 0.54$; $p < 0.05$), and the number of crackles were only reliable at trachea ($ICC_{1,2} = 0.79$; $p < 0.05$).

Please insert Table 2 here

Good absolute between-days reliability with no systematic bias was found in the Bland and Altman plots for number of crackles and normal respiratory sounds. However, large limits of agreement were found at trachea for all respiratory sound parameters and for %Wh in all locations, especially during expiration. Figure 1 and 2 shows the Bland-Altman plots obtained at posterior right and left chest, respectively. The remaining plots can be found in the supplementary material.

Please insert Figure 1 and Figure 2 here

Validity

Concerning criterion validity, significant low-to-moderate negative correlations ($-0.53 < r_s < -0.32$; $p < 0.05$) between FEV₁%predicted and adventitious respiratory sounds were found, especially for the number of crackles during inspiration. Significant correlations were also found for normal respiratory sounds, being negative for inspiratory F50 and positive for I_{max}, especially during inspiration. Table 3 presents the correlations between FEV₁%predicted and computerised respiratory sounds.

Please insert Table 3 here

AUCs of all variables analysed ranged from 0.27 to 0.81, indicating “no discrimination” to “acceptable discrimination”. Higher AUCs were found for inspiratory number of crackles recorded at posterior right (AUC=0.78; 95%CI=0.51-1.00; $p < 0.001$) and left (AUC=0.81; 95%CI=0.68-0.93; $p < 0.001$) chest (Figure 3). To differentiate between participants with mild-to-moderate from participants with severe-to-very severe airflow limitation, cutoff points of 0.1 (sensitivity=81%; specificity=71%) and of 0.5 (sensitivity=74%; specificity=80%) for the mean number of crackles at posterior right and left chest, respectively, were identified. Results from the ROC analysis of all computerised respiratory sounds are in the supplementary material.

Please insert Figure 3 here

Concerning construct validity, significant low positive ($r_s < 0.48$; $p < 0.05$) correlations were found between patient-reported outcome measures and computerised respiratory sounds. Values for all correlations are in the supplementary material.

Concerning predictive validity, both adventitious and normal respiratory sounds showed no ability to predict exacerbations up to one year, with AUCs ranging from 0.00 to 0.58 ($p > 0.05$). None of the computerised respiratory sound parameters were predictors of the time until the first exacerbation ($p > 0.05$; hazard ratios between 0.95 and 1.04).

Minimal detectable change

The respiratory sounds parameters presenting adequate reliability and validity were inspiratory number of crackles at posterior right chest, inspiratory F50 at trachea and anterior left chest, and expiratory I_{max} at anterior right chest. The MDC₉₅ was 2.41 (SEM=0.87; MDC%=175.38%), 55.27 (SEM=19.94; MDC%=41.22%), 29.55 (SEM=10.66; MDC%=31.86%) and 3.98 (SEM=1.43; MDC%=35.47%) for number of crackles, F50 at trachea, F50 at anterior left chest and I_{max}, respectively.

Discussion

To the authors' best knowledge, this is the first study describing computerised respiratory sounds reliability and validity according to the COSMIM guidelines. The main findings indicate that respiratory sounds i) present moderate-to-excellent within-day reliability and moderate-to-good between-days reliability; ii) are valid to express lung function, especially inspiratory number of crackles at posterior chest; iii) correlate poorly with patient-reported outcome measures; iv) do not predict COPD exacerbations and v) present high values of MDC.

Moderate-to-excellent within-day reliability was found for all respiratory sound parameters, which is in line with data previously reported (ICC_{1,3} from 0.66 to 0.89) (10). Regarding to between-days reliability, slightly lower values were found. This was expected, as it is known that better reliability is achieved when repeated tests are performed within short periods of time (35). The number of inspiratory crackles, recorded at posterior chest, and inspiratory and expiratory %Wh, recorded at anterior chest, were the most reliable parameters. It is known that COPD is characterized by

changes in airflow mechanics targeting mainly the smaller airways (1), thus inspiratory crackles at posterior regions have been indicated as the most common and persistent finding in these patients (10,36). Wheezes are also a usual characteristic of patients with bronchial obstruction (37). In the present study, inspiratory wheezes were slightly more reliable than expiratory wheezes, which may be explained by their genesis. Patients with COPD usually experience expiratory low frequency wheezes (also known as rhonchi) in upper airways that are generally produced by increased sputum and are easily removed by cough (36,38). On the other hand, inspiratory wheezes are more related with severe airway obstruction (37), which characterizes most of our sample, and thus are more difficult to change with respiratory manoeuvres.

Similar to what has been previously described for within-day reliability (10), the agreement assessed with the Bland-Altman method was found to be acceptable for mean number of crackles and normal respiratory sounds intensity. However, high limits of agreement were found at trachea and expiratory wheezes, possibly due to their dependence on airflow and respiratory manoeuvres. Therefore, this anatomic location and respiratory sound parameter may not be suitable to use in interventions studies of patients with stable COPD.

Overall, respiratory sounds showed low-to-moderate correlations with FEV₁%predicted, being the strongest correlations found for the number of crackles at posterior regions. The presence of inspiratory crackles at the posterior chest also showed to be discriminative of participants with more and less severe airflow limitation. FEV₁ mainly expresses obstruction in larger airways (2), while COPD pathogenesis primarily targets small airways (1). On the other hand, crackles are an acoustic phenomenon that, when heard over distal lung regions, are associated with inflammation or oedema (39) of smaller airways. Thus, whilst high correlations between these outcome measures would not be expected, the significant correlations found, highlight crackles potential to indicate peripheral airway obstruction in patients with COPD.

Significant and lower correlations were also found between FEV₁%predicted and I_{max} at the anterior chest. Similar results have been reported (40) and might be related with a shift of regional ventilation from the lower- to upper-lung locations as a function of

the degree of hyperinflation presented in more severe COPD patients. This type of hyperinflation can be reversed with therapy (41) and inspiratory maximum intensity recorded at the anterior chest may be an adequate outcome measure to assess these changes. F50 has been the respiratory sound parameter most described as related with lung function (42-44) however, in this study only poor correlations were found between F50 and FEV₁. Different methodologies used across studies, i.e., no additional maneuvers or forced expiratory maneuvers or chemical substances that cause airway obstruction during respiratory sound recordings, might explain this incongruence. Nevertheless, our results mimic those from Malmberg et al., (1995) (45) where a similar protocol has been used. Therefore, protocol standardisation to record respiratory sounds is needed to advance knowledge in this field.

Regarding construct validity, significant and negative, although low, correlations were found between respiratory sounds and patient-reported outcome measures. Similar results have been found for FEV₁ ($0.14 < r < 0.41$) (5) and for respiratory sounds ($0.33 < r < 0.57$) (46) in previous studies and further confirms that clinical outcome measures significantly differ from the individuals' experience of the disease effects on health status and hence, should not be used isolated (47).

Respiratory sounds presented no ability to predict exacerbations up to one year after the baseline assessment. Although changes in the tracheobronchial tree are closely related with changes in respiratory acoustics (48), such associations are likely to be unravelled in time periods close to exacerbations, when the beginning of the inflammatory and/or infectious process has occurred. Additionally, although COPD is mainly characterised by changes in smaller airways, COPD exacerbations are frequently triggered by upper respiratory tract infections (49), thus it is reasonable that predictions of these events are better detected by changes in larger airways. Indeed, recent studies have shown that computerised respiratory sounds, recorded at trachea, have potential to predict exacerbations in the short term (i.e., 5 days ahead of medical attention), however such predictions have been determined based on complex analysis (principal component analysis), that cannot be easily understood and applicable by clinicians in clinical practice (50). Our results have shown that F50 and I_{max} recorded at upper anatomical locations are valid and reliable parameters and can be more easily determined or even perceived by clinicians during auscultation.

Thus, we recommend future studies on COPD telemonitoring to explore the efficacy of these normal respiratory parameters to early detect exacerbations.

High MDC values were found for all respiratory sound parameters which might be related with patients' high inter-variability, as reported in previous studies (10). These results highlight the importance of supporting health-care professionals' clinical decisions in the interpretation of respiratory sound changes at an individual level and in combination with other patient-reported outcome measures. Nevertheless, this was the first study to calculate MDC for respiratory sounds and provides a valuable cut-off point to represent minimum detectable change in repeated measures beyond the threshold of error.

Limitations

This study has some potential limitations that need to be discussed. Flows and/or volumes were not controlled during respiratory sounds recordings and it is known that respiratory sound acoustics depends on volume and rate of respiratory maneuvers (48). However, this can be arguable for the purposes of this study as it has been previously demonstrated that even without airflow control, respiratory sounds present adequate reliability ($ICC > 0.70$) and are almost as reliable as during recordings at controlled flows (10). Moreover, this study was designed to be as close to clinical practice as possible, and currently, equipment for airflow monitoring is expensive, little portable and requires trained professionals for its interpretation, which hinders its use in such settings.

This study followed the COSMIN methodological recommendations to test the suitability of an outcome measure to be implemented in the clinical practice. The COSMIN was originally developed for health-related, patient-reported outcome measures, such as questionnaires (11). Therefore, the application of the COSMIN as a tool for guiding methodology of studies testing clinical outcome measures can be questioned. Nonetheless, in the absence of guidelines specifically designed to conduct such studies, the COSMIN is indicated as an adequate alternative tool (51).

Conclusion

The number of crackles recorded at posterior locations and the normal respiratory sounds recorded at trachea and anterior regions are reliable and valid parameters to assess and monitor patients with stable COPD. Nonetheless, results from criterion and

construct validity showed that computerised respiratory sounds should not be used isolated, but rather integrated with other clinical and patient reported outcome measures, as they may fill the gap of assessing small airways and contribute towards a patient's comprehensive evaluation.

Acknowledgments

This work was supported by Fundo Europeu de Desenvolvimento Regional (FEDER) through Programa Operacional Competitividade e Internacionalização (COMPETE) and Fundação para a Ciência e Tecnologia (FCT) under the project UID/BIM/04501/2013 and SFRH/BD/101951/2014 and partially supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. (CAPES) – grant number 88881.134901/2016-01. The authors would also like to acknowledge to Marta João Oliveira for her assistance in the edition of images.

References

1. The Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease - 2017 Update. In: The Global Initiative for Chronic Obstructive Lung Disease, Inc.; 2017.
2. McNulty W, Usmani OS. Techniques of assessing small airways dysfunction. *Eur Clin Respir J.* 2014;1:10.3402/ecrj.v3401.25898.
3. Zwick RH, Burghuber OC, Dovjak N, et al. The effect of one year outpatient pulmonary rehabilitation on patients with COPD. *Wien Klin Wochenschr.* 2009;121(5):189.
4. Calverley P, Boonsawat W, Cseke Z, Zhong N, Peterson S, Olsson H. Maintenance therapy with budesonide and formoterol in chronic obstructive pulmonary disease. *Eur Respir J.* 2003;22(6):912-919.
5. Jones PW. Health status measurement in chronic obstructive pulmonary disease. *Thorax.* 2001;56(11):880-887.
6. Celli BR, Decramer M, Wedzicha JA, et al. An Official American Thoracic Society/European Respiratory Society Statement: Research questions in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 2015;191(7):e4-e27.
7. Gavriely N, Nissan M, Cugell DW, Rubin AH. Respiratory health screening using pulmonary function tests and lung sound analysis. *Eur Respir J.* 1994;7(1):35-42.
8. Jacome C, Oliveira A, Marques A. Computerized respiratory sounds: a comparison between patients with stable and exacerbated COPD. *Clin Respir J.* 2015;[Epub ahead of print].
9. Oliveira A, Pinho C, Marques A. Effects of a respiratory physiotherapy session in patients with LRTI: a pre/post-test study. *Clin Respir J.* 2015;[Epub ahead of print].
10. Jácome C, Marques A. Computerized respiratory sounds are a reliable marker in subjects with COPD. *Respir Care.* 2015;60(9):1264-1275.
11. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60(1):34-42.
12. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012;21(4):651-657.
13. Morris N, Sabapathy S, Adams L, Kingsley R, Schneider D, Stulbarg M. Verbal numerical scales are as reliable and sensitive as visual analog scales for rating dyspnea in young and older subjects. *Respiratory physiology & neurobiology.* 2007;157(2):360-365.
14. Boulet L-P, Coeytaux RR, McCrory DC, et al. Tools for assessing outcomes in studies of chronic cough: CHEST guideline and expert panel report. *Chest.* 2015;147(3):804-81415. Doherty DE, Belfer MH, Brunton

- SA, Fromer L, Morris CM, Snader TC. Chronic obstructive pulmonary disease: consensus recommendations for early diagnosis and treatment. *J Fam Pract.* 2006;55(11):S1-S1.
16. Mahler DA, Ward J, Waterman LA, McCusker C, ZuWallack R, Baird JC. Patient-reported dyspnea in COPD reliability and association with stage of disease. *Chest.* 2009;136(6):1473-1479.
 17. Bestall J, Paul E, Garrod R, Garnham R, Jones P, Wedzicha J. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax.* 1999;54(7):581-586.
 18. Jones P, Harding G, Berry P, Wiklund I, Chen W, Leidy NK. Development and first validation of the COPD Assessment Test. *Eur Respir J.* 2009;34(3):648-654.
 19. Vannuccini L, Earis J, Helisto P, et al. Capturing and preprocessing of respiratory sounds. *Eur Respir Rev.* 2000;10(77):616-620.
 20. Rossi M, Sovijarvi A, Piirila P, Vannuccini L, Dalmaso F, Vanderschoot J. Environmental and subject conditions and breathing manoeuvres for respiratory sound recordings. *Eur Respir Rev.* 2000;10(77):611-615.
 21. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J.* 2005;26(2):319-338.
 22. Pinho C, Oliveira A, Oliveira D, Dinis J, Marques A. LungSounds@ UA interface and multimedia database. *International Journal of E-Health and Medical Communications (IJEHMC).* 2014;5(1):81-95.
 23. Kraman SS, Wodicka GR, Oh Y, Pasterkamp H. Measurement of respiratory acoustic signals: effect of microphone air cavity width, shape, and venting. *Chest.* 1995;108(4):1004-1008.
 24. Vyshedskiy A, Murphy R. Crackle pitch rises progressively during inspiration in pneumonia, CHF, and IPF patients. *Pulm Med.* 2012;2012.
 25. Pasterkamp H, Powell RE, Sanchez I. Lung sound spectra at standardized air flow in normal infants, children, and adults. *Am J Respir Crit Care Med.* 1996;154(2):424-430.
 26. Hossain I, Moussavi Z. Finding the lung sound-flow relationship in normal and asthmatic subjects. Paper presented at: Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE2004.
 27. Sulzer J, Schüttler F, Penzel T, Wichert Pv. A simple method for detecting bronchial breathing sound. Paper presented at: Proceedings of the 22nd International Conference on Lung Sounds (October 15–17, 1997). International Lung Sounds Association, Nippon Medical School, Tokyo, Japan1997.
 28. Oliveira A, Sen I, Kahya YP, Afreixo V, Marques A. Computerised respiratory sounds can differentiate smokers and non-smokers. *J Clin Monit Comput.* 2017;31(3):571-580.

29. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil.* 1998;12(3):187-199.
30. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;327(8476):307-310.
31. J F. In: J F, ed. *Design and analysis of clinical experiments.* New York: John Wiley & Sons; 1986:1-32.
32. Mukaka M. A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J.* 2012;24(3):69-71.
33. Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied logistic regression.* Vol 398: John Wiley & Sons; 2013.
34. Portney LG, Watkins MP. *Foundations of clinical research: applications to practice.* Vol 2: Prentice Hall Upper Saddle River, NJ; 2000.
35. Shin SH, Ro du H, Lee OS, Oh JH, Kim SH. Within-day reliability of shoulder range of motion measurement with a smartphone. *Manual therapy.* 2012;17(4):298-304.
36. Jacome C, Marques A. Computerized respiratory sounds in patients with COPD: a systematic review. *COPD.* 2015;12(1):104-112.
37. Meslier N, Charbonneau G, Racineux JL. Wheezes. *Eur Respir J.* 1995;8(11):1942-1948.
38. Bohadana A, Izbicki G, Kraman SS. Fundamentals of lung auscultation. *N Engl J Med.* 2014;370(8):744-751.
39. Piirila P, Sovijarvi AR. Crackles: recording, analysis and clinical significance. *Eur Respir J.* 1995;8(12):2139-2148.
40. Mineshita M, Kida H, Handa H, et al. The correlation between lung sound distribution and pulmonary function in COPD patients. *PloS one.* 2014;9(9):e107506.
41. Mineshita M, Matsuoka S, Miyazawa T. Effects of bronchodilators on regional lung sound distribution in patients with chronic obstructive pulmonary disease. *Respiration.* 2014;87(1):45-53.
42. Malmberg LP, Sovijarvi AR, Paajanen E, Piirila P, Haahtela T, Katila T. Changes in frequency spectra of breath sounds during histamine challenge test in adult asthmatics and healthy control subjects. *Chest.* 1994;105(1):122-131.
43. Spence DP, Bentley S, Evans DH, Morgan MD. Effect of methacholine induced bronchoconstriction on the spectral characteristics of breath sounds in asthma. *Thorax.* 1992;47(9):680-683.
44. Anderson K, Aitken S, Carter R, MacLeod JE, Moran F. Variation of breath sound and airway caliber induced by histamine challenge. *Am Rev Respir Dis.* 1990;141(5 Pt 1):1147-1150.
45. Malmberg LP, Pesu L, Sovijarvi AR. Significant differences in flow standardised breath sound spectra in patients with chronic obstructive

- pulmonary disease, stable asthma, and healthy lungs. *Thorax*. 1995;50(12):1285-1291.
46. Jácome C, Marques A. Computerized Respiratory Sounds: Novel Outcomes for Pulmonary Rehabilitation in COPD. *Respir Care*. 2017;62(2):199-208.
 47. Westwood M, Bourbeau J, Jones PW, Cerulli A, Capkun-Niggli G, Worthy G. Relationship between FEV1 change and patient-reported outcomes in randomised trials of inhaled bronchodilators for stable COPD: a systematic review. *Respir Res*. 2011;12:40.
 48. Pasterkamp H, Kraman SS, Wodicka GR. Respiratory sounds. Advances beyond the stethoscope. *Am J Respir Crit Care Med*. 1997;156(3 Pt 1):974-987.
 49. Wedzicha JA, Donaldson GC. Exacerbations of chronic obstructive pulmonary disease. *Respir Care*. 2003;48(12):1204-1213.
 50. Fernandez-Granero MA, Sanchez-Morillo D, Leon-Jimenez A. Computerised Analysis of Telemonitored Respiratory Sounds for Predicting Acute Exacerbations of COPD. *Sensors (Basel)*. 2015;15(10):26978-26996.
 51. Bartels B, de Groot JF, Terwee CB. The six-minute walk test in chronic pediatric conditions: a systematic review of measurement properties. *Phys Ther*. 2013;93(4):529-541.

Figure legend

Figure 1. Bland and Altman plots of number of crackles and wheeze occupation rate (%Wh) collected at session 1 (S1) and session 2 (S2) at posterior right and left chest. Solid lines represent the zero value dashed lines show the associated bias and 95% upper (ULA) and lower (LLA) limits of agreement.

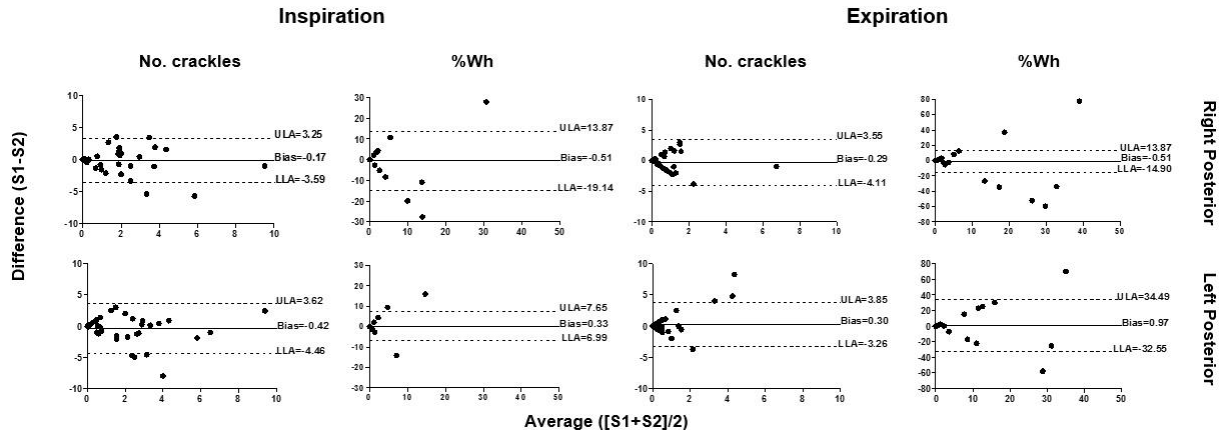


Figure 2. Bland and Altman plots of median frequency (F50) and maximum intensity (Imax) collected at session 1 (S1) and session 2 (S2) at posterior right and left chest. Solid lines represent the zero value dashed lines show the associated bias and 95% upper (ULA) and lower (LLA) limits of agreement.

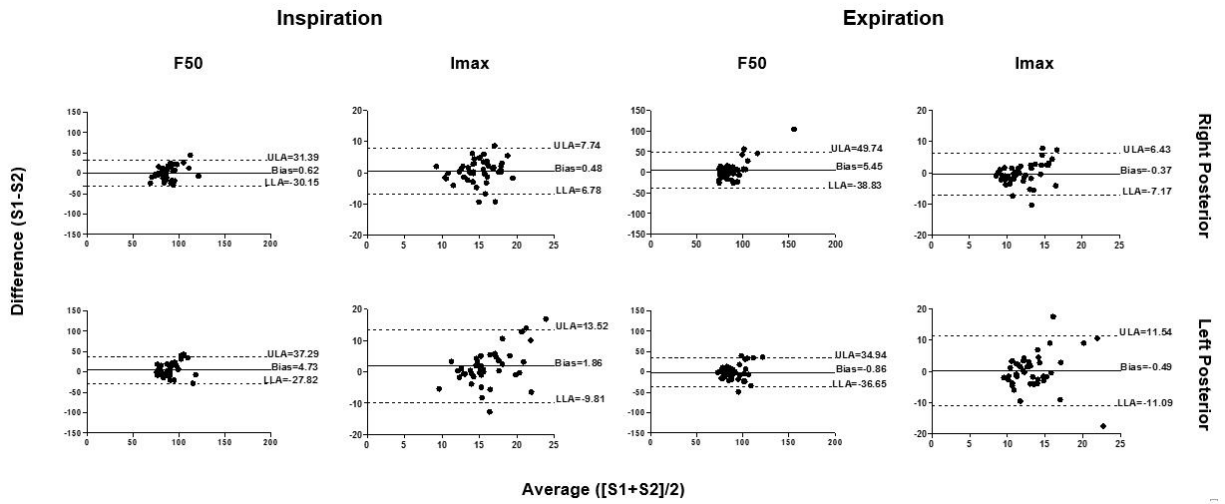
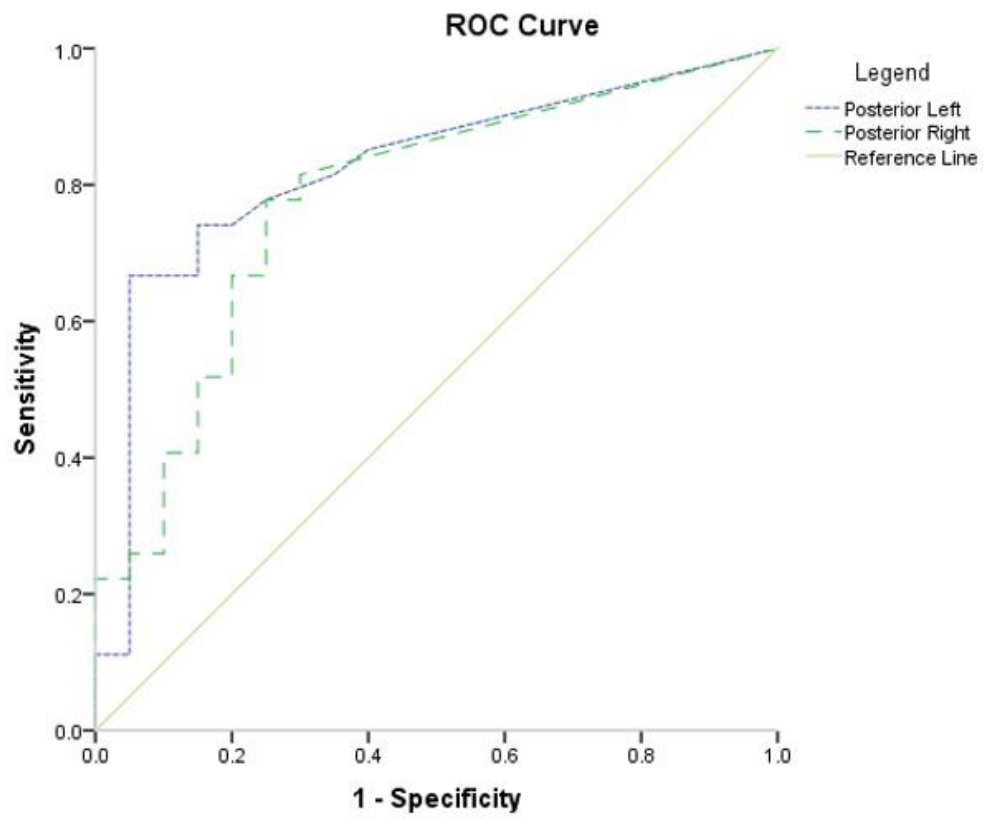


Figure 3. Receiver operator characteristics (ROC) of the inspiratory number of crackles at posterior right and left chest to differentiate between participants with mild-to-moderate airflow limitation and participants with severe-to-very severe airflow limitation.



Tables and captions

Table 1. Sample characterization.

Characteristics	Total (n=50)	GOLD stages I-II (n=21)	GOLD stages III-IV (n=29)	p-value
Age, years	67.26±9.31	67.29±11.22	67.24±7.87	0.987
Gender (male), n(%)	36 (72)	13 (62)	23 (79)	0.213
BMI, kg/m ²	27.26±8.22	29.93±11.26	25.32±4.28	0.049*
Smoking status, n(%)				0.056
Current	7 (14)	5 (24)	2 (7)	
Former	26 (52)	7 (33)	19 (66)	
Never	17 (34)	9 (43)	8 (28)	
Packs/year, M[IQR]	50 [32-77]	48 [24-54]	50 [33-90]	0.294
Exacerbations/year, n(%)				0.638
0	18 (36)	6 (29)	12 (41)	
1	9 (18)	4 (19)	5 (18)	
≥2	23 (46)	11 (52)	12 (41)	
FEV ₁ , L	1.24±0.53	1.65±0.53	0.95±0.28	<0.001*
FEV ₁ , %predicted	49.52±19.67	69.10±10.65	35.34±10.04	<0.001*
FEV ₁ /FVC, %	49.76±13.24	58.71±8.65	43.28±12.24	<0.001*
GOLD stages, n(%)				
I	3 (6)			
II	18 (36)			
III	21 (42)			
IV	8 (16)			
CAT, M[IQR]	13 [8-21]	12 [9-18]	14 [7-23]	0.582
mMRC, M[IQR]	2 [1-2]	1 [1-2]	2 [1-3]	0.004*
NS, M[IQR]				
Cough	1 [0-3]	2 [0-5]	1 [0-3]	0.233
Wheezing	2 [0-4]	2 [0-3]	3 [0-5]	0.236

*p<0.05

Values are presented as mean±standard deviation, unless otherwise stated.

Legend: BMI, body mass index; CAT, COPD Assessment Test; FEV₁, forced expiratory volume in one second; FVC, forced vital capacity; GOLD, Global Initiative for Chronic Obstructive Lung Disease; IQR, interquartile range; M, median; mMRC, Modified British Medical Research Council questionnaire; SD, standard deviation; NS, numerical scale.

Table 2. Between-days reliability (ICC_{1,2}) for normal and adventitious respiratory sounds.

	Inspiration				Expiration			
	No. crackles	%Wh	F50	I _{max}	No. crackles	%Wh	F50	I _{max}
Trachea	0.38	0.62*	0.79*	0.62*	0.79*	0.69*	0.71*	0.84*
	[-0.12-0.69]	[0.31-0.79]	[0.62-0.83]	[0.32-0.79]	[0.62-0.88]	[0.44-0.83]	[0.47-0.84]	[0.72-0.91]
Anterior	0.63*	0.92*	0.51*	0.60*	0.44	0.54*	0.54*	0.78*
right	[0.33-0.796]	[0.86-0.96]	[0.12-0.73]	[0.27-0.78]	[-0.02-0.69]	[0.16-0.74]	[0.16-0.75]	[0.60-0.88]
Anterior	0.88*	0.96*	0.86*	0.73*	0.36	0.70*	0.75*	0.60*
left	[0.78-0.94]	[0.92-0.96]	[0.73-0.93]	[0.48-0.86]	[-0.22-0.66]	[0.43-0.85]	[0.52-0.87]	[0.24-0.79]
Posterior	0.79*	0.57*	0.47*	0.65*	0.42	-0.01	0.39	0.43
right	[0.62-0.89]	[0.21-0.76]	[0.03-0.71]	[0.35-0.81]	[-0.05-0.68]	[-0.85-0.45]	[-0.11-0.66]	[-0.03-0.69]
Posterior	0.74*	0.48*	0.41	0.15	0.25	0.07	0.31	0.07
left	[0.52-0.86]	[0.04-0.72]	[-0.10-0.68]	[-0.58-0.54]	[-0.39-0.59]	[-0.73-0.50]	[-0.29-0.63]	[-0.72-0.50]

*p<0.05

Values are presented as ICC_{1,2}[95% confidence interval].

Legend: %Wh, wheeze occupation rate; F50, median frequency; I_{max}, maximum intensity.

Table 3. Correlations between lung function (FEV₁ %predicted) and computerised respiratory sounds.

	Inspiration				Expiration				FEV ₁ %predicted
	No. crackles	%wh	F50	I _{max}	No. crackles	%wh	F50	I _{max}	
Trachea	r _s =-0.07	r _s =-0.09	r _s =-0.35*	r _s =0.28*	r _s =-0.20	r _s =-0.37*	r _s =-0.18	r _s =0.26	
Anterior right	r _s =-0.11	r _s =-0.09	r _s =-0.18	r _s =0.32*	r _s =-0.16	r _s =-0.20	r _s =0.04	r _s =0.32*	
Anterior left	r _s =-0.42*	r _s =-0.13	r _s =-0.37*	r _s =0.36*	r _s =-0.53*	r _s =-0.18	r _s =0.04	r _s =0.03	
Posterior right	r _s =-0.44*	r _s =-0.23	r _s =0.06	r _s =0.23	r _s =-0.11	r _s =-0.21	r _s =0.02	r _s =0.28	
Posterior Left	r _s =-0.42*	r _s =-0.22	r _s =0.02	r _s =0.07	r _s =-0.16	r _s =-0.12	r _s =0.14	r _s =-0.08	

*p<0.05

Values are presented as Spearman's correlations.

Legend: %wh, wheeze occupation rate; F50, median frequency; I_{max}, maximum intensity.