RESPIRATORY CARE Paper in Press. Published on March 29, 2016 as DOI: 10.4187/respcare.04547

Reproducibility of the Sputum Color Evaluation Depends on the Category of Caregivers

Gregory Reychler PhD PT, Emmanuel Andre MD, Laurence Couturiaux PT, Kinga Hohenwarter MD, Giuseppe Liistro PhD MD, Thierry Pieters MD, and Annie Robert PhD MSc

BACKGROUND: Sputum production and purulence were proposed as criteria for justifying the use of antimicrobial agents. The Sputum Color Chart was developed and validated to standardize purulence of sputum evaluation. The aim of this study was to observe the reproducibility of the Sputum Color Chart from different categories of health caregivers. METHODS: The color of 10 sputum samples was evaluated using photographs for intra- and inter-reliability. The observation was repeated 3 times. Eighteen volunteers from 6 categories of health caregivers (student in physiotherapy, senior chest physiotherapist, junior resident in pulmonology, medical microbiologist, pulmonologist, and general practitioner) were investigated. RESULTS: Poor inter-rater reliability was observed for all categories with the exception of senior chest physiotherapists. The best intra-rater reliability was observed for microbiologists and senior chest physiotherapists. We found a great proportion (>40%) of important discrepancies in 2 categories (junior pulmonologist and general practitioner). The proportion of non-discrepancy between evaluators varied between 10 and 40%, depending on the category. CONCLUSIONS: Even if the Sputum Color Chart is a useful tool for the clinician in the context of clinical deterioration, it presents non-uniform reliability regarding the caregivers and their category. *Key words: sputum; reproducibility; evaluation; COPD; intra-rater; inter-rater.* [Respir Care 0;0(0):1–•. © 0 Daedalus Enterprises]

Introduction

Sputum is a common symptom in various pulmonary diseases. Indeed, sputum production is typical in cystic fibrosis or COPD for physiopathological reasons, and patients with bronchiectasis frequently present with a long-standing productive cough.[1]

Sputum production is also associated with inflammatory processes in the airways[2] and with the progression of obstruction[3] in smokers. Additionally, it is related to the mortality rate in patients with COPD.[4]

Sputum color is related to the degree of infection. Bacterial colonization has been demonstrated to be low (5%), moderate (43.5%), and very important (86.4%) in mucoid, mucopurulent, and purulent sputum, respectively ($P < .001$).[5] The purulence of sputum has also been associated with the bacterial load during exacerbations in subjects with COPD.[6]

Drs Reychler, Liistro, and Pieters are affiliated with the Institut de Recherche Expérimentale et Clinique (IREC), Pôle de Pneumologie, ORL, and Dermatologie, Université Catholique de Louvain, Brussels, Belgium and the Service de Pneumologie, Cliniques Universitaires Saint-Luc, Brussels, Belgium. Dr Reychler is also affiliated with the Institut Parnasse-ISEI, Brussels, Belgium. Dr Andre is affiliated with the Institut de Recherche Expérimentale et Clinique (IREC), Pôle de Microbiologie, Université Catholique de Louvain, Brussels, Belgium and the Service de Microbiologie, Cliniques Universitaires Saint-Luc, Brussels, Belgium. Mr Couturiaux is affiliated with the Institut Parnasse-ISEI, Brussels, Belgium. Dr Hohenwarter is affiliated with the Department of Hygiene and Microbiology, Klinikum Wels-Grieskirchen, Wels, Austria. Dr Robert is affiliated with the Institut de Recherche Expérimentale et Clinique (IREC), Pôle Epidémiologie et Biostatistique (EPID), Université Catholique de Louvain, Brussels, Belgium.

The authors have disclosed no conflicts of interest.

Correspondence: Gregory Reychler PhD PT, Pneumology Unit, Cliniques Universitaires St-Luc (UCL), Avenue Hippocrate 10, 1200 Brussels, Belgium. E-mail: gregory.reychler@uclouvain.be.

DOI: 10.4187/respcare.04547

RESPIRATORY CARE Paper in Press. Published on March 29, 2016 as DOI: 10.4187/respcare.04547

The decision to prescribe an antimicrobial agent in patients with an exacerbation is often difficult. Sputum production and purulence were proposed as criteria for defining an exacerbation and justifying the use of antimicrobial agents.[7] Usually, the prescription is ordered empirically, based on clinical evidence of purulence. Hence, the color of sputum could play an important role in clinical practice in lung diseases with productive cough even if that role is still debated.[8] It has been proposed to take into account the sputum color to determine the clinical impact of COPD (the current consequences of the disease for the patient).[9] However, the color can be difficult to evaluate, and when reported by patients, it is not reliable.[10]

For this purpose, Murray et al[5] developed and validated a quantitative method (Sputum Color Chart) to standardize the evaluation. It allows clinicians to report sputum color by providing an accurate representation of the 3 major grades of color.

Such validated tools are necessary to facilitate management. A good inter-rater reproducibility was observed between the physicians and patients.[5] However, some other psychometrics properties must be investigated as they are for questionnaires. A complete tool validation requires the evaluation of its reproducibility. Reproducibility concerns the degree to which repeated measurements in steady state provide similar answers. It includes reliability and agreement. These properties are usually evaluated by the weighted Cohen's kappa coefficient and the Bland-Altman analysis, respectively.[11] Reliability concerns the power of distinction between patients.[11] Agreement concerns the absolute measurement error, which refers to clinically important changes.[11] The aim of this study was to observe the reproducibility of the Sputum Color Chart for 3 evaluators from 6 categories of health caregivers who can be routinely faced with sputum evaluation in their clinical practice.

QUICK LOOK

Current knowledge

Sputum color is related to the degree of infection, and the purulence of sputum has been associated with bacterial load in subjects with COPD exacerbation. The Sputum Color Chart was developed and validated to standardize the purulence of sputum evaluation.

What this paper contributes to our knowledge

Among health caregivers, we found poor inter-rater reliability for all categories except senior chest physiotherapists. Although the Sputum Color Chart may be useful in the context of assessing clinical deterioration, it is not reliable across health caregiver categories.

Methods

Evaluators

Volunteers with different clinical backgrounds and professional experience were recruited in a tertiary hospital. Eighteen volunteers from 6 categories of caregivers were investigated: 3 students in the last year of physiotherapy, 3 senior chest physiotherapists, 3 junior residents in pulmonology, 3 medical microbiologists, 3 pulmonologists, and 3 general practitioners. The single exclusion criterion was a volunteer who presented with a diagnosis of color vision deficiency. All of the volunteers were unfamiliar with the Sputum Color Chart.

Design

Ten consecutive sputum samples in doublet were collected from hospitalized patients independent of their characteristics and bacteriology. The selection of samples was based on consecutive bacteriological analysis requests. The test was ordered by a physician independent of the study. One sample was used for the study, and the other one was sent to the microbiology laboratory for bacteriological analysis. The 2 inclusion criteria were the presence of expectoration and a diagnosis of chronic respiratory disease. These 10 sputum samples were placed on a white support and were photographed. The pictures were sent by electronic message to the evaluators for reading with the Sputum Color Chart[5] on a similar screen. The Sputum Color Chart is a chart using photographs of sputum representing the 3 typical gradations of color (mucoid [clear], mucopurulent [pale yellow or green], and purulent [dark yellow or green]). The reading was repeated 3 times.

Statistics

Statistics were evaluated using SPSS 22.0 (IBM Corporation, Armonk, New York). A descriptive analysis was performed to describe the readings. Discrepancies (defined by the maximal difference between 3 readings) were observed in the triplicate readings of the same sputum by one evaluator and in the first reading by 3 evaluators of the same health caregiver category. Eighteen raters were included in the study because it was suggested that kappa can be validly applied when the number of subjects being rated is $>2 \times C^2$ (where C is the number of categories in the assessment tool).[12]

Cohen's kappa coefficient was calculated and Bland-Altman analysis was conducted to evaluate the reliability and the agreement, respectively. Kappa values were interpreted as follows: >0.80 was very good, 0.61–0.80 was good, 0.41–0.60 was moderate, 0.21–0.40 was fair, and <0.21 was poor.[13] Bland-Altman analysis was performed

RESPIRATORY CARE Paper in Press. Published on March 29, 2016 as DOI: 10.4187/respcare.04547

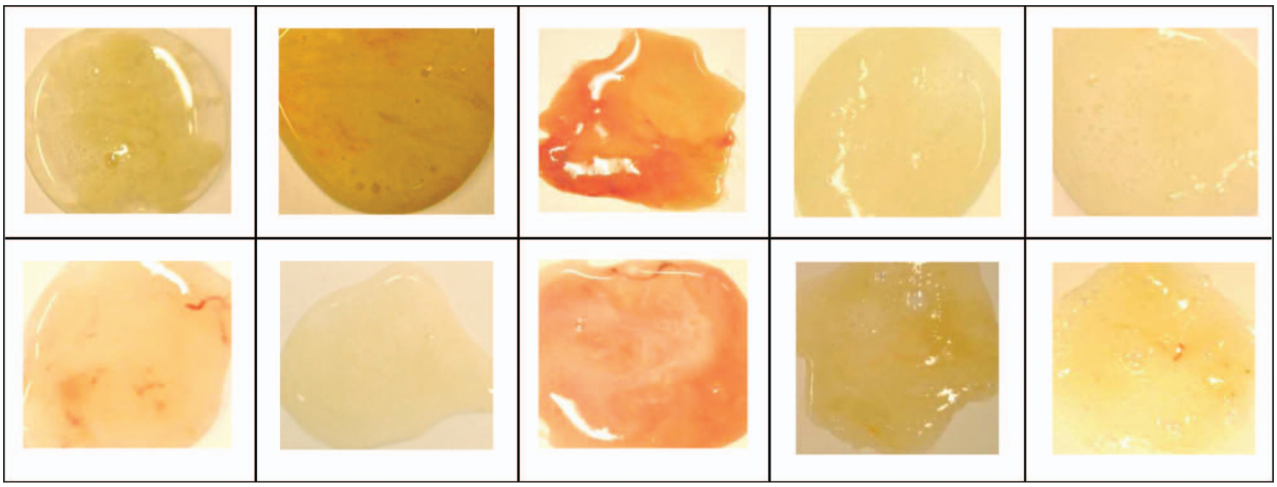

Fig. 1. Illustration of sputum samples.

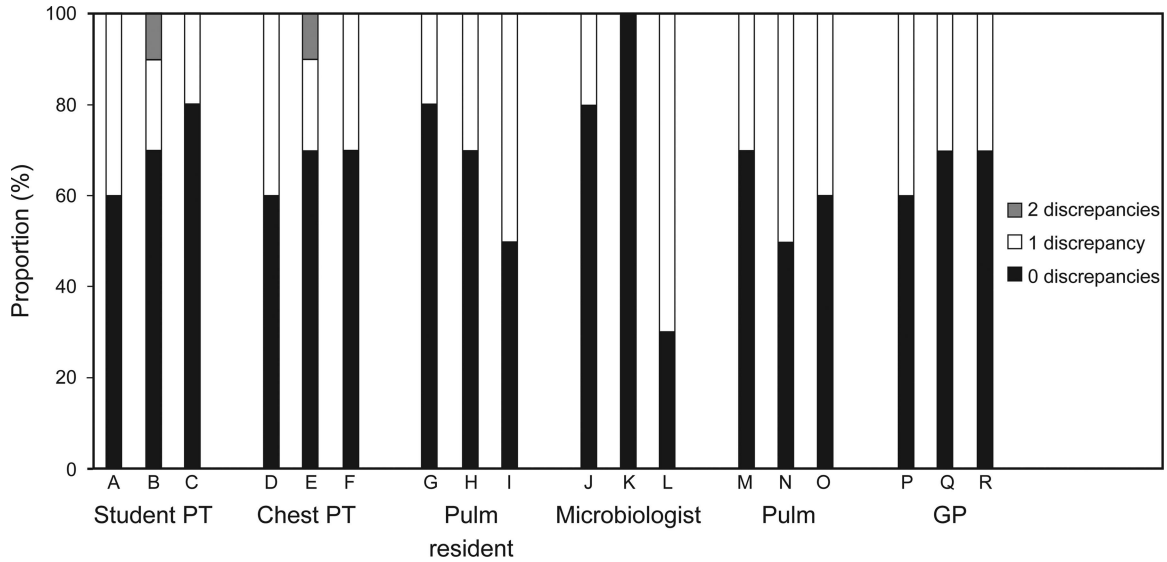

Fig. 2. Intra-rater reproducibility. Percentage by number of discrepancies between the 3 secretion readings of each evaluator from the 6 categories. Each letter represents an individual caregiver. PT = physiotherapist; Pulm = pulmonologist; GP = general practitioner.

for the greatest difference between 3 measurements of the same evaluator or the same health caregiver category and the mean of the 3 measurements. A Kruskal-Wallis test was performed to compare the results between the different categories of health caregivers for the first evaluation.

Results

Bacteriology and Subjects

Subjects were between 18 and 67 y old. They had various respiratory diseases (cystic fibrosis [$n = 3$] and COPD [$n = 7$]), and all were hospitalized for exacerbation.

Samples (Fig. 1) were characterized as follows: a normal flora for 6 samples, *Pseudomonas aeruginosa* in 2 samples, and *Staphylococcus aureus* in 2 samples.

Intra-Rater Reproducibility for the 6 Categories of Health Caregivers

The discrepancies, the reliability, and the agreement between the 3 readings of each evaluator from the 6 categories are presented in Figure 2 and Table 1. Only 20 of 54 Cohen's kappa coefficients were <0.60. We observed an absence of discrepancy from 30–100% of the secretions for all evaluators, independent of the category of caregivers. The proportion of the maximal discrepancy (=2) between the 3 readings was very low or null for all of the categories. Only one

RESPIRATORY CARE Paper in Press. Published on March 29, 2016 as DOI: 10.4187/respcare.04547

Table 1. Intra-Rater Reproducibility: Cohen's Kappa and Bland-Altman Between the 3 Sputum Readings of Each Evaluator From the 6 Categories

Microbiologist	Evaluator	Cohen's Kappa Coefficients			Bias, Mean	Limits of Agreement	
		R1 vs R2	R1 vs R3	R2 vs R3		Lower	Upper
Student PT	A	0.00	0.00	0.21	0.40	−0.61	1.41
	B	0.55	0.55	0.85	0.00	−1.60	1.60
	C	0.41	0.62	0.74	0.20	−0.63	1.03
Chest PT	D	0.61	0.29	0.55	0.40	−0.61	1.41
	E	0.60	0.64	0.62	0.00	−1.60	1.60
	F	0.58	0.35	0.78	−0.10	−1.21	1.01
Pulm resident	G	0.52	0.84	0.67	−0.10	−1.21	1.01
	H	0.21	0.35	0.55	0.00	−1.31	1.31
	I	−0.32	−0.18	0.62	0.40	−0.61	1.41
Microbiologist	J	0.58	0.80	0.80	0.40	−0.61	1.41
	K	1.00	1.00	1.00	0.00	0.00	0.00
	L	0.27	0.38	0.22	−0.30	−1.91	1.31
Pulmonologist	M	1.00	0.40	0.40	0.40	−0.61	1.41
	N	0.00	0.00	0.09	0.50	−0.53	1.53
	O	0.35	−0.32	0.55	0.00	−1.31	1.31
GP	P	0.50	0.68	0.50	0.40	−0.61	1.41
	Q	0.74	0.52	0.21	−0.10	−1.21	1.01
	R	1.00	0.54	0.54	0.10	−1.01	1.21

R1–R3 = readings 1–3
Student PT = student in last year of physiotherapy
PT = chest physiotherapist
Pulm resident = junior resident in pulmonology
GP = general practitioner

Table 2. Inter-Rater Reproducibility: Cohen's Kappa and Bland-Altman Between the 3 Evaluators From the 6 Categories for the First Reading of the 10 Secretions

Category	Cohen's Kappa Coefficients			Bias, Mean	Limits of Agreement	
	E1 vs E2	E1 vs E3	E2 vs E3		Lower	Upper
Student PT	0.00	0.00	0.04	1.10	−0.62	2.82
Chest PT	0.80	0.62	0.40	−0.40	−1.41	0.61
Pulm resident	0.00	0.17	0.05	0.40	−0.61	1.41
Microbiologist	0.50	0.06	0.06	0.40	−0.61	1.41
Pulmonologist	0.00	−0.09	0.00	−0.20	−1.75	1.35
GP	0.38	0.68	0.22	−0.50	−2.40	1.40

E1–E3 = evaluators 1–3
Student PT = student in last year of physiotherapy
PT = chest physiotherapist
Pulm resident = junior resident in pulmonology
GP = general practitioner

evaluator among the students in physiotherapy and the physiotherapists had 2-point discrepancies for one secretion.

Inter-Rater Reproducibility for the 6 Categories of Health Caregivers

The reliability and the agreement between the 18 evaluators from the 6 categories are presented in Table 2. Comparison of the first reading of the 10 sputum samples of each evaluator showed a great proportion of important discrepancies (=2) in 2 categories (40 and 50% for students in the last year of physiotherapy and general practitioners, respectively) (Fig. 3). Only 3 of 18 Cohen's kappa coefficients were >0.60, and 2 of them were observed between 2 chest physiotherapists. All but the pulmonologists and the senior chest physiotherapists showed 2 dis-

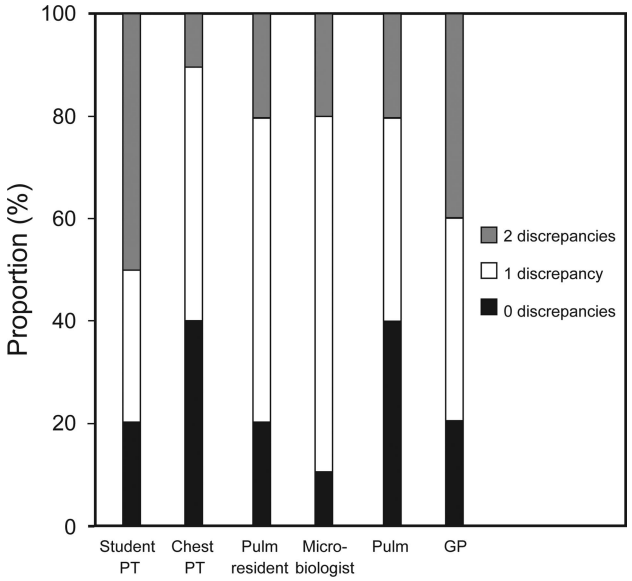

Fig. 3. Overall inter-rater reproducibility. PT = physiotherapist; Pulm = pulmonologist; GP = general practitioner.

crepancies for >20% of the secretions. The proportion of non-discrepancy between evaluators was very low and varied between 10 and 40%, depending on the category. The comparison between the categories showed a significant difference in the sputum readings ($P = .02$).

Discussion

We evaluated the intra- and inter-rater reliability and the agreement of the Sputum Color Chart for 3 evaluators from 6 categories of caregivers who can be routinely faced with sputum evaluation in their clinical practice. Our study highlighted a non-uniform reliability and agreement with the Sputum Color Chart. This observation was not related to the category of caregivers.

Sputum color is considered to be a key element in clinical evaluation of subjects with respiratory diseases. Indeed, sputum color appears to be regularly used in routine practice and in research.[14,15] Changes in sputum are included in the widely accepted Anthonisen definition of exacerbations in COPD[7] and in their evaluation.[16] In a large study on COPD exacerbations, sputum color was used as a marker of exacerbation, and the authors demonstrated that sputum purulence was strongly associated with bacterial growth.[17] Although this criterion has lower sensitivity and more limited applicability in children than in adults due to the difficulty of inducing children to expectorate,[18] it is also mentioned in the literature regarding exacerbations in pediatric subjects. Indeed, sputum color is one of the clinical criteria that improves specificity of the non-cystic fibrosis bronchiectasis exacerbation definition in children.[19] Moreover, Goeminne et al[20] used the Sputum Color Chart to demonstrate a relationship between sputum purulence from non-cystic fibrosis bronchiectasis subjects and inflammatory markers in sputum, history of *P. aeruginosa*, any form of colonization, or modified Brody scores. They defined the Sputum Color Chart as a tool that may predict the degree of inflammation, gelatinolytic activity, and disease severity in these subjects.[20] When children with cystic fibrosis are interviewed about symptom improvement after exacerbations, they report, among other improvements, a sputum color change.[21]

As self-assessment of sputum color by patients was demonstrated to be unreliable,[10] so there is a need for validated tools. Although the Sputum Color Chart previously demonstrated a good inter-rater reliability between doctors and subjects,[5] no study had previously investigated the intra-rater reliability for different health caregivers and the inter-rater reliability between different health professionals. These properties are important, since they are major factors in the therapeutic decision,[9] notably the prescription of antibiotics.[7] Indeed, in a study, it appeared that subjects with COPD experiencing an exacerbation with purulent sputum were treated with antibiotics, and those with mucoid sputum were not. The evaluation of sputum purulence was based on sputum color evaluation with a 9-point color chart. The authors concluded that sputum color assessment could contribute to avoiding unnecessary antibiotic therapy.[6]

We observed poor global intra-rater reliability in our sample of raters. This means that a difference in sputum color was perceived by mistake by a lot of caregivers. Since a sputum color change implies a reason for the physician to prescribe antibiotics, we can speculate a non-justified prescription based on this parameter. However, when looking specifically by category of caregivers, better intra-rater reliability was observed for microbiologists and senior chest physiotherapists than for other categories of caregivers (13 of 18 Cohen's kappa coefficients of both of these categories were >0.41, which means a moderate to very good reliability). A very good reliability (Cohen's kappa >0.80) was found for microbiologists. This could be explained by the dedicated routine practice based on the evaluation of sputum for this category of caregivers. The clinical experience probably plays a role, since it was previously shown to be related to a better agreement in different kinds of evaluations,[22,23] even if it is not systematically verified for all tools.[24-26]

The intra-rater discrepancies for all of the evaluators are heterogeneous. Only one reading of sputum produced 2 discrepancies for one junior and one senior physiotherapist. Bias of agreement is null for 1 of 3 of the evaluators and always <0.5, which means less than one classification of discrepancy. Also, this is not related to the category of caregivers.

Poor inter-rater reliability was observed for all categories with the exception of senior chest physiotherapists.

The level of training could be an influential element on the inter-reliability. Its influence was previously demonstrated on auscultation reliability.[27] Indeed, the students in physiotherapy and in pulmonology showed poor inter-reliability, with Cohen's kappa coefficient <0.20. However, we found a similar inter-reliability for pulmonologists. This can be considered surprising due to the experience of this category of caregivers regarding sputum observation. Moreover, it is questioning because the antibiotic prescription is based partially on this observation. In clinical practice, when an antibiotic prescription is based only on change in color sputum, it could be useful to combine sputum color evaluation from a physician and a microbiologist.

Global poor intra- and inter-rater reliability on the Sputum Color Chart reflects an inherent challenge with difficult evaluation of the color of secretions. Raters' disagreements on sputum color explain the negative Cohen's kappa coefficients that were calculated for a series of evaluations. Negative Cohen's kappa coefficients result when agreement occurs less often than predicted by chance alone. This suggests genuine disagreement between raters or an underlying issue with the instrument itself.[28]

Some limitations regarding this study need to be addressed. First, the sputum samples were presented as photographs, and this could influence the interpretation of the sputum color compared with true secretions. However, the Sputum Color Chart also uses photographs providing accurate representation of the 3 major grades of color, and the computer screen was always the same to avoid color modification due to the screen. Second, there was no control in the presentation of sputum. Third, generalizing our results to other raters should be performed with caution. Reliability could also differ according to other rater categories and other levels of skill. The individual level of training (as demonstrated previously for auscultation)[25,26] of our raters potentially influenced our results, and our randomized sample might not represent all raters with similar disciplines and training. But this means that sputum color evaluation is complex.

Conclusions

Even if the Sputum Color Chart is a useful tool for the clinician in the context of clinical deterioration, it presents a non-uniform reliability regarding the caregivers and their category.

REFERENCES

1. King PT, Holdsworth SR, Freezer NJ, Villanueva E, Holmes PW. Characterisation of the onset and presenting clinical features of adult bronchiectasis. Respir Med 2006;100(12):2183-2189.
2. Mullen JB, Wright JL, Wiggs BR, Paré PD, Hogg JC. Structure of central airways in current smokers and ex-smokers with and without mucus hypersecretion: relationship to lung function. Thorax 1987; 42(11):843-848.
3. Stănescu D, Sanna A, Veriter C, Kostianev S, Calcagni PG, Fabbri LM, Maestrelli P. Airways obstruction, chronic expectoration, and rapid decline of FEV1 in smokers are associated with increased levels of sputum neutrophils. Thorax 1996;51(3):267-271.
4. Lange P, Nyboe J, Appleyard M, Jensen G, Schnohr P. Relation of ventilatory impairment and of chronic mucus hypersecretion to mortality from obstructive lung disease and from all causes. Thorax 1990;45(8):579-585.
5. Murray MP, Pentland JL, Turnbull K, MacQuarrie S, Hill AT. Sputum colour: a useful clinical tool in non-cystic fibrosis bronchiectasis. Eur Respir J 2009;34(2):361-364.
6. Stockley RA, O'Brien C, Pye A, Hill SL. Relationship of sputum color to nature and outpatient management of acute exacerbations of COPD. Chest 2000;117(6):1638-1645.
7. Anthonisen NR, Manfreda J, Warren CP, Hershfield ES, Harding GK, Nelson NA. Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease. Ann Intern Med 1987;106(2):196-204.
8. Brusse-Keizer MG, Grotenhuis AJ, Kerstjens HA, Telgen MC, van der Palen J, Hendrix MG, van der Valk PD. Relation of sputum colour to bacterial load in acute exacerbations of COPD. Respir Med 2009;103(4):601-606.
9. Soler-Cataluña JJ, Alcázar-Navarrete B, Miravitlles M. The concept of control of COPD in clinical practice. Int J Chron Obstruct Pulmon Dis 2014;9:1397-1405.
10. Daniels JM, de Graaff CS, Vlaspolder F, Snijders D, Jansen HM, Boersma WG. Sputum colour reported by patients is not a reliable marker of the presence of bacteria in acute exacerbations of chronic obstructive pulmonary disease. Clin Microbiol Infect 2010;16(6):583-588.
11. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007; 60(1):34-42.
12. Soeken KL, Prescott PA. Issues in the use of kappa to estimate reliability. Med Care 1986;24(8):733-741.
13. Altman DG. Practical statistics for medical research. London: Chapman and Hall; 1991;179-276.
14. Diego AD, Milara J, Martinez-Moragón E, Palop M, León M, Cortijo J. Effects of long-term azithromycin therapy on airway oxidative stress markers in non-cystic fibrosis bronchiectasis. Respirology 2013; 18(7):1056-1062.
15. Llor C, Moragas A, Miravitlles M., ESAB study. Usefulness of a patient symptom diary card in the monitoring of exacerbations of chronic bronchitis and chronic obstructive pulmonary disease. Int J Clin Pract 2012;66(7):711-717.
16. Murray MP, Turnbull K, Macquarrie S, Hill AT. Assessing response to treatment of exacerbations of bronchiectasis in adults. Eur Respir J 2009;33(2):312-318.
17. Allegra L, Blasi F, Diano P, Cosentini R, Tarsia P, Confalonieri M, et al. Sputum color as a marker of acute bacterial exacerbations of chronic obstructive pulmonary disease. Respir Med 2005;99(6):742-747.
18. Kapur N, Masters IB, Chang AB. Exacerbations in noncystic fibrosis bronchiectasis: clinical features and investigations. Respir Med 2009; 103(11):1681-1687.
19. Kapur N, Masters IB, Morris PS, Galligan J, Ware R, Chang AB. Defining pulmonary exacerbation in children with non-cystic fibrosis bronchiectasis. Pediatr Pulmonol 2012;47(1):68-75.
20. Goeminne PC, Vandooren J, Moelants EA, Decraene A, Rabaey E, Pauwels A, et al. The Sputum Colour Chart as a predictor of lung inflammation, proteolysis and damage in non-cystic fibrosis bronchiectasis: a case-control analysis. Respirology 2014;19(2):203-210.
21. Abbott J, Holt A, Morton AM, Hart A, Milne G, Wolfe SP, Conway SP. Patient indicators of a pulmonary exacerbation: preliminary re-

ports from school aged children map onto those of adults. J Cyst Fibros 2012;11(3):180-186.
22. Hermansson LM, Bodin L, Eliasson AC. Intra- and inter-rater reliability of the assessment of capacity for myoelectric control. J Rehabil Med 2006;38(2):118-123.
23. Whatman C, Hing W, Hume P. Physiotherapist agreement when visually rating movement quality during lower extremity functional screening tests. Phys Ther Sport 2012;13(2):87-96.
24. Miles A, Huckabee ML. Intra- and inter-rater reliability for judgement of cough following citric acid inhalation. Int J Speech Lang Pathol 2013;15(2):209-215.
25. Allingame S, Williams T, Jenkins S, Tucker B. Accuracy and reliability of physiotherapists in the interpretation of tape-recorded lung sounds. Aust J Physiother 1995;41(3):179-184.
26. Morrow B, Angus L, Greenhough D, Hansen A, Olivier O, Shillington L, et al. The reliability of identifying bronchial breathing by auscultation. Int J Ther Rehabil 2010;17(2):69-75.
27. Brooks D, Thomas J. Interrater reliability of auscultation of breath sounds among physical therapists. Phys Ther 1995;75(12):1082-1088.
28. Juurlink DN, Detsky AS. Kappa statistic. CMAJ 2005;173(1):16.