UiT The Arctic University of Norway

Faculty of Health sciences, Department of Community Medicine.

Identification and prevalence of adventitious lung sounds in a general adult population

Juan Carlos Aviles Solis

A dissertation for the degree of Philosophiae Doctor - November 2019

Table of Contents

List of Figures

List of Tables

Foreword

Immediately after I finished my medical training at the University Hospital in Monterrey, Mexico I decided to join a project where I worked a whole year as a general practitioner in a remote area near the border between Mexico and Guatemala. No phone, no internet and the nearest hospital eight hours by road (when it was even open). There had never been a clinic before in this little town and the resources were scarce. Most of the medicine I learned in the university hospital, was of no use there, since we had no access to practically any diagnostic tools. But I knew well my semiology, I had my hands and I had my stethoscope. I got surprisingly far with these tools.

However, when I was in need to make more subtle clinical differences with the stethoscope I noticed that in my thick book of internal medicine there was not very good information about how much I could or should trust my findings. Most evidence presented was related to the newest diagnostic techniques, something that the patients I served would probably never have access to. I thought it would have been usefull to have better evidence on the tools I had available. It is probably the case that only a small amount of the total population in our planet has access to “state of the art” medical attention. Research and good evidence is also needed for the doctors treating people with more modest resources. This is why I found this project so interesting and decided to apply for the PhD position.

The “modest medicine” is not only a problem of developing countries. Resources are finite everywhere and even in the most affluent countries (like the case of Scandinavia) it is not a realistic scenario to use all the available technology for every single case. People in Norway know this very well and have decided to operate its health system by the principle of “Lowest Effective Level of Care” (LEON by its initials in Norwegian) and made the primary care physicians the corner stone of it. Norway has today one of the most effective health care systems in the world. If we want to achieve more efficient healthcare systems we need to make the best of the resources that already are in place. Making research of “old” and widely available diagnostic tools is just as important as investing in new ones.

I hope this work can contribute at least a little with good quality evidence for a diagnostic tool that is cheap, easy to use and is available virtually everywhere you can find a doctor.

Ackowledgements

I would like to express my gratitude to my supervisor Dr. Hasse Melbye. He was very supportive and patient throughout my period as a PhD candidate. He always went the extra mile in order to try to help me achieve my goal.

I would like to thank my second supervisor Peder Halvorsen for his comments in this work. They were very helpful and contributed for this work to be better.

Special thanks to Dr. Hans Pasterkamp and Dr. Cristina Jácome for the good teamwork with in the long process of classifying all the recordings. I would like to also thank Dr. Sophie Vanbelle for providing guidance and support in measuring agreement. Her insight and the methods she developed were crucial to complete this thesis.

A big thank you to all the staff and participants at the Tromsø Study. Without the Tromsø study such a large survey in lung sounds with so many quality datapoints would have been pretty much impossible.

I would like to thank my co-workers at the general practice milieu in Tromsø for creating a very nurturing environment that facilitated the flow of ideas and discussion during my time there.

I would like to thank my partner Ida and my little daughter Frida. Even though changing diapers and playing in the living room were not directly part of this thesis, it helped me to get through it.

English Summary

The stethoscope is a well-known diagnostic tool and a symbol of medicine itself. Despite its large popularity, the current evidence to justify its use is not very strong. The aim of this work is to describe the prevalence of adventitious lung sounds (wheezes and crackles) in a general population. In addition, we explored how different sources of variation such as inter-observer agreement and different breathing modes affected the prevalence of lung sounds and the reliability of its classification.

Therefore, we obtained lung sound recordings from 4033 participants in the 7th survey of the Tromsø at six locations in the thorax. In addition, the participants performed a spirometry and completed questionnaires regarding symptoms and self-reported disease. We observed a crude prevalence of adventitious lung sounds in 28% of the participants; 18 % had wheezes, 13% had crackles. We observed that age, female sex, self-reported asthma, and current smoking predicted the occurrence of expiratory wheezes. In the case of inspiratory crackles, significant predictors were age, current smoking, rheumatoid arthritis mMRC ≥2, low oxygen saturation and FEV1 Z-score. Age was the strongest predictor of crackles. Neither the presence of wheezes nor crackles were associated with the presence of airway infections in the week before the examination.

We also explored the interobserver agreement of seven groups of 28 doctors from different countries classifying 120 lung sound recordings. The probability of agreement for crackles varied between 65% and 87% in the different groups of observers. Congers kappa ranged from 0.20 to 0.58 and four of seven groups reached a k ≥0.49. For wheezes, we observed a probability of agreement between 69% and 100% and kappa values from 0.09 to 0.97. Four out of seven groups reached a k≥0.62.

It was in our interest to explore if the use of spectrograms could improve the classification of lung sounds. We conducted a study where we asked 23 medical students to classify the same lung sounds with and without spectrograms. Fleiss kappa values for the multirater agreement were k=0.51 and k=0.56 (p=.63) for wheezes without and with spectrogram, respectively. For crackles, we observed k=0.22 and k=0.40 (p=<0.01) in the same order. Compared to an expert panel's classification, 13 out of 23 students had a positive change in kappa when classifying wheezes (one with p<.05), and 16 out of

23 (two with p<.05) when classifying crackles. All the statistically significant changes were in the direction of improved kappa values (0.52 - 0.75).

In addition, we tested the possibility for variation in the prevalence of adventitious lung sounds in a subsample of 116 participants in the Tromsø Study breathing at spontaneous airflow velocity vs standardized airflow velocity at 1.5 L/s. We found that expiratory wheezes were present in 18 (16%) participants at spontaneous breathing and in 23 (20%) at standardized breathing. Inspiratory crackles were present in 19 participants at spontaneous breathing (16%) and in 18 (16%) at standardized breathing. The prevalence was not significantly different between the two methods. However, only nine participants in the case of wheezes, and five participants in the case of crackles were detected by both methods. The agreement of the two methods was k= 0.32 for expiratory wheezes and k=0.13 for inspiratory crackles.

Norsk sammendrag

Stetoskopet er et utbredt diagnostisk verktøy og et viktig symbol for legeyrket. Til tross for å være veldig populært mangler vi sterk evidens av dets diagnostiskevne til å rettferdiggjøre bruken av stetoskopet i et moderne legekontor. Hovedmålet med denne doktorgradsavhandlingen var å beskrive forekomsten av unormale lungelyder (piping og knatring) i en alminnelig befolkning. Vi undersøkte i tillegg hvordan forskjellige variasjonskilder kunne påvirke prevalens av lungelyder og reliabiliteten av lungelyd klassifikasjonen.

Vi samlet opptak av lungelyder fra 4033 deltakere i den syvende Tromsøundersøkelsen. Lungelydene ble tatt opp fra 6 forskjellige steder på brystkassen. Deltagerene fullførte en spirometri og svarte på spørsmål om symptomer og sykdom. Vi observerte unormale lungelyder hos 28% av deltagerene ; 18 % hadde piping, 13% hadde knatring. De variablene som best predikerte ekspiratorisk piping var alder, kjønn, astma og nåværende røyking. Når det gjaldt knatring var alder, nåværende røyking, reumatoid artritt, dyspné, lav oksygen metning og lav FEV1 Z-score signifikante prediktorer. Alder var den sterkeste prediktoren for forekomst av knatring. Det var ingen assosiasjon mellom piping eller knatring og symptomer på nedre luftveisinfeksjoner.

Vi undersøkte også interobserver enighet mellom syv forskjellige grupper med fire leger fra forskjellige land. De klassifiserte 120 lungelydopptak. Enigheten varierte fra 65% og 87% mellom de forskjellige gruppene. Vi observerte Congers kappa mellom 0.20 og 0.58. Fire av syv grupper oppnådde $k \geq 0.49$ når legene klassifiserte knatring. For piping varierte enigheten mellom 69% og 100% og kappa fra 0.09 til 0.97. Fire av syv grupper oppnådde $k \geq 0.62$.

Vi var også interessert i å finne ut om bruken av spektrogram kunne forbedre enigheten av lungelyd klassifikasjonen. Vi gjennomførte en studie der vi spurte 23 medisinstudenter om å klassifisere de samme lungelydene to ganger, en med og en uten spektrogram. Fleiss kappa for piping var $k=0.51$ med spektrogram og $k=0.56$ uten ($p=.63$), og for knatring henholdsvis $k=0.22$ og $k=0.40$ ($p=<0.01$) Sammenlignet med fasit hadde 13 av 23 studenter en positiv endring i kappa når de klassifiserte piping

(en med p<.05), og 16 av 23 (to med p<.05) når de klassifiserte knatring. Alle de statistiske signifikante endringer var retning av økt enighet med fasit (kappa

Vi testet i tillegg om forskjellige pustemønster endret prevalensen av lungelyder hos 116 deltakere fra Tromsøundersøkelsen. Vi observerte ekspiratorisk piping hos 18 (16%) deltakere med spontant pustemønster (tilfeldig lufthastighet) og 23 (20%) som pustet med en standardisert lufthastighet på 1.5 L/s. Inspiratorisk knatring var til stede hos 19 (16%) deltakere med spontant pustemønster og 18 (16%) som pustet med standard lufthastighet. Prevalensen var ikke forskjellig mellom de to metodene, og det ble med begge metoder kun oppdaget ni deltakere med piping og fem med knatring. Enighet mellom de to metodene var k= 0.32 for ekspiratorisk piping og k=0.13 for inspiratorisk knatring.

List of papers

Paper I

Juan Carlos Aviles-Solis, Sophie Vanbelle , Peder Halvorsen, Nick Francis, Jochen W L Cals, Elena A Andreeva, Alda Marques, Päivi Piirilä, Hans Pasterkamp and Hasse Melbye. International perception of lung sounds: a comparison of classification across some European borders. BMJ Open Respiratory Research 2017;4:e000250. doi: 10.1136/bmjresp-2017-000250

Paper II

Juan Carlos Aviles-Solis, Ingrid Storvoll, Sophie Vanbelle and Hasse Melbye. Impact of spectrograms on the classification of wheezes and crackles. Submitted.

Paper III

Cristina Jácome, Juan Carlos Aviles-Solis, Åshild Uhre, Hans Pasterkamp and Hasse Melbye. Adventitious and Normal Lung Sounds in the General Population: Comparison of Standardized and Spontaneous Breathing. Respiratory Care Nov 2018, 63 (11) 1379-387; DOI: 10.4187/respcare.06121

Paper IV

Juan Carlos Aviles-Solis, Cristina Jácome, Anne Davidsen, Raimonda Einarsen, Sophie Vanbelle, Hans Pasterkamp and Hasse Melbye. Prevalence and clinical associations of wheezes and crackles in the general population. The Tromsø Study. BMC Pulm Med. 2019 Sep 11;19(1):173. doi: 10.1186/s12890-019-0928-1.

Abreviations

ACC – American College of Cardiology

AHA – American Heart Association

AdLS - Adventitious Lung Sounds

BMI - Body Mass Index

CI – Confidence Interval

COPD - Chronic Obstructive Lung Disease

FEV_1 – Forced expiratory volume in one second

GLI – Global Lung Function Initiative

HF – Heart Failure

Hz - Hertz

ILD – Interstitial Lung Disease

IPF – Idiopathic Pulmonary Fibrosis

k – Kappa Coefficient

LVEF – Left Ventricular Ejection Fraction

LLN – Lower Limit of Normal

mMRC – Modified Medical Research Questionnaire

ms – Milliseconds

OR – Odds ratio

PEFR – Peak Expiratory Flow Rate

RA-Rheumatoid arthritis

SpO_2 – Peripheral capillary oxygen saturation,

UiT - UiT, The Arctic University of Norway

1 Introduction

1.1 A brief historical note on auscultation and the stethoscope

Diagnosis is the investigation or analysis of the cause or nature of a condition, situation, or problem.(1) When applying this concept into medicine, we can define diagnosis as the art or act of identifying a disease from its signs and symptoms.(1) Through time, medicine has made an effort to characterize diseases and document their unique features in order to be able to distinguish healthy from diseased and diseases from one another. This, with the main goal to provide appropriate treatment and relief. Some of these characteristics can be identified by plain sight, or during interrogation. However, some other might be elusive and tools are needed in order to observe them.

The use of diagnostic tools has changed according to the understanding of the physiological mechanisms used to explain a condition and the epidemiology of the diseases at the time. In the 19th century, pulmonary diseases such as pneumonia and tuberculosis were among the main causes of death.(2) Doctors at this time put a lot of effort into improving their ability to diagnose these diseases and developed techniques of auscultation and percussion of the chest.(3)

Back in the 18th century, auscultation occurred in a direct manner where the doctor had to push his ear to the chest of the patient. This situation was distressing for a young Parisian doctor named René Laënnec. One day during an encounter with a female overweight patient, he found the task of chest auscultation difficult. He pushed his ear against the naked and voluptuous bosom of his patient, but this solution was not effective and made him extremely uncomfortable. To solve this problem and inspired by some children playing in the park, he rolled some pieces of paper

Figure 1.- "The Examination" by Jules-Abel Faivre, 1898. Wellcome Collection, London. CC by 4.0.

to a cylinder shape and with his ear on one extreme, and the chest of the patient on the other, he managed to satisfactorily auscultate the patient.[4] The trick worked so well that he later developed a wooden cylinder to auscultate his patients. He called his invention stethoscope from “Stethos” which means chest, and “skopein” which means to see.

Figure 2.- Laennec’s stethoscope.
By Science Museum London / Science and Society Picture Library, CC BY-SA 2.0.

By reading this curious anecdote it might be easy to imply that the main achievement of Laënnec was the invention of a wooden cylinder to auscultate the chest. However, his work with auscultation was a lot more comprehensive. He invested a large amount of time comparing the auscultation findings with pathological observations in the lungs of patients who had died of respiratory disease.[4] He observed a clear association between diseased lungs and the presence of auscultation findings. Through this work, he gave the stethoscope a fact-based ability to distinguish disease from non- diseased. The ability to diagnose. These observations are probably the reason why he called his invention stethoscope. Because the sounds he heard gave him the ability to “see” the changes he had observed in the lungs at the pathology laboratory.

Figure 3.- Canman Stethoscope. Science Museum, London. Wellcome collection, London. CC by 4.0

The stethoscope has changed since its invention. In 1853, Dr. George P. Cammann used rubber tubes and earpieces to substitute the wooden cylinder.[5] In 1925 Bowles and Sprague developed a headpiece with a bell and a diaphragm to select different frequency ranges.[6] In 1970’s 3M developed a tunable diaphragm for the same purpose.[7] Nowadays, it is possible to find electronic versions of the stethoscope. These devices are very diverse and can perform many different functions depending on the model. Most of them are able to amplify, record, visualize, store and share lung sounds. These devices together with the increased availability of processing power in the form of personal computers and mobile telephones open the possibility for new types of analysis of lung sounds. Nonetheless, to present day the clinical principle is still the same: To look for the presence of adventitious lung sounds. Both electronic and analogue stethoscopes seem to perform equally good for this purpose.(8)

However, the epidemiology of mortal diseases has changed since the invention of the stethoscope. People do not die from tuberculosis and pneumonia as often as before. In addition, new diagnoses associated with adventitious lung sounds not identified at Laennec's time have emerged. For example Chronic Obstructive Pulmonary Disease (COPD), pulmonary fibrosis, interstitial lung disease, etc. Thus, the observations made by Laennec are incomplete in a modern medical context. In addition, there are diagnostic tools that literally let us see ("skopein") into the lungs. For example, x-rays, CT scans, MRI's and ultrasound.

With limited concrete evidence about the utility of the stethoscope in daily clinical practice, and many new and sophisticated competitors it is natural to question whether the stethoscope is a helpful tool in the diagnosis of chest disease and to what extent the stethoscope still has a place in modern medicine.

1.2 Use of lung auscultation in contemporary clinical practice

The stethoscope is one of the most available medical tools even in developing countries.(9) Medicine schools teach auscultation, and every book on physical examination includes a chapter about this subject.

However, 200 years after its invention there is controversy among clinicians about how useful this tool really is. Some people see it as an old ritual with gaps in clinical evidence that resists disappearing under the shadow of newer and more accurate technology.(10, 11) Others still see value in its simplicity, cost, availability and ease of use and training.(12, 13)

In spite of the controversy, the sales of stethoscopes increase with every year. The total value of the stethoscope market worldwide was USD 327.7 million in 2016, and is expected to expand at a constant annual growth rate of 4.7% over 2025.[14]

1.3 Adventitious lung sounds: terminology, classification and physiology.

Classification and terminology of lung sounds have been a source of debate over centuries,[15] and a well-known cause for variation in its classification thus hindering its reliability.[16] The international Lung Sounds Association has developed a classification of lung sounds,[17] which I will use in the following descriptions. These sounds can be broadly classified into a) Normal lung sounds and b) Adventitious lung sounds (AdLS). The latter are divided into i) continuous and ii) discontinuous sounds.

a) Normal lung sounds

This is a continuous and soft sound. It is present at most of inspiration and the beginning of expiration. The sound mainly arises from the turbulence in the central airways.[18] This sound is also referred to as vesicular sound due to a past and erroneous belief that the sound originated from the alveoli.[18, 19]

b) Adventitious Lung sounds

i) Continuous lung sounds

Continuous adventitious lung sounds are usually known as wheezes. This type of sound is relatively long (>100 milliseconds). It has a musical tone because of its sinusoidal waveform with a frequency and harmonics added to it.[20] Wheezes can be further classified according to its frequency into high frequency and low frequency wheezes. Wheezes are generated by the flutter of narrowed or collapsed airways that let flow bolus of air at a determinate frequency and are generated in the 2nd to 7th generation of bronchial branches.[21] Wheezes can appear in inspiration and/or expiration.

The term rhoncus is also used for low frequency wheeze.[17] Even though low frequency wheezes and rhoncus share some characteristics, it is suggested that the physiological mechanisms behind rhoncus are different than those for low frequency wheezes. The involvement of secretions in the airway is suggested to play a role in the generation of rhonchi.[16, 18, 20]

ii) *Discontinuous lung sounds*

Crackles are short (5-15 ms) explosive sounds that have been described in different terms like the sound obtained by shaking a container of moderately heated salt.[18] The classification suggested by the International Lung Sound Association makes a distinction between high frequency or fine crackles and low frequency and coarse crackles based on its frequency and amplitude of its waveform.[17] Fine crackles are also known as Velcro crackles.[22] However, these characteristics seem to be hard to recognize by clinicians and it may not be recommended to make this distinction for clinical purposes.[23]

Crackles occur because of the sudden opening of collapsed airways, causing an instant equalization of the air pressure from which sound originates.[24] Crackles can be present in inspiration and/or expiration.

iii) *Mixed sounds*

Sqwaks are sounds with both musical and non-musical components and it resembles the most to a short wheeze.[18] This sound is thought to occur by the oscillation of the walls in the peripheral airways at inspiration.[25]

c) *Respiratory sounds not originating from the lungs*

i) *Stridor*

This is a continuous high pitch sound with musical quality and of short duration.[26] It is present mostly at inspiration and is produced by the sudden closing of the epiglottis, trachea, the main bronchi or the vocal chords.

ii) *Pleural rub*

This is a discontinuous sound, caused by the friction between the visceral and parietal pleurae when inflammation is present.[18] It is described as the sound produced when rubbing two pieces of leather against each other. Its appearance coincides with both inspiration and expiration.[25]

1.4 Adventitious lung sounds and the diagnosis of lung diseases

Crackles and pneumonia

Crackles appear probably due to the reduction of the airways lumen and airway collapse caused by the congestion and engorgement of the lung tissue during the congestion phase in the pathophysiological process of pneumonia.[24, 27] The increase of bronchial secretion, which intermittently block the passage of air during inspiration, could also play a role in the presence of crackles.[19] Crackles are present in approximately 32% – 65% of the adult patients presenting to primary care with a confirmed diagnosis of pneumonia.[28-31] This sign has showed to be a significant predictor of pneumonia in different primary care based prediction models.[32] Although crackles have a moderate to low sensitivity and specificity on its own, the perceived diagnostic weight is high thus affecting the rate of antibiotic prescription and specialist referrals.[29, 33]

Crackles and interstitial lung disease

Different types of interstitial lung disease have been associated with the presence of crackles. Thickening of the distal airways and the substitution of the elastic tissue by fibrotic tissue may result in an increased collapsibility of the airways,[34] facilitating the appearance of crackles.[35] Idiopathic pulmonary fibrosis (IPF) is probably the best-known example where inspiratory fine crackles are almost considered diagnostic of this disease in absence of other causes.[36, 37] Asbestosis is another good example where crackles are considered an early sign of the disease.[38, 39] In a study using lung biopsy as gold standard, Epler et al found that 60% of the patients with changes related to interstitial lung disease had fine crackles on auscultation.[39] In a recent prospective case control study, Sgalla et al found that the presence of bilateral inspiratory velcro crackles had a strong correlation (OR 13.46, 95% CI 5.85–30.96, $p < 0.001$) with the presence of radiological patterns of Interstitial Lung disease.[34]

Crackles and airflow limitation

The presence of crackles has been found to be helpful in the diagnosis of airflow limitation but mostly in conjunction with other elements of the clinical examination.[40-42] Crackles can be present in patients with COPD due to the loss of elastic supportive tissues of the distal airways contributing to its collapse during expiration and suddenly reopen in inspiration.[43]

The prevalence of crackles in COPD is estimated at 15% using direct auscultation.[41] Prevalence as high as 71% have been reported using computerized analysis of lung sounds.[44] In one study, the presence of crackles alone had a sensitivity of 22% and a specificity of 96% for a diagnosis of COPD.[41] Other studies found that crackles do not contribute much to the prediction of airflow limitation in COPD.[45, 46]

However, the presence of crackles might be useful in the monitoring of COPD exacerbations. Jácome et al report that in patients with COPD the number of crackles increases during acute exacerbations compared to stable state using automated computer analysis.[47] It has been postulated that the presence of crackles precedes acute exacerbations also using computerized analysis.[48] Nonetheless, the evidence on this subject comes from small and highly selected samples and is thus inconclusive.

Wheezes and airflow limitation

Wheezes present in asthma due to the contraction of the smooth muscle thus narrowing the diameter of the peripheral airways which generates oscillation of the bronchial walls by rapid flow of gas.[25] The presence of wheezes is a clinical feature commonly associated with airways obstruction whether this is reversible as in the case of asthma,[20] or irreversible as in the case of COPD.[41]

The presence of wheezes is significantly associated with a lower peak expiratory flow rate (PEFR) and characteristics like loudness and high pitch were associated with even lower PEFR.[49] However, the relationship between the level of obstruction measured by PEFR and the presence of wheezes does not have a fixed threshold and is not very stable.[49]

The presence of wheezes relates to the probability of having airway obstruction measured by FEV1.[41, 46] However, changes as big as 35 % drop in FEV_1 are required in order for wheezes to appear in provocation tests.[20] Nonetheless, wheezes might be absent in severe obstruction.[50] These results might explain the low sensibility (around 15 %) of wheezes to predict airflow limitation.[45, 46] However wheezes have a high specificity for airway obstruction (99%).[45]

Asthma is associated with the presence of wheezes, intermittently.[20] It is a cardinal clinical sign for its diagnosis, and is actively used in the monitoring of adequate control of asthma.[51]

Wheezes has been estimated to be present in 21% of patients with COPD.[41] The severity of COPD is associated with the presence of wheezes. Oshaug et al reported that the frequency of wheezes on auscultation increases from about 12% in mild cases of COPD to almost 40% in severe cases of COPD.[52]

Other sounds and disease

The presence of sqwaks is associated with the presence of allergic pneumonitis and pneumonia.[53, 54] Stridor is associated with laringomalacia, tracheomalacia, bronchomalacia, croup, paralysis of the vocal chords, anaphylaxis or a mass obstructing the main airways.[18] Pleural rubs appear because of pleural inflammation. Pleural rubs can also be heard in the presence of malignant diseases of the pleura.[18]

1.5 Adventitious lung sounds and the diagnosis of heart diseases

The current guidelines for the diagnosis of heart failure (HF) state that clinical history and physical examination (including auscultation) make the corner stone in the initial diagnosis of HF.[55] A survey among European primary care physicians showed that general practitioners considered the presence of crackles as an important sign of HF.[56]

In heart failure, crackles originate when there is an equalization of gas pressure during the opening of distal airways, narrowed by peribronchial edema.[25] Crackles occur more often at the bases of the lungs and during late inspiration. The prevalence of crackles during stable state HF is estimated at 24% among patients with systolic heart failure and 18% in patients with heart failure with normal ejection fraction. (57) However, the frequency of crackles increases along with the severity of HF and they are associated with significant dysfunction of the left ventricle.[58, 59] A study done by Pfitzenmeyer et al showed a high prevalence of crackles (77%) in patients with the most advanced stage of heart failure (Stage D ACC/AHA).[60] Observations from two studies including patients with an aggravated disease course, (emergency rooms incomings), reported a sensitivity of crackles for heart failure of 69% and 93%, respectively.[61, 62]

Beyond the initial diagnostic setting, crackles could also play an important role in prognosis. The presence of crackles and the extent to which they are present relates to mortality independently of LVEF values.[57, 63] Crackles also correlate well to 60 days hospital readmission, in patients with HF as main diagnosis at discharge.[64]

1.6 Limitations and challenges of the use of adventitious lung sounds in clinical practice

In spite of its obvious advantages of low cost, easy to use and nearly universal availability, the stethoscope has shortcomings as a diagnostic tool in the clinic.

The appearance of neither wheezes nor crackles is pathognomonic of any particular disease of physiological process. As exposed in the past sections, many different diseases are associated with AdLS. In addition, wheezes and crackles can also be present in healthy individuals.[44, 65-67] I have previously described that AdLS have moderate diagnostic sensitivities at its best, making it less than an ideal tool for screening purposes.

Another significant problem is that there are significant sources of variation, which can be present in the process of recollecting, interpreting and classifying lung sounds.

The use of different terminology is an important problem. Even though different health professionals may agree that they hear the same sound phenomenon, they might have different names for it.[15] The situation gets more complicated when different cultures and languages are involved.[16]

In addition, the interpretation of sound is a subjective task. Considerable inter-observer variation has been reported in several studies.[30, 46, 68-74] This variation is not experience dependent.[72] With new technological developments and increased processing power of computers and portable devices it is possible to apply new methods to analyze sound signals. One of these methods is the use of spectrograms of the sound.[75] The spectrograms are a visual representation of sounds where it shows time in the X axis, frequency in the y axis and the intensity is represented by color. Spectrograms are calculated with the help of a mathematical procedure (Fast Fourier Transformation).[76] When analyzing spectrograms of lung sounds recordings we can observe that wheezes and crackles have recognizable patterns (figure 7). Wheezes appear as long horizontal lines since they sustain a determinate frequency over a period of time while crackles show as vertical lines due to their short duration. Using an additional sensorial input to interpret AdLS could seem a reasonable solution to increase the reliability in the classification of AdLS. Andrés et al found that the spectrograms do have a positive impact on how medical students assign a diagnosis with the help of lung sounds.[13] However, the design of this study did not isolate the effect of spectrograms and therefore this affirmation is inconclusive. There are very few studies on this subject. Therefore, we wanted to find out whether the use of spectrograms would help to medical students to better classify lung sounds.

How auscultation is carried out may also play a role. One particular component that can potentially have a large inter-individual variation is the speed at which a patient inhales or exhales (Airflow velocity). Patients can modify this parameter *ad libitum* without doctors having any objective control of it. Airflow velocity changes can affect the rate of wheezes since the presence of wheezes is dependent on a critical airflow velocity.[49] Changes in lung volume could also have some effect in the presence of crackles. Some collapsed airways might not open at tidal or subtidal volumes, while they might do when increasing inspiratory or expiratory volumes. For the sake of pragmatism, doctors usually ask their patients to take a deep breath with an open mouth. However, a lot of the research about the physiology of lung sounds uses fixed airflows in laboratory environments prioritizing the repeatability of the experiments. This poses a double problem. First, the reliability of AdLS in a clinical environment might be compromised if the inter-intra individual variation is large enough with a consequently direct effect in the presence of wheezes and crackles. Second, the valuable research in lab conditions might not be applicable in a busy general practitioner office.

Even though General Practitioners might be positive to use the stethoscope in their daily routine, there is a lack of evidence about how useful lung auscultation actually is in this context. To date most of the studies have investigated AdLS in small and highly selected samples. Most of the existing studies do not take into account the presence of wheezes and crackles in healthy people, probably overestimating their specificity. To my knowledge there are no studies describing the prevalence of AdLS in the general population including how AdLS relate to measurements of lung function.

2 Aims of the thesis

The main objective of this thesis was to describe the occurrence of adventitious lung sounds (wheezes and crackles) in the general adult population and their relationship to self-reported disease, respiratory symptoms, and measurements of lung function such as spirometry and oxygen saturation. In addition, I wanted to explore the reliability of the method we used to classify lung sounds in terms of inter-observer agreement, impact of the use of spectrograms during classification of the recordings, and how the use of spontaneous breathing can influence the prevalence of AdLS compared to the use of standardized respiratory airflow.

3 Overview of the Papers

In study I we explored the variation of inter-observer agreement in a sample of 28 observers from different countries when classifying lung sound recordings from 20 subjects (n=120). These lung sound recordings were supported by visual representation of the sounds called spectrograms.

In study II we wanted to see if the use of spectrograms had an impact on the classification of lung sound recordings. We tested this hypothesis using a sample of 23 medical students classifying 30 lung sound recordings with and without the use of spectrograms.

A big amount of research in lung sounds is performed with subjects breathing at a fixed airflow velocity. This is not the case in clinical practice. In study III we tried to see if the use of spontaneous breathing airflow, as used in clinical praxis, had an impact in the prevalence of wheezes and crackles compared to a fixed airflow which is used for research purposes. We used lung sounds from 116 adults (40 year and older) at spontaneous breathing and fixed airflow velocity. The sounds were classified by four lung sound researchers aided by spectrograms.

In study IV we calculated the prevalence of adventitious lung sounds and their associations with self-reported disease, respiratory symptoms and measurements of lung function in a sample of 4033 adults from a population based study.

4 Material and methods

In this thesis, I will refer to the subjects from which lung sounds were recorded as “participants”, and to the subjects involved in the interpretation and classification of lung sounds as “observers”.

4.1.1 Participants

We selected our participants from two completely different settings. For study I and II the questions to be answered were in relationship to the reliability in the classification of AdLS. The main outcome was dependent on the observers and the participants only provided good enough material for classification. For this reason, we recorded lung sounds in adults attending a heart and lung rehabilitation program since they provided the best chance to find AdLS.

For study III, we wanted to explore how the airflow variations affected the prevalence of AdLS and took a small subsample of participants in study IV. Since the protocol we chose to record lung sounds in study IV did not control for airflow velocity it could potentially influence its conclusions. With that in mind we attempted to obtain a representative subsample of study IV.

In study IV we wanted to estimate a prevalence of AdLS in a general population, thus the question of representativeness was fundamental. A representative sample from an adult general population was important because the results would be more relevant to primary care providers, although less to clinicians working at hospitals where the patients have a longer or more complicated course of disease. A general sample would also provide us with a mixture of healthy and diseased patients that would allow us to compare the levels of the variables of interest in healthy and diseased patients with AdLS.

Study I and II

For study I, we recorded lung sounds from a convenience sample of 20 subjects 40 years or older. To recruit them, we held a presentation about lung sounds at a rehabilitation center for heart and lung related diseases (Lung cancer, COPD, heart failure, etc) in North Norway. At the end of the presentation, we asked if someone would be interested in participating in our study. Fourteen participants agreed to participate. The participants from the rehabilitation center were 67.4 years old in average (44-84) and nine of them were female. To hold a balanced sample (concerning prevalence of normal lung sounds), we obtained the rest of our recordings from six self-reported healthy employees at our university aged 51.8 years on average (46-67) and five were female. We collected age, gender and self-reported history of heart or lung disease. No personal information was registered that could link the sounds to the individual subjects. The research project was presented to the ethical committee in south east Norway who concluded that the Research Project was outside the remit of the Act on Medical and Health Research 2008. This was because the study would not generate new knowledge about health and disease, but rather information about the variation of the skills in auscultation of lung sounds.

In study II, we used a sample of the sounds recorded in study I. In order to reduce the influence of prevalence in the kappa coefficients we draw a sample of 30 sounds with an aimed prevalence of 50% of the recordings containing AdLS and 50% normal respiratory sounds. We used the expert classification of the recordings used in study I.

Study III and IV

For study III and IV we recruited participants from the seventh survey of the Tromsø Study (2015-2016). The Tromsø Study is a periodical epidemiological survey that has taken place regularly since 1974.[77] All the inhabitants 40 years and older living in the municipality of Tromsø received a postal invitation to participate (n= 32 591). From this group 13 304 individuals were preselected to participate in the second visit, 10 150 due to random selection and 3 154 specially invited due to previous participation in the study. There was an attendance of 65% (n=21 083) to the first visit. Among them, 9 253 received the invitation to the second visit and 90% (n= 8 346) attended. We recorded lung sounds in 72% of the participants at the second visit (n= 6 035). Due to limited time and resources to classify all the sound recordings our study sample only included participants attending the second visit in 2015, and the participants attending the second visit in 2016 who were also randomized to an echocardiographic examination (n= 4033). A complete overview of the participants included in study IV is presented in figure 3.

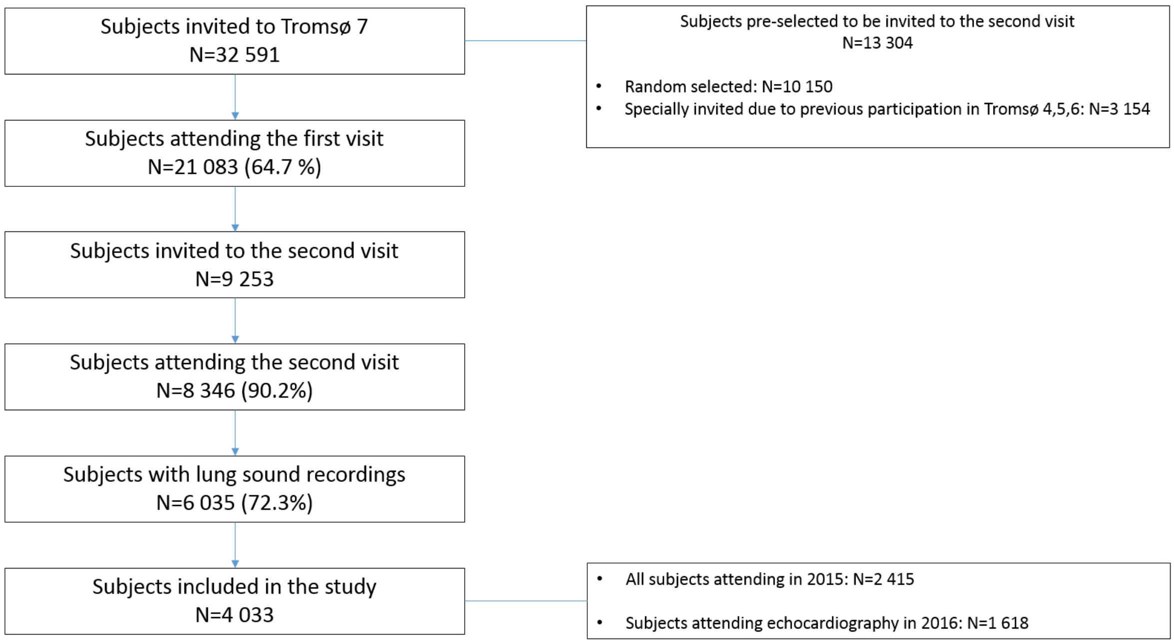


Figure 4.- Flow diagram of participants in study IV.

In study III, we used a random subsample of 116 participants from study IV. Recruitment of the participants happened consecutively in a four-week period in 2016 among those not randomized for echocardiography.

All the study participants provided written consent. The Regional Committee for Medical and Health Research Ethics in North Norway (REK) approved the study.

At the first visit, the participants filled in questionnaires about self-reported disease and life style, and blood samples were taken. The second visit consisted of a more detailed examination where among other things, the participants performed a spirometry and oxygen saturation was measured. Due to staff limitation about half the participants were randomized to an echocardiographic examination.

Age sex and weight

Height and weight were measured in light clothing and without shoes. Weight was rounded up to nearest 100 grams. Body mass index (BMI) was calculated (kg/m^2). The age reached up to December 31st 2015 was registered.

Smoking status

The participants answered the following question about smoking habits:

Do you/did you smoke daily? **(Never/ Yes, now/ Yes, previously)**

We divided the participants into never smokers, current smokers and past smokers according to their answer.

Self-reported disease

The participants completed a questionnaire in their first visit regarding the presence of disease. This questionnaire was electronic and the participants used tablets to answer. The questions we included in our dataset were the following:

Have you ever had, or do you have high blood pressure? ***(no, yes now/yes previously)***

Do you have, or have you had a heart attack? ***(yes/no)***

Have you ever had, or do you have heart failure? ***(no/yes now/yes previously)***

Have you ever had, or do you have atrial fibrillation? ***(no/yes now/yes previously)***

Have you ever had, or do you have chronic bronchitis/emphysema/COPD? ***(no/yes now/yes previously)***

Do you have, or have you had asthma? ***(no/yes now/yes previously)***

Have you ever had, or do you have rheumatoid arthritis? ***(no/yes now/yes previously)***

These variables were dichotomized and the answers “yes now” and “yes previously” were accounted as “yes”.

Symptoms of respiratory disease and dyspnea

Before performing spirometry the patients had to answer the following question about symptoms of airways disease present at the moment:

Have you had symptoms of common cold, bronchitis or other airway infection the last 7 days?

(Yes/No)

To explore for recent occurrence of dyspnea we asked this question:

How is your breathing today compared to normal?

(I breathe more easily/I breathe as normal/ I am more short of breath than normal)

This variable was dichotomized. One group included those participants who breathe as normal or more easily. The other group consisted of the participants who were shorter of breath than normal. We were interested in the group with an acute symptomatic worsening and the difference between the first two categories (easily and normal) was not in our interest.

Before spirometry, we applied the modified Medical Research Council dyspnea scale (mMRC). The answers were dichotomized into mMRC≥2 and mMRC≤1 since we considered this cut-off to be clinically significant.[51]

Spirometry

The participants performed spirometry with the use of SensorMedics Vmax 20c Encore (VIASYS Healthcare Respiratory Technologies, Yorba Linda, CA, USA). Spirometry was conducted according to the standards of the American Thoracic Society (ATS)/ European Respiratory Society (ERS). We used the reference values from the Global Lung Function Initiative (GLI).[78] Calibration was performed every morning. The participants were seated and wore a nose clip.

Pulse oximetry

We used a pulse oximeter Onyx II model 9550 (Nonin Medical, Inc., Plymouth, MN, USA). The highest value after three measurements was registered. $Sp0_2$ values <80% were discarded due to uncertain validity (n=2).

4.1.2 Recording of the sounds

To record lung sounds we used a microphone MKE 2-EW placed in the tube of a Littmann Master Classic II stethoscope (3M, Maplewood, MN, USA) at a distance of 10cm from the headpiece. In study I and II the microphone was connected to a digital sound Handy recorder H4n (Zoom, Tokyo, Japan) through a wireless system, EW 112-P G3-G (Sennheiser electronic, Wedemark, Germany).

In study III and IV the microphone was connected to the same wireless system. The wireless system transmitted the signal to an external sound card (Scarlett 2i2, Focusrite Audio Engineering Ltd., High Wycombe UK) which connected to the computer audio input. The computer used custom developed software to label the sounds (participant ID and recording site). The program also allowed us to start the recording with the help of a wireless control (R700, Logitech Europe S.A., Lausanne Switzerland). (figure 4)

The audio files were in '.wav' format and recorded at a sample rate of 44 100Hz and 16 bit depth in a single monophonic channel. We did not perform post-processing of the sound files or implement filters.

Figure 5.- Equipment used to perform lung sound recordings in study III and IV.

We placed the membrane of the stethoscope against the exposed thorax of the subjects. We asked the subjects to breathe deeply while keeping their mouth open (Spontaneous breathing). We started the recording with an inspiration and continued for approximately 15s trying to capture three full respiratory cycles with good quality sound. In study III, the recordings were shorter (10 seconds). In all the studies, the person recording sounds used a headphone as an audio monitor to evaluate the quality of the recording. When too much noise or cough was heard during the recording, a second attempt was performed. We repeated this procedure at six different locations (figure 5) These locations were selected to be similar to the usual auscultation exam used by doctors in clinical praxis and to reduce the amount of noise in the recordings.(79)

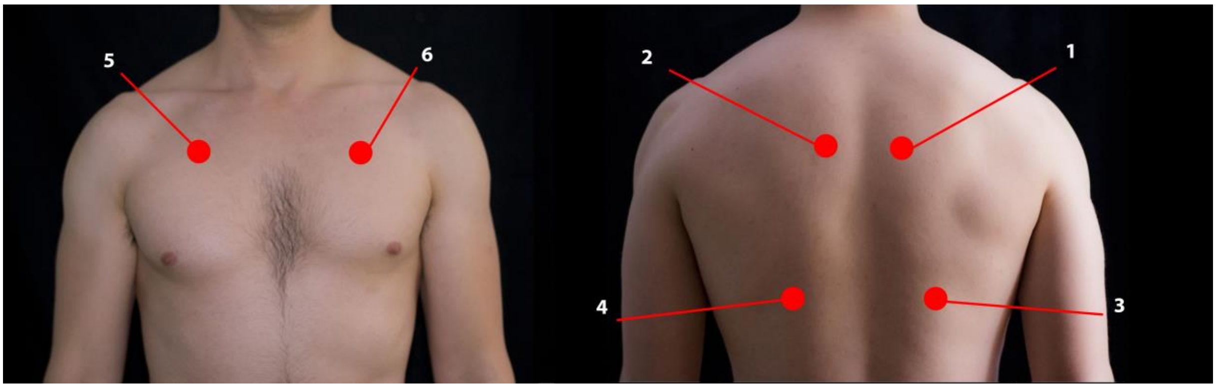


Figure 6.- Recording sites. (1 and 2) Between the spine and the medial border of the scapula at the level of T4–T5; (3 and 4) at the middle point between the spine and the mid-axillary line at the level of T9–T10; (5 and 6) at the intersection of the mid-clavicular line and second intercostal space.

In study III, we recorded lung sounds twice in the participants included. The first recordings were done according to the standard procedure previously described. The second set of recordings was performed while the participants had to inspire and expire at a target airflow of 1.5 l/s. In order to achieve this, the participants had to breathe through the mouth piece of a portable spirometer (Ndd Easy on-PC Spirometry System, Zurich, 8 Switzerland). They received visual feedback through a portable computer screen; a bar which size was proportional to the airflow velocity turned from yellow to green when the desired airflow was reached (figure 6). The patients were instructed to breath so the bar would turn green. The visual feedback and the recording of the airflow were provided by the research software WBreath v3.41.4.1 (Ndd Medizintechnik AG, Zurich, Switzerland). All participants wore a nose clip.

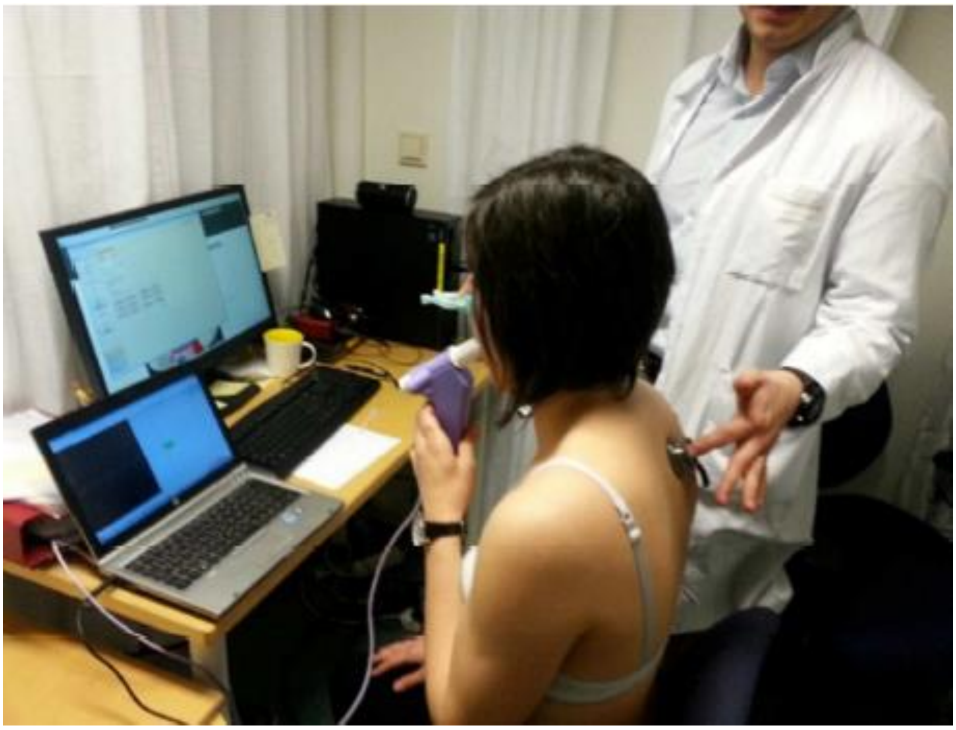


Figure 7.- Lung sound recording with airflow measurement.

4.1.3 Observers

Study I

We recruited seven groups of four participants each (n=28). Four groups consisted of general practitioners from different European countries (Norway, United Kingdom, Russia and The Netherlands). In addition, there was a group of sixth year medical students from UiT, one group of pulmonologists working at the University Hospital of North Norway and a group of experts (researchers) in the field of lung sounds.

Study II

We made an open invitation to medical students at UiT at the fourth, fifth and sixth year of medical school. The invitation was distributed via the university's on-line learning platform (Fronter) and visits to the classrooms. The students interested in participating sent an e-mail to be part of the study. We registered 30 students who were interested to participate in the project. Two decided not to participate before the start of the study, two did not show up and three did not complete the classification session due to lack of time. Thus, we obtained a final sample of 23 students, 19 women and four men. Fourteen participants were fourth year students, one was from the fifth year and eight were from the sixth year. The answers of the experts in study I were used to create a reference standard based on a majority rule (see page 34 “Study II”).

Study III and IV

Six observers were involved in the classification of the lung sounds. I, as observer 1, classified the recordings of all the participants included in the study. Two medical doctors and one experienced researcher in lung sounds served as observer 2. They divided all the recordings between them selves, so that each recording was classified by two observers. In addition, two experts (experienced researchers) in lung sounds participated as observers in the resolution of disagreements between observer 1 and observer 2.

4.1.4 Classification of lung sounds

To classify the recordings the observers could mark the type of AdLS heard during inspiration and expiration as independent dichotomous variables (present/absent). This allowed us to treat the events independently in order to have a more precise estimate of the agreement between observers. For example, if observer A heard expiratory wheezes and inspiratory crackles and observer B heard only inspiratory crackles they would agree that the recording contained crackles, but disagree about the presence of wheezes. The database that we built from the observers’ classifications could differentiate between agreement in each category of AdLS present in the recording or the agreement of the recording overall. The variables were the same for studies I, II, III and the first two steps of the classification in study IV. In the third step of classification in study IV we used a modified scheme with different classification variables (figure 10).

Study I

Twenty-eight observers received a power point (Microsoft, Redmond, WA, USA) presentation with 120 videos with sound recordings and their respective spectrograms. Age, gender and recording location, but no clinical information, were presented about the subjects. The spectrograms showed time on the x-axis, frequency on the y-axis and intensity by color saturation (figure 7). We first asked the observers to classify the lung sounds as normal or abnormal. If abnormal, they had to further classify them as containing crackles, wheezes (including rhonchi) or other abnormal sounds. It was possible to mark more than one option. The observers specified whether the abnormalities occurred in inspiration or expiration. In addition, they could mark if there was noise present in the recording that made the classification difficult.

We offered two options for answering the survey: an electronic form in Microsoft Access (Microsoft, figure 8), or a printed version of the questionnaire.

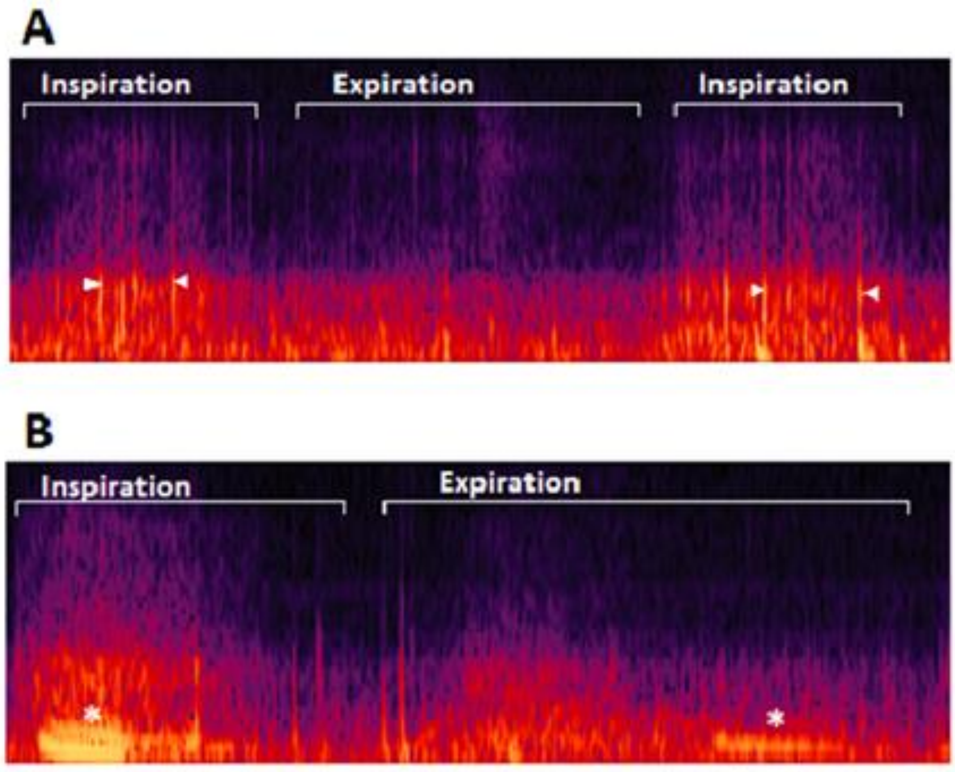


Figure 8.- Spectrograms of lung sound recordings showing crackles (A, arrowheads) and wheezes (B, stars).

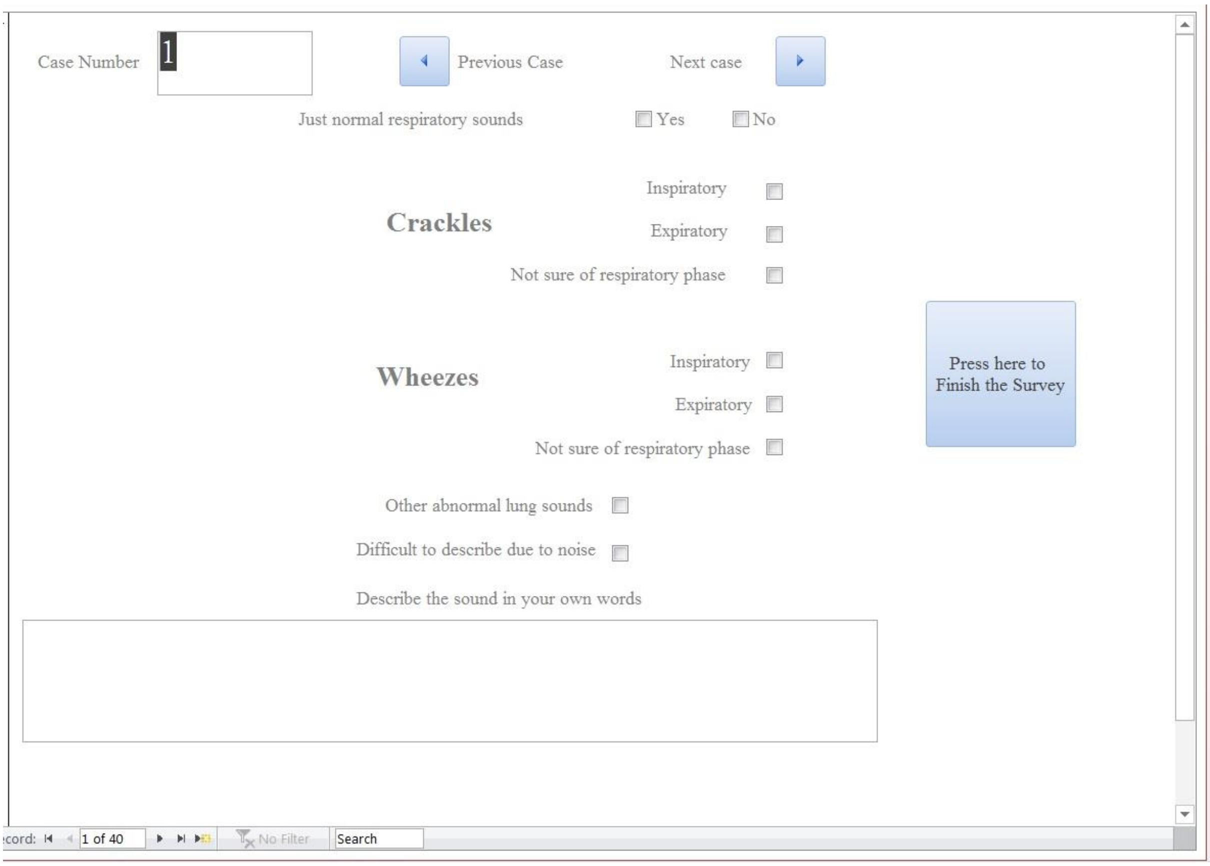


Figure 9.- Screenshot of the electronic version of the classification scheme used in study I.

Study II

We selected 30 recordings from study I containing 15 recordings with normal respiratory sounds, 9 recordings containing crackles and 6 recordings containing wheezes according to the classifications of the expert group of from study I. The reference standard was established by a majority criterion where it was necessary that three out of four experts called the presence of AdLS to be considered a positive finding. Otherwise, we registered the recording as normal.

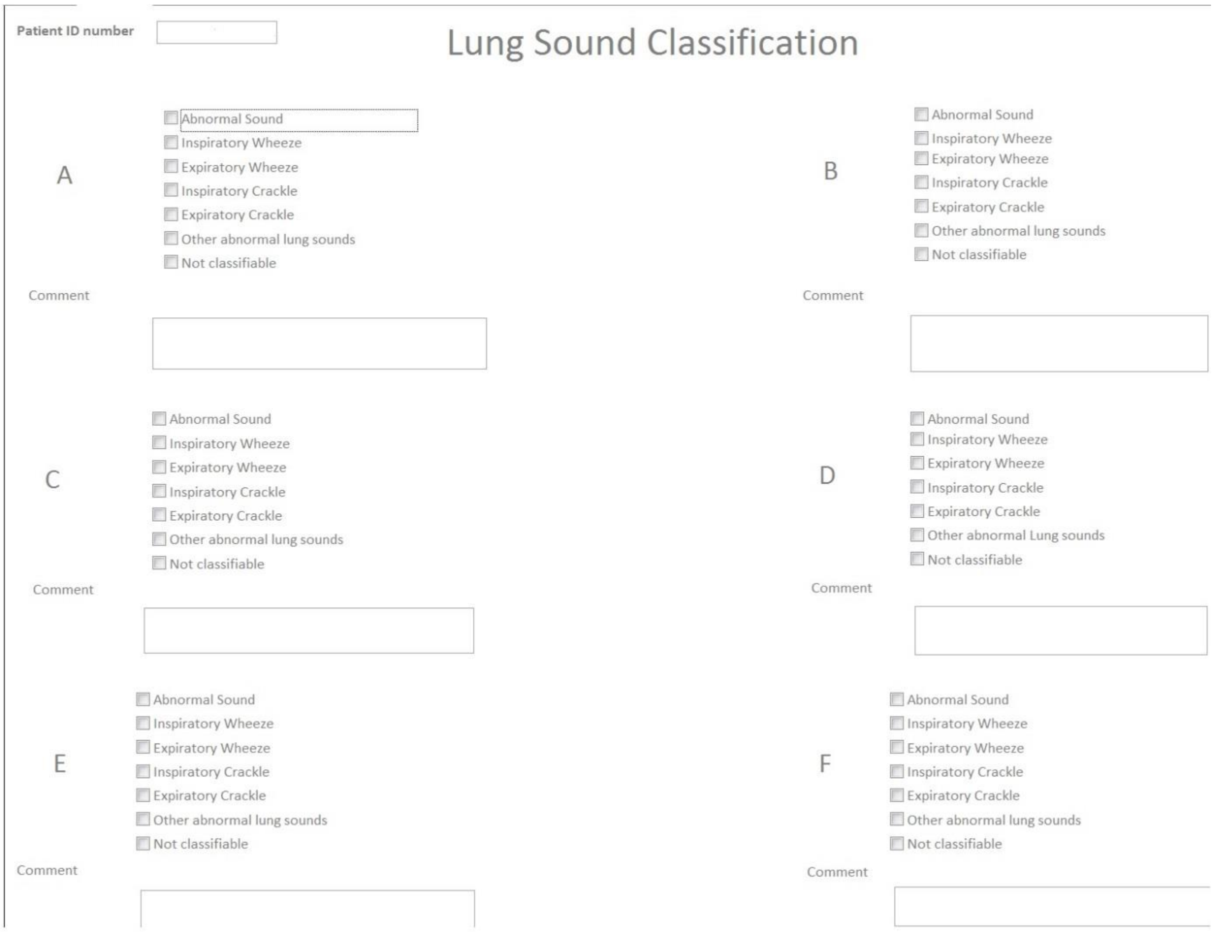
The classification session took place at a medicine school auditorium with a sound system and a screen projector to show the spectrograms. The students were free to sit anywhere in the auditorium. At the beginning of the session, we presented one example of normal lung sounds, one example of crackles and one example of wheezes with their respective spectrograms on the screen. We showed the students how these sounds looked in the spectrograms (figure 7). After this introduction, we played the 30 recordings in a random order in two sections, first, sound only. In the second section, the sound simultaneously with spectrograms on the classroom screen. There was a 20 minutes pause between the two sections. We presented no additional information beyond the sound and the spectrograms. The observers were not aware that the same sounds were played in both sections.

Each recording was played two times and the students had up to 30 seconds to classify it before the next recording was shown. The observers used their personal computers and an online classification scheme (Questback AS, Norway) when classifying the recordings. In this scheme, the observers had to specify if the recording contained only normal respiratory sounds. If this was not the case, the observers had to further specify if the recording contained wheezes, crackles or other sounds and if they appeared during inspiration or expiration. It was also possible to mark the recording as containing too much noise to be classified. At the end of the classifications session, Questback generated a report in .xml format.

Study III and IV

In these studies, the recordings were classified in a three-step procedure.

First, a pair of independent observers classified all the included sound recordings. I classified all the recordings while three other observers, two physicians and one lung sound researcher served as the second observer as described above (page 33). The observers listened to the recordings with a headset and simultaneously looked at the sounds spectrograms using Adobe Audition 5.0 (Adobe Systems, San Jose, CA, USA). They registered their findings in an electronic form (Access, Microsoft Corporation, Redmond WA, USA). (figure 9) First, they evaluated whether the recordings contained only normal respiratory sounds. If this was not the case, they specified if the recording contained wheezes (including rhonchi), crackles or other adventitious lung sounds and whether these were heard in inspiration or expiration. They reported if noise made the classification difficult. The observers could listen to the sounds as many times as needed. They were blinded to any kind of information about the participant.


Patient ID number

Lung Sound Classification

A
- Abnormal Sound
- Inspiratory Wheeze
- Expiratory Wheeze
- Inspiratory Crackle
- Expiratory Crackle
- Other abnormal lung sounds
- Not classifiable

Comment

B
- Abnormal Sound
- Inspiratory Wheeze
- Expiratory Wheeze
- Inspiratory Crackle
- Expiratory Crackle
- Other abnormal lung sounds
- Not classifiable

Comment

C
- Abnormal Sound
- Inspiratory Wheeze
- Expiratory Wheeze
- Inspiratory Crackle
- Expiratory Crackle
- Other abnormal lung sounds
- Not classifiable

Comment

D
- Abnormal Sound
- Inspiratory Wheeze
- Expiratory Wheeze
- Inspiratory Crackle
- Expiratory Crackle
- Other abnormal Lung sounds
- Not classifiable

Comment

E
- Abnormal Sound
- Inspiratory Wheeze
- Expiratory Wheeze
- Inspiratory Crackle
- Expiratory Crackle
- Other abnormal lung sounds
- Not classifiable

Comment

F
- Abnormal Sound
- Inspiratory Wheeze
- Expiratory Wheeze
- Inspiratory Crackle
- Expiratory Crackle
- Other abnormal lung sounds
- Not classifiable

Comment

Figure 10.- Classification scheme used in study III and in the first and second step on the classification in study IV.

Secondly, all the recordings in which the two observers disagreed were evaluated once more in meetings with the two initial observers and a third experienced observer. The three observers listened to the sounds and solved disagreements through consensus. When consensus was difficult to reach the sounds were submitted for classification in the final step.

In the final step, all recordings classified as containing adventitious respiratory sounds (1257 containing wheezes and 894 crackles) were re-classified once more. This time we used two pairs of observers consisting of one junior and one senior lung sound researcher. One pair classifying crackles and another pair classifying wheezes. These observers had the possibility to mark the findings as “certain”, “possible” or “absent”. Other characteristics of the sound were further described but remained out of the present analysis (figure 10). If crackles or wheezes were classified as certain by both observers or certain by one observer and possible by the other, the classification remained “present”. The recording was changed into absent when crackles or wheezes were marked as “absent” or “possible” by both observers. The recordings where the observers disagreed (present versus absent) were discussed in a face-to-face meeting with all the four observers and a voting was done where three out of four was required to classify an adventitious sound as “present”. At the same session, we classified all the sounds categorized as difficult or as “other sounds” in step two (n=41).

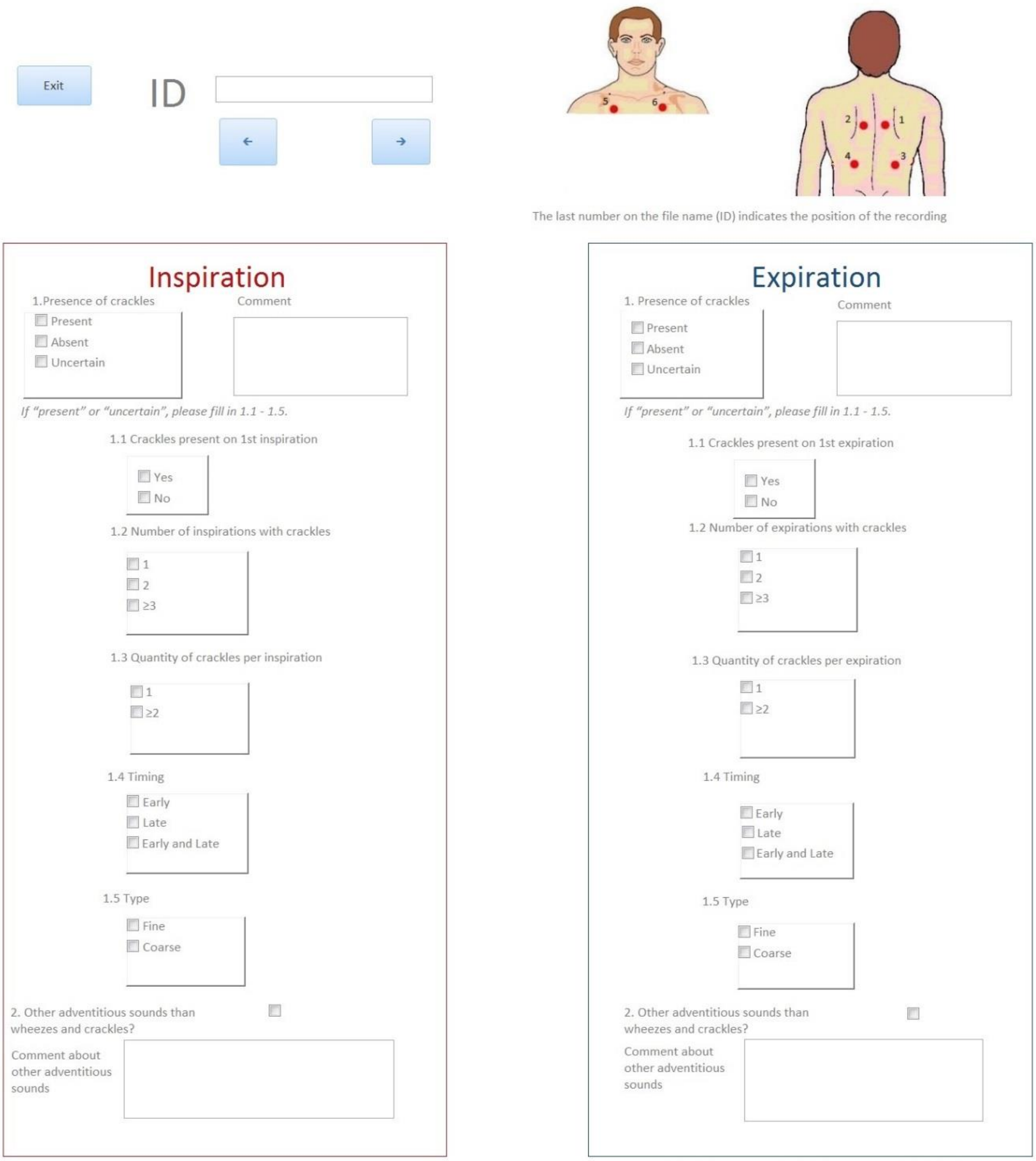


Figure 11.- Classification scheme used in the third step of the classification in study IV.

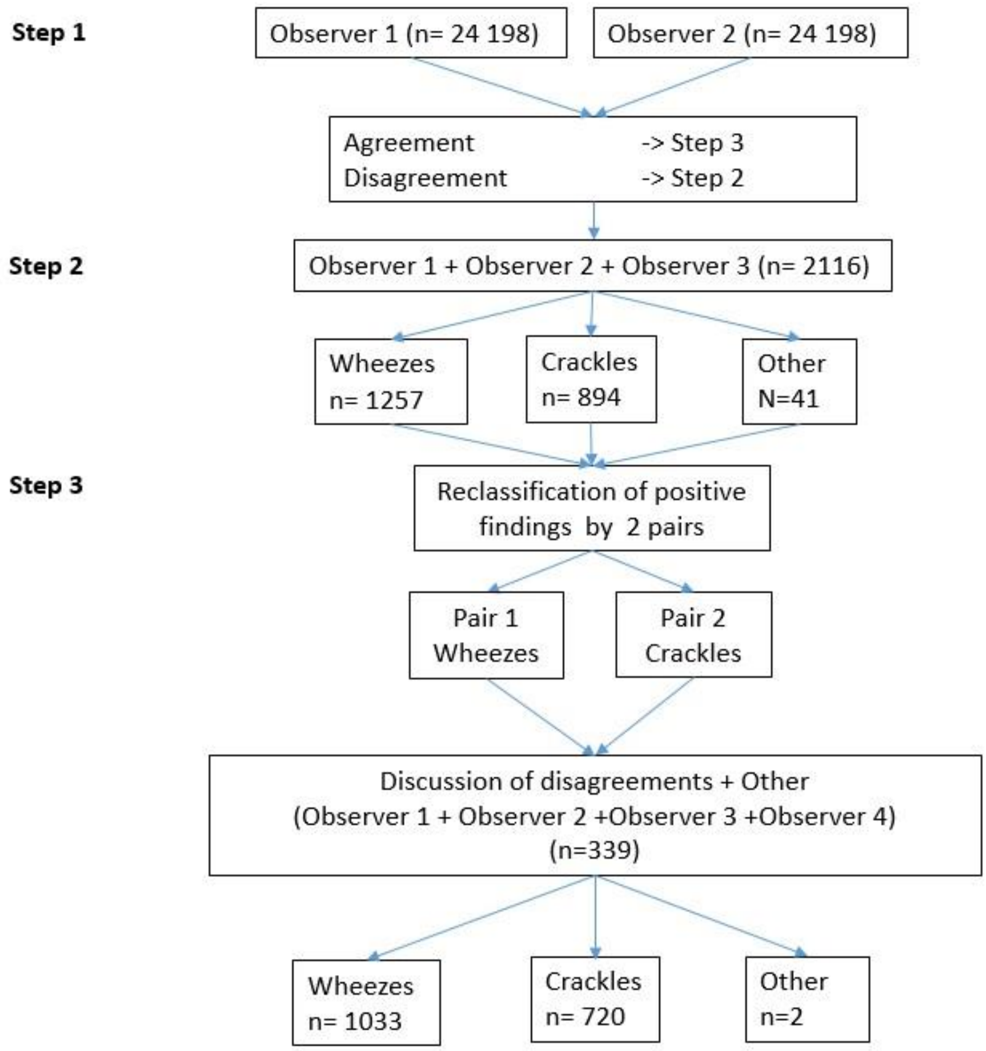


Figure 12.- Flow diagram of the classification steps with the number of recordings considered.

4.1.5 Statistical analysis

Throughout the four studies, the quantification of agreement was an important issue since the findings on auscultation were our main outcome. Since agreement is not a wide subject in the usual statistics course for PhD students, I took contact with the statistician Sophie Vanbelle at Maastricht's University who has agreement as her specialty. Our data in studies I, III and IV had a particular challenge because each participant had six recordings. The agreement estimates should account for this participant dependency to avoid bias. Vanbelle had developed a method to calculate agreement for this type of multilevel data in an R statistical package named "multiagree" and guided me through the analysis of study I and IV. The statistical package for the analysis of kappa was developed for use in the program "R" (80). The statistical analyses for each study will be described in detail in the following paragraphs.

The most intuitive measure of agreement is measuring the percentage of agreement of two (or more) observers when classifying an X amount of items. However, there is a possibility that a proportion of the agreement observed is due to pure chance rather than "true" agreement. In order to overcome this problem, Cohen develop a statistic to estimate agreement beyond chance. He named the measure of agreement Kappa and described it as follows:

"The coefficient k is simply the proportion of chance-expected disagreements which do not occur, or alternatively, it is the proportion of agreement after chance agreement is removed from consideration"[(81)]

The limits of kappa can be pragmatically defined from 0 to 1 where 0 represents agreement by chance and 1 represents perfect agreement (negative values are possible, but it is outside the scope of this thesis to describe this). There is no fixed cut-off of what represents "good" agreement. Landis & Koch proposed a table of to classify the strength of agreement (see table 1).[(82)]

Kappa statistic	Strength of agreement
<0.00	Poor
0.00 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost Perfect

Table 1.- Table with arbitrary divisions of kappa coefficients as described by Landis and Koch.[(82)]

However, many researchers no longer recommend to use it since the divisions are arbitrary and with a relevance depending on the problem under study.[(83)] Kappa should be regarded not only as a descriptive statistic but more as an effect size measurement with intended use in inferential statistics.[84] Therefore, other properties are equally important such as the calculation of its confidence intervals and tests of statistical significance.[83, 84] There are variations of the kappa coefficient for paired observers (Conger), or multiple observers (Fleiss).[83]

Study I

We calculated the probability of agreement and multirater Conger's kappa using the delta method for the analysis of multilevel data.[85] Conger's kappa coefficient was chosen over Fleiss' kappa due to pairwise comparisons. We analyzed the intragroup agreement in each of the seven groups of observers when classifying the recordings for the presence of wheezes and crackles disregarding the breathing phase. We used the statistical software 'R' V.3.2.1 together with the package 'multiagree' for the statistical analysis of kappa statistics.[80] In order to permit the comparison of the agreement levels between and within groups, within and between-group agreement levels were summarized in a matrix. The diagonal elements represent the mean agreement level between all possible pairs formed by two observers in the same group, and the off-diagonal elements represent the mean agreement level between all possible pairs with one observer in one group and the second observer of the pair in another group. This information was summarized in correlograms using the R package 'Corrplot'.[86]

Study II

We calculated Cohen kappa for the agreement between the observers and the experts, and Fleiss kappa for all the observers as a group. We then compared the kappa values obtained in the sections with and without the use of spectrograms and calculated p values to explore for statistically significant differences using an adaption of Hotelling's T^2 test described by Vanbelle, S.[87] In this analysis, the recordings were clustered by the individual they were recorded from. We used Holm's correction procedure to adjust p values for multiple hypothesis testing. We used R version 3.2.1 and the package "magree" to perform all the calculations.[88] Significance level was set at $p < 0.05$. In addition, we calculated sensitivity and specificity of each participant using the experts' classification as the gold standard. We tested for significant differences in sensitivity and specificity with and without the intervention using paired Wilcoxon signed rank test with continuity correction.

Study III

Descriptive statistics were used to characterize the sample. Based on the airflow signal recorded with the research software WBreath v3.41.4.1, mean inspiratory and expiratory peak airflow per subject were determined (computing the mean of the six recordings at standardized airflow). To analyze the relative reliability of peak airflow, the Intraclass Correlation Coefficient was determined.(89) Relative reliability was computed using the scores obtained from six recordings. To determine the agreement on the presence of crackles/wheezes between breathing modes, the percentage of agreement and Cohen's kappa were used.

Study IV

We calculated age standardized prevalence of wheezes and crackles for women and men using the population distribution from the municipality of Tromsø per January 2018.(90) Then the outcome was divided into three categories, wheezes and no crackles, crackles and no wheezes, and both wheezes and crackles, irrespective of respiratory phase. We calculated prevalence by sex, age, and other participant characteristics and used linear models to explore statistically significant differences between the groups. Tukey procedure was used to account for multiple testing. The continuous variables were dichotomized with cutoff values for age ≥65 years, for oxygen saturation of ≤95%,(91) mMRC score ≥2,(92) and FEV_1 below lower limit of normal (LLN) according to the Global Lung Initiative reference.(78)

We studied the relationship of demographic and clinical variables with the presence of wheezes and crackles using univariable logistic regression. In this analysis wheezes was counted as present also when accompanied by crackles and vice versa. The following outcome variables were considered separately: (1) presence of any wheeze, (2) presence of wheezes only at the inspiratory phase, (3) presence of wheezes at the expiratory phase and (4) presence of wheezes at the expiratory phase in two or more recording sites. For crackles, the outcomes were (1) presence of any crackle, (2) presence of inspiratory crackles, (3) presence of inspiratory crackles at two or more locations, (4) presence of expiratory crackles only. In these analyses, the categorical variables measuring FEV_1 as percentage of predicted and oxygen saturation were substituted by continuous data to avoid loss of information. We divided age by ten and kept it as a continuous variable.

We entered all the variables that showed a statistical significant correlation for each outcome in the univariable analyses into multivariable logistic regression models. We performed a backward elimination procedure with a threshold of $p<.05$ to obtain the best fitting models for each outcome. We plotted Receiver Operator Characteristics (ROC) curves for all the final models and calculated the area under the curve (AUC) with the r package "pROC".[(93)] Multicollinearity in the final models was assessed using variance inflation factor with the statistical package "car".[(94)] We used R statistical computing version 3.2.1 package to perform all the calculations.[(80)] Results were considered significant at 5% level.

5 Summary of the papers and main results

The main goal of this PhD thesis was to estimate the prevalence of AdLS (more specific wheezes and crackles) in a general population (study IV). Nevertheless, other important questions needed to be explored to assure optimal quality of the estimates in study IV. The first one concerned the reliability of the classification of AdLS. How large is the agreement between clinicians when classifying lung sound recordings? (study I). We needed to test the reliability of our classification system and look for potential sources of variation. However, in our classification system we used sound spectrograms, in addition to sound recordings, as a helping tool with the hope to achieve a better agreement across observers. We aimed to achieve high external validity with respect to clinical practice, but the use of spectrograms is not common among clinicians. Therefore, we wanted to investigate the impact of using spectrograms when classifying lung sound recordings (study II). Our recording protocol for study IV had no quantifiable control of the participants' breathing airflow velocity. We chose to do this in order to make it similar to the auscultation routine at the doctor's office. This came at the cost of limited comparison with previous studies in lung sounds, which usually do control for airflow velocity. We performed study III to have an idea about how this decision might affect our results in study IV.

5.1 Paper I

The probability of agreement for crackles varied between 65% and 87% in the different groups of observers. Congers kappa ranged from 0.20 to 0.58 and four of seven groups reached a k≥0.49. For wheezes, we observed a probability of agreement between 69% and 100% and kappa values from 0.09 to 0.97. Four out of seven groups reached a k≥0.62. We found it likely that variation in the level of agreement stemmed partly from variation in the use of terminology when we observed that the variation in the Russian part of the study could be explained by a difference in the classification systems. (See paper I) We concluded that digital recordings with the use of spectrograms is a method suitable for research of lung sounds. However, to improve agreement it would be necessary to implement training of raters, standardization of the terminology, multiple independent observations and consensus agreement.

5.2 Paper II

The students observed a mean prevalence of wheezes of 9.7 (6 – 15) without spectrograms and 8.3 (5 – 12) with spectrograms. In the case of crackles the students observed a mean prevalence of 11.5 (4 – 22) and 10.9 (5 – 18) in the same order. The mean proportion of agreement (%) and Cohen kappa (k) with the experts for all 23 participants classifying wheezes without spectrograms was 82 % and k=0.56. We observed 88 % and k=0.68 with the use of spectrograms. In the case of crackles we observed a proportion of agreement of 72 % and k=0.38 without spectrograms. With the use of spectrograms 80 % and k=0.56. Fleiss kappa values for the multirater agreement were k=0.51 and k=0.56 (p=.63) for wheezes without and with spectrogram, respectively. For crackles, we observed k=0.22 and k=0.40 (p=<0.01) in the same order. Compared to the expert panel's classification, 13/23 students had a positive change in kappa when classifying wheezes (one with p<.05), and 16/23 (two with p<.05) when classifying crackles. All the statistically significant changes were in the direction of improved kappa values (0.52 - 0.75). The median sensitivity for wheezes did not present a significant change but the specificity was higher in the classification with the use of spectrograms (p=0.002). In the case of crackles, there was a significant increase in sensitivity (p=0.03) when using spectrograms but without significant change in specificity.

5.3 Paper III

We found expiratory wheezes were present in 18 (16%) participants at spontaneous breathing and in 23 (20%) at standardized breathing. Inspiratory crackles were present in 19 participants at spontaneous breathing (16%) and in 18 (16%) at standardized breathing. The prevalence was not significantly different between the two methods. However, only nine participants in the case of wheezes, and five participants in the case of crackles were detected by both methods. The agreement of the two methods was k= 0.32 for expiratory wheezes and k=0.13 for inspiratory crackles. By this, we conclude that the mode of breathing has an impact on which participants present AdLS. Although adventitious sounds were found with similar frequency between the modes of breathing, less than half of these subjects were identified by both. However, spontaneous breathing was not inferior to standardized breathing in reflecting lung disease.

5.4 Paper IV

In our large sample from the Tromsø study the crude prevalence of AdLS was 28%; 18 % had wheezes, 13% had crackles. Expiratory wheezes and inspiratory crackles were the most common findings. The age standardized prevalence of wheezes was 19% for women and 15 % for men. For crackles, the standardized prevalence was 11% and 9% for women and men, respectively. In the multivariable analysis, we observed that age, female sex, self-reported asthma, and current smoking predicted the occurrence of expiratory wheezes. FEV_1-Z score was a significant predictor of the occurrence of inspiratory wheezes. Age was the only predictor of expiratory crackles. In the case of inspiratory crackles, significant predictors were age, current smoking, rheumatoid arthritis mMRC ≥2, low oxygen saturation and FEV_1 Z-score. Age was the strongest predictor of crackles. (Figure 12) The presence of AdLS was not associated with the presence of airway infections in the week before the examination.

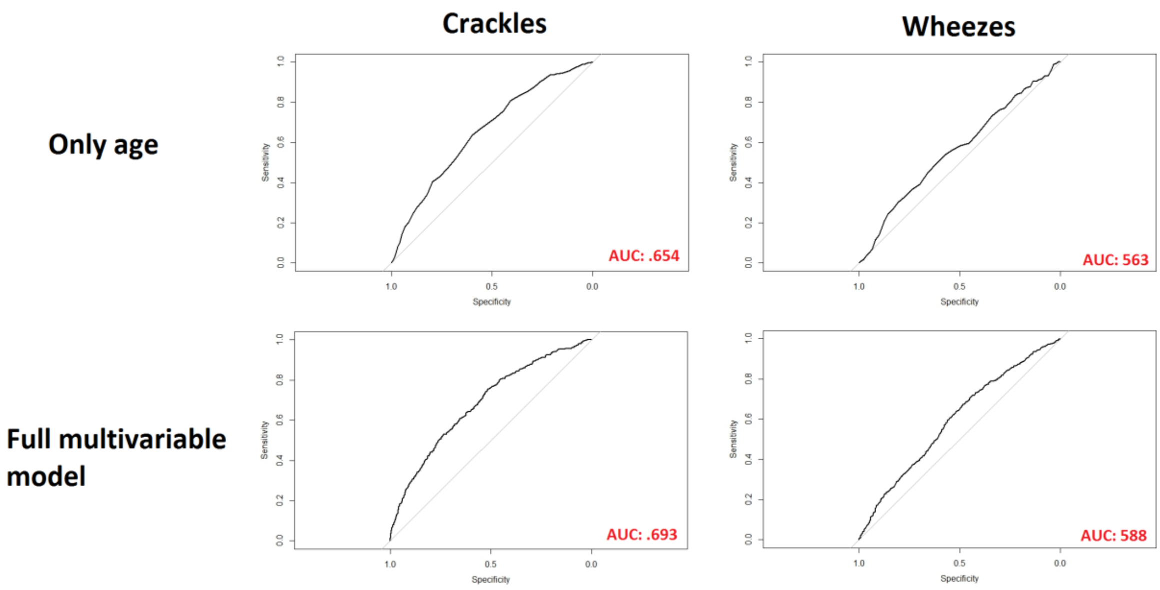


Figure 13.- AUC for crackles (a) and wheezes (b) with age as independent variable. The figures in the second row show the AUC for the complete multivariable models presented in study IV for crackles (c) and wheezes (d).

6 Discussion

6.1 Methodological considerations

6.1.1 Design

A cross-sectional study was the logical choice of design to estimate the prevalence of AdLS. To execute our study within the Tromsø study gave us the advantage that the people invited to participate belonged to a delimited geographical area. This gave us the opportunity to define a concrete population at risk making it easier to account for missed cases and achieve a more reliable denominator to calculate the prevalence proportion. In addition, we had the possibility to access a large variety of variables collected in a standardized and professional manner.

In study I, we chose a cross-sectional design with seven clusters of equal size with different backgrounds, experience and nationalities. It can be argued that these groups were too different to be compared. However, the main objective was not to find an estimate of the "true" variation in the classification of lung sounds. Instead, we wanted to explore how big the variation could be in a complex sample and to see how we could improve it. I think this design worked good for this purpose.

In study II, we applied a lung sound survey with and without spectrograms accompanying the sounds in order to estimate the influence of spectrograms. We did this because spectrograms are not part of common clinical praxis which might limit the external validity of study IV. In this design, we needed to keep a balance between a large enough number of recordings to classify (in order to achieve a good statistical power) but small enough to not discourage participation. The sample size turned out to be a possible limitation since only differences in kappa >.30 for crackles and >.40 for wheezes showed to be statistically significant. If we had used a larger sample of sounds smaller improvements could have become statistically significant. Another weakness was the study is subject to testing bias.[95] This means that the improvement could have been due to the observers taking the test a second time and getting better at it (learning effect).

In study III, we used a subsample of the participants included in study IV. Based on the descriptive characteristics of the participants we can say that they were similar in age, smoking habits, FEV1, mMRC and the presence of heart and lung disease as the total sample used in study IV. Therefore, we can conclude that the sample in study III was representative of the participants in study IV. In study III, we assumed that the low agreement in the presence of AdLS in the same individual with the two different methods is due to our intervention (the introduction of a fixed respiratory airflow velocity). However, the quasi-experimental design employed to address this question had one important weaknesses: The lack of a control group. This could potentially hamper the internal validity of study III.[95] The presence of AdLS may vary over short periods of time, and we did not control for temporal change in AdLS independently of the intervention. A possible alternative would have been to include a group of participants performing recordings twice at spontaneous breathing with a similar time interval as in the intervention group. If the variation remained high in spite of same breathing then it would be likely that the variation was due to variation of AdLS rather than change in breathing mode.

Temporal variation of AdLS may also be a limitation in the design of study IV. We did not perform a second test using the same method on a different time point to explore if the positive findings were replicable. Further studies of temporal variation in AdLS is in my opinion highly recommendable.

6.1.2 Inter-observer agreement and misclassification error

The classification of lung sounds is a subjective task and variation between two or more observers is expected. The outcome variables in study IV are the result of a classification process of the lung sound recordings. If the classification had a low quality, then the quality of our outcome and analysis will be low as well. In the following paragraphs I will explore the issues of reliability and possible bias in our classification.

Since the observers had no knowledge about the participants' health status, a systematic misclassification due to expectations should not happen, at least not when the first of the six recordings was observed. The classification of the other five recordings might have been influenced by each other.

This source of bias should not hamper our results, since we controlled for the hierarchical structure of the data in the calculation of kappa coefficients.[(88)] Misclassification bias would affect our results by diluting the strength of the associations in our analysis.[(96)] To avoid/minimize the occurrence of systematic error we presented the sounds in an anonymous fashion where the only information about the subject was an ID number. This was a necessary step to avoid that certain groups of participants would get a higher/lower chance of getting a positive finding (Old, sick, smoker, etc). However, this comes at a cost of a reduced representation of clinical praxis where doctors do have this information in advance.

To avoid systematic misclassification dependent on the classifier, we decided that two observers should classify all the recordings independently. At this step, one classifier remained constant and three other medical professionals shared the position of the second classifier. We noticed that the difference in the rate of AdLS between the observers was somewhat reduced after the consensus meetings. Acquiring experience in the classification of the material could have contributed to reduce disagreement. (figure 13).

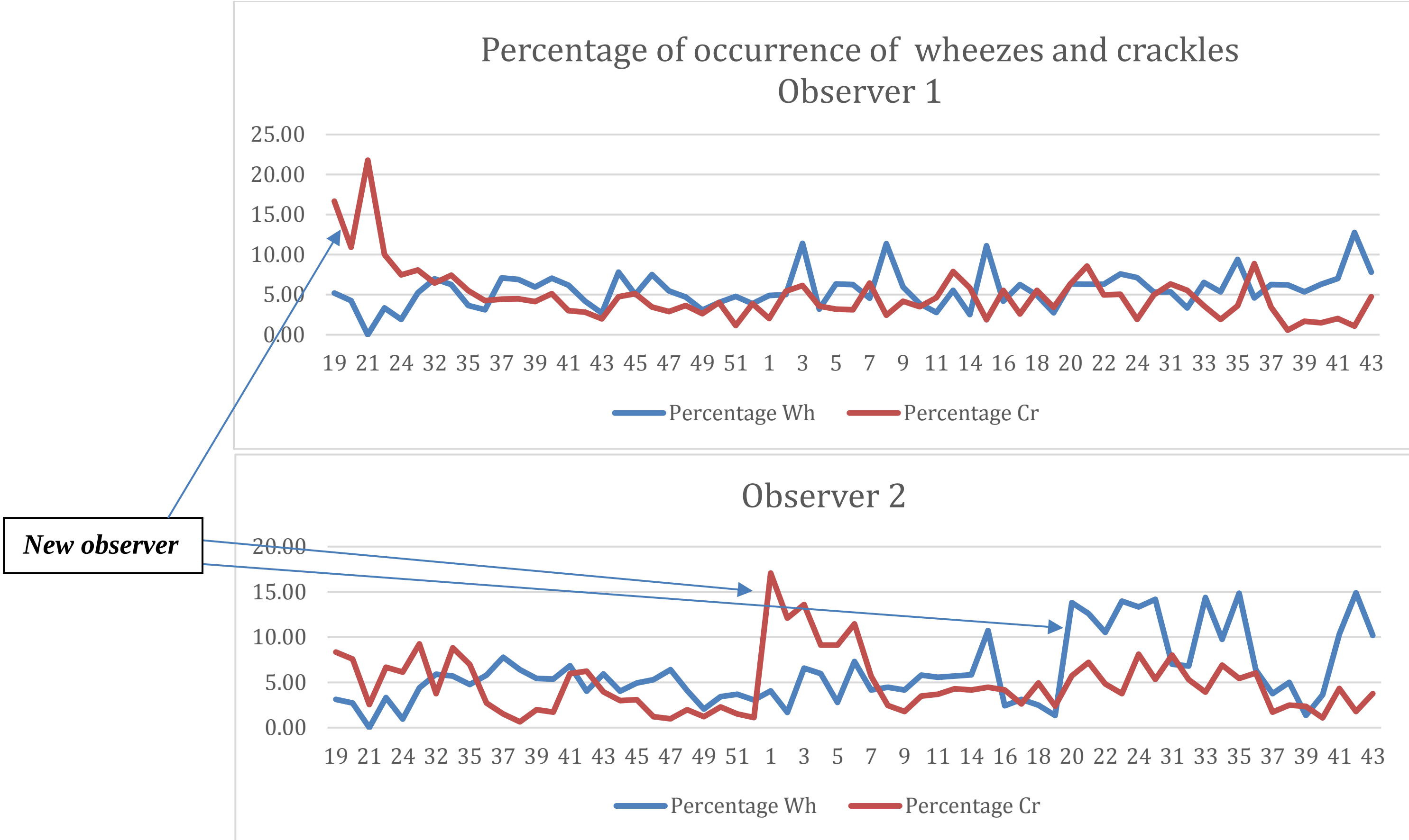


Figure 14.- Percentage of wheezes and crackles by calendar week of data collection.

In my opinion, this situation should not have a significant impact in the results of study IV since disagreements were discussed and resolved in the second step of the classification. If we had used four observers each classifying 25% of the sample, then the prevalence rate of AdLS would have a higher chance for presenting observer-dependent variation. Our method was good to avoid over-registration of AdLS, but there is a chance that both observers missed some AdLS. Then the AdLS prevalence we report could be somewhat smaller than in reality.

At step two of the classification process all the disagreements were discussed in a face-to-face session. The two involved classifiers plus an experienced lung sound researcher participated in this session (n=3). The panel made decisions by consensus. This mechanism was employed to avoid systematic error by one of the observers, but this face-to-face session could have produced some residual bias due to personal interaction. For example, some of the participants might try to avoid being too strict or avoid “conflict” with the other observers.

To minimize this source of bias we implemented a third stage of classification where all the positive findings should be independently reclassified once more by two observers. At the third step of the classification, we chose to discard as positive findings the recordings marked as “possible” by two observers. If we had included a high number of uncertain finding we could have experience a dilution of the associations, thus changing our conclusions. However, in a sensitivity analysis, they were classified as “present” but the results were similar, and the conclusions in study IV remained unchanged.

6.1.3 Selection bias

For study III and IV we used a sample from the Tromsø study. This survey obtained a response rate of 65%. Over half the invited population attended and therefore we can say that the sample is representative of the total population in Tromsø.[77] However, it is likely that people being too sick to attend the survey would refrain from participation. By intention, the segment of the population who were 60 years or older and individuals of female sex were overrepresented. We have controlled for this situation by reporting the standardized prevalence corrected for age and sex. However, some residual selection bias cannot be ruled out.

Our final number of classified participants represents 19% of the participants in Tromsø 7 and 48% of those attending the second visit. We reached our final sample by a series of random processes. We originally aimed to record and classify lung sounds in all the participants attending the second visit (n= 8 346). However, we recorded lung sounds in only 6035 (72%). This was partly due to lost days at the recording site by the inability of having a technician in place. We assume that the days where the recording technician was unavailable happened at random. Some participants “dropped out” since they did not have time for all the second visit examinations. Another issue was that after starting the classification process, we realized that doing all the individual classifications, the face-to-face meetings, and all the three steps in the classification would become more time consuming than expected. Therefore we decided to classify all the participants recorded in 2015 (n=2415) and all the participants randomized to an echocardiographic examination (n=1618). The echocardiographic examination was randomly allocated to 30% of the participants in the second visit. Since all these processes were more or less random and we had no previous knowledge of the health status of the participants, the possibility of a selection bias in our sample can be regarded as low.

In the case of study III, we recruited the subsample in a consecutive continuous period of four weeks during the data collection period. The prevalence of wheezes and crackles in both study III and IV were similar and therefore we think that the sample in study III is representative of the sample in study IV, although those selected for echocardiography were not included in study III.

6.1.4 Respiratory flows and volumes

Even though the prevalence rates did not vary much by the two modes of breathing (spontaneous and standardized), we observed that only five of 19 participants presenting inspiratory crackles and nine of 18 participants presenting expiratory wheezes were identified by both methods.

If we extrapolate this observation to study IV, we can then say that if we had performed our AdLS recording protocol with a fixed airflow velocity of 1.5 l/s we would have a similar frequency but would have detected two thirds of different participants with crackles and another different half with wheezes. This could have had an impact on the associations we described in study IV.

We could argue that the persons with reliable and repeatable positive findings probably were less healthy than those detected with only one method. However, in this study neither of the two methods was better to predict disease-related endpoints. Nonetheless, it is important to remark that the number of individuals with positive findings in both methods might be too small (crackles n= 5, wheezes n=9) to find statistically significant differences in variables used to diagnose disease such as decreased FEV_1, or high dyspnea scores (mMRC).

6.1.5 Information bias

In study IV, the collection of lung sound recordings started on the 7th of may 2015 and ended on the 28th of October 2016. The prevalence of lung sounds could in theory be affected by conditions that have a seasonal variation such as temperature,[97, 98] pollen,[99] dust and air pollutants concentration,[100, 101] and influenza season.[102, 103]

This problem is not possible to avoid due to the logistical challenge of recording lung sounds in a large amounts of participants. However, it is possible to identify season-related variation in the prevalence rates of AdLS and possibly correct for it.

I divided the participants with AdLS by winter and summer months. I considered winter months the period from October to April, and summer from May to September. I explored for differences in the prevalence rate of positive findings between these two periods using logistic regression models to test for significant differences. Results of the analysis are reported in table 2.

	Cr +	**Cr -**
Summer	241	1514
Winter	270	1852

OR : Summer (reference), winter (0.91; 95% CI 0.76 – 1.10), p=.356

	Wh +	**Wh -**
Summer	305	1450
Winter	386	1736

OR : Summer (reference), winter (1.06; 95% CI 0.90 – 1.25), p=.511

Table 2.- Distribution of participants by season and the presence of crackles and wheezes and results of logistic regression analysis.

We can observe that the prevalence rate does not show a statistically significant variation in this analysis. I can therefore conclude that season related variation did not have a significant impact in our results.

6.2 Clinical considerations

6.2.1 Impact of the reliability of the classification of lung sounds in clinical practice.

We observed some variation in the classification of AdLS among medical professionals in study I. We found an absolute agreement of 97% (k=0.46) for inspiratory crackles and 96% (k=0.56) for expiratory wheezes. Although these values did not show perfect agreement, they are in line with other widely used diagnostic procedures. For example: Gleason score pathology reports (k=0.36-0.46),[(104)] BI-RADS evaluation of mammography (k=0.48);[(105)] Electrocardiograms (k=0.29-0.54);[(106)] Evaluation of clinical signs in lower respiratory tract infections in children (k=0.25-0.48);[(107)] Identification of seizures in electroencephalography (k=0.58);[(108)] Fatty liver detection by ultrasound (k=0.54) and lung ultrasound to detect B-lines (k=0.55).[(109)] However, the fact that agreement is similar to other diagnostic procedures does not make it valid *per se*. I will approach this discussion first from an epidemiological and then from a clinical point of view.

From an epidemiological perspective, we have to analyze whether lung auscultation can differentiate positive from negative cases in a certain population. The kappa values we observed in our studies are not optimal, but not a product of pure chance either. It is important to note that AdLS did not have a 50 % prevalence (In study IV: wheezes= 18%, crackles 13%), which would be ideal to obtain highest possible kappa values. As discussed by Kraemer et al (84) a kappa value of 0 indicates either that the heterogeneity of the patients in the population is not well detected by the instrument or that the subjects in the populations are homogenous (0% or 100% prevalence of the trait). Therefore, if the population is more homogeneous it reduces the ability of the instrument to differentiate positive from negative cases. The influence of prevalence on Kappa might seem as a disadvantage,[(110)] but it is in fact a desired property that helps us to assess how the instrument (auscultation in this case) will perform in a general population.[(84)] Consequently, the Kappa values in study IV should not be regarded as a measure of the reliability of auscultation alone, but as the reliability of auscultation in a population with a similar prevalence of AdLS. From our results in study IV, we can conclude that auscultation has the ability to differentiate positive from negative cases of AdLS in the general population well beyond chance, but still with some limitations. These limitations are influenced by both the amount of measurement error plus the variation of the population.[111] Therefore, the reliability of auscultation will be different in populations with different prevalence.[112]

From a clinical standpoint, kappa would not be the best measure to quantify agreement. Clinicians make decisions one patient at a time independently of the distribution of the trait in a certain population and an absolute measure of agreement, like the probability of agreement, would be a more useful tool to evaluate the clinical performance.[113] In study I, we found that clinicians in the groups where translation was not an issue, have a probability >80% to agree on crackles and >90% on wheezes. These probabilities were similar within and between groups. This means that clinicians classify the phenomena (wheezes and crackles) in a similar manner despite their different background. Nonetheless, this means that they would disagree in about 10-20 out of 100 cases and this shall not be overlooked. However, the clinical implications of such misclassification error are hard to estimate from the data presented. New studies could help clarify this question.

I can therefore conclude that the use of lung auscultation has a reliability that makes it acceptable for use in the clinical practice and is valid when used in a general population. However, it is far from perfect. Beyond moderate inter-observer agreement, we also found some evidence that suggests a high variability in AdLS in the same subjects. Therefore, the gravity of the clinical decisions made with the aid of auscultation should be in accordance with these limitations. The studies included in this thesis cannot give a definitive answer on what does lung auscultation adds to the other parts of the physical examination. Therefore more studies in clinical usefullness of AdLS are necessary.

6.2.2 The use of spectrograms in the classification of lung sounds

We observed some improvement in the agreement between medical students and lung sound experts in the classification of crackles and wheezes when aided by spectrograms. The agreement within the medical students also had a statistical significant increase. I therefore conclude that the use of spectrograms might be useful for improving the classification of lung sounds.

This improvement in the classification of lung soud recordings might have influenced the results in study IV. This might have improved our sensibility to detect AdLS. The choice of using spectrograms might remove some of the external validity of study IV since the use of spectrograms is not common in clinical practice.

For several reasons spectrograms seems like an attractive tool for education and training of health personnel. First, the production of spectrograms is an easy and cheap task with the technology available now. In addition, the visualization of the sounds could give the students the opportunity to pair the acoustic phenomena with a picture in order to ease the recognition of AdLS. A visual aid could also facilitate group discussions about AdLS. The training programs on auscultation could be more efficient by including the use of spectrograms. Nevertheless, this is just an assumption based on a small study. Larger studies with better designs are needed to investigate how useful this tool would actually be for educational purposes.

In the case of daily clinical practice, I dare to say that the use for this tool would be rather limited. It is hard for me to imagine that busy primary care physicians would invest time into looking at spectrograms of each of the auscultation procedures in their office when the gain of doing so might not be very large compared to standard auscultation for an experienced physician.

6.2.3 AdLS and variation with airflow velocities in clinical practice.

Clinicians in daily practice perform chest auscultation at spontaneous breathing airflow velocity and therefore we assume that the conclusions in study IV are valid in that clinical context (with respect to breathing mode). Our results also suggest that lung sound research with higher airflow velocities than relaxed deep breathing might have a limited validity at the doctor's office.

Due to the limitations in our study designs, we cannot rule out that the variation we found is due to the instability of AdLS rather than the change in airflow velocity. Jácome and Marques showed that the presence of AdLS is reliable in patients with a diagnosis of COPD at spontaneous airflow (0.4 – 0.6 l/s).[114] However, the situation could be different in a different subset of patients. I find it likely that patients with AdLS due to a manifest lung disease would have more stable AdLS compared to patients with subclinical lung conditions or healthy individuals.   However, this cannot be confirmed with the data we have presented and a proper study with repeated recordings is necessary to answer this question.

6.2.4 Prevalence of wheezes and crackles in the general population and main clinical associations

We found that about one in four in our sample had AdLS where 18% had wheezes and 13% had crackles in at least one location. This means that primary care physicians have a chance of 1/4 to hear AdLS if they auscultate every adult in their practice. However, we have actively used spectrograms during the classification process. In addition, we have used several observers and consensus meetings. This could have had influenced our results by increasing the sensibility for the presences of AdLS. This protocol would be hard to reply in a clinical context.

In as much as 73% of the participants with wheezes and 72% with crackles the sounds were heard at only one of the six recording sites. We observed that the strength of the clinical associations presented in study IV increased in the participants with AdLS in more than one recording site compared to those who only had AdLS in one recording site. My interpretation of these results is that the increase in the number of positive recording sites has a stronger association with clinically verifiable disease.

6.2.5 Associations with age

Age was the stronger predictor of AdLS, particularly of crackles. This variable remained significant in all the analyses we performed. Age also was the variable with a very high contribution to the AUC in the multivariable models. (figure 12)

Lung function shows a sustained reduction through age.(115) The reduction of elastic and supportive tissue around the distal airways contributes to an earlier collapse during expiration and a sudden opening at inspiration.(115) These changes could explain the relationship between age and crackles.(24) In the case of wheezes, the relationship with age is not that clear and it might be related to a higher burden of obstructive disease or advanced airway remodeling due to the increased amount of years since diagnosis.(116)

The association with age is an interesting finding from a clinical perspective. Crackles can appear due to an acute disease such as pneumonia,[28] but it is possible that they can also appear as a consequence of a long term degeneration. Therefore, the clinical value of crackles in relationship to acute diseases might be reduced by increasing age of the patient. The utility of crackles in older adults in the absence of acute disease should be investigated for possible diagnostic or prognostic usefulness.

6.2.6 Associations with symptoms and self-reported disease

The presence of AdLS was not associated with the presence of airway infections in the week before the examination. The presence of crackles is a significant predictor for the prescription of antibiotics,[117] but as mentioned above there is a good chance that crackles were present before and independently of an airways infection, especially in older adults. It is possible that the use of crackles as criterion for the prescription of antibiotics contributes to the over prescription of antibiotics among the elderly.

We found a significant association between wheezes and self-reported asthma. This a well-known association that has largely been examined with questionnaires asking about self-reported wheezing.[20]

We observed an association of the presence of crackles with rheumatoid arthritis (RA) which I found interesting. RA is associated with a higher risk of interstitial lung disease[118] and/or COPD.[119, 120] Lung involvement is suggested to occur early after the diagnosis of RA, and up to half the patients with recently diagnosed RA have radiological changes in the lungs.[121] Also smoking is a known risk factor for both COPD[122] and RA.[123] However, the association we found seemed to be independent of smoking since both smoking and RA were statistically significant variables in the multivariable model with any crackles as outcome in study IV. It is important to mention that we had no objective measure to corroborate the presence of an RA diagnosis given by the patient. Therefore, this association should be interpreted with caution.

The presence of dyspnea measured by mMRC was associated with inspiratory wheezes and inspiratory crackles. In some of our multivariable models this association disappeared when SpO_2 and or FEV_1 were included in the multivariable model. Low SpO_2 and or FEV_1 are causes of dyspnea.[124] This could indicate that reduced FEV_1 and/or SpO_2 explained much of the associations between mMRC ≥ 2 and

AdLS. However, there was an exception in the multivariable model for inspiratory crackles where mMRC ≥2 remained significant in the model. Inspiratory crackles are also related to the presence of heart failure[57] and the significant association with mMRC ≥2 adjusting for SpO_2 and FEV_1 could perhaps be explained by heart disease [124, 125] which was not explored in study IV.

We found no associations with self-reported heart failure in our analyses. This is a surprising finding since there is a vast literature reporting crackles in the presence heart failure.[55] This could be explained by the fact that heart failure is underdiagnosed in the general population.[126] This relationship should be re-examined with objective measurements of heart function.

6.2.7 Associations with smoking status

We observed that the use of tobacco was related to the presence of wheezes and crackles. In the case of expiratory wheezes only current smoking was significantly associated in the multivariable models while both current and past smoking was associated with inspiratory crackles. Smoking causes damage to the lungs by means of local inflammation, increase of secretions and airway remodeling. Thus, resulting in an accelerated age-related decline of FEV_1, and increased risk of COPD, chronic bronchitis, asthma and a more frequent presence of respiratory symptoms (including wheezing) in general.[127] The association between tobacco smoke and crackles could be explained by two mechanisms. First, due to the increase of secretion in the airways which could intermittently obstruct the distal airways. Second, by the reduction of supportive tissue that results from continuous remodeling of the distal airways making them more prone to collapse during expiration. In the case of wheezes, the smoke of tobacco could play a role by causing inflammation of the airway mucosa[128] and increased contractility of bronchial smooth muscle[129] in acute and sub-acute exposure. This could explain the fact that we only observed a significant relation with current, but not past smoking using expiratory wheezes as outcome.

6.2.8 Associations with FEV_1

We found that the presence of wheezes and inspiratory crackles was significantly associated with a lower FEV_1-Z score. Also, an increase of one unit in the FEV_1 Z-score resulted in a significantly lower probability to have crackles at two or more recording sites. The appearance of crackles may therefore point to a loss of lung function before FEV_1 reaches the lower limit of normal. We found that the age adjusted association of FEV_1 Z-score with inspiratory crackles (OR 0.76; 95%CI 0.69 – 0.84, $p<0.001$) remained unchanged even when removing all the participants (n= 286) with a FEV_1 Z-score under LLN (0.77; 0.67 – 0.87, $p<0.001$). Nevertheless, despite these interesting associations, not all the participants with very low lung function had crackles. It has been hypothesized that COPD can present as several different phenotypes.[(130)] One epidemiological study by Huang et al already suggests that wheezes are associated with different phenotypes of COPD.[(131)] It would be interesting if future research could approach the presence of crackles or wheezes from this perspective and find out whether the presence of AdLS could help the clinician into identifying a certain phenotype of COPD and if relates to prognostic outcomes or influences treatment strategy.

6.2.9 Associations with Oxygen saturation

Low oxygen saturation was significantly associated with the presence of inspiratory crackles. Crackles are related to the sudden opening of closed airways or to air movement through mucus that obstructs airways.[(24)] These conditions are also likely to impair ventilation/perfusion matching, the most common cause of hypoxemia. This could explain the relationship in our study.[(115)] The presence of crackles in the absence of known lung or heart disease could be an indicator of impaired ventilation. One study by Räsanen et al showed a significant decrease in SpO_2 and an increase in the power spectra of lung sounds in the band from 150 to 1200 Hz (which includes the presence of crackles) during acute lung injury.[(132)] In this study they were able to reverse the change in lung sound spectral power and SpO_2 to pre injury levels by applying Positive End-Expiratory Pressure in the ventilator machine.

Based on the epidemiological data obtained in the present studies and physiological/experimental data it is possible to see a relationship between the presence of crackles and the probability of low oxygen saturation. The presence of crackles in an asymptomatic patient should trigger a measurement of SpO_2. More research would be necessary to answer whether this strategy could be of benefit for the patient.

7 Conclusion and future research

I can conclude that the use of lung auscultation has an inter-observer variability that deems to be acceptable in daily clinical praxis and in line with other diagnostic methods. However, it is not an optimal tool for screening of lung diseases due to the low sensitivity of wheezes and crackles to specific diagnostic categories. We found that wheezes and crackles are common findings in adults and its prevalence increases significantly with increasing age. Our findings suggest that having wheezes or crackles in one auscultation site is not worrisome in itself. However, the finding of AdLS in two or more recording sites is more likely to suggest reduced lung function. The findings of AdLS should be analyzed in conjunction with adequate history taking and complete physical examination in order to have a comprehensive clinical picture. It is possible that AdLS are related to chronic changes in the lungs. Thus, the prescription of antibiotics based on the presence of crackles during an acute lower respiratory tract infection might not be advisable, because it is likely that crackles were already in place before the infectious episode, especially in older adults.

We have analyzed how AdLS mainly related to the presence of self-reported disease, and markers of lung function. However, the limitations of our studies leaves some important questions unanswered.

1.- We used a cross-sectional design to stablish the prevalence of AdLS, therefore we cannot say anything about the presence of AdLS over time. Some evidence suggests that the presence of crackles changes during acute exacerbations of COPD and heart failure. A prospective study of patients with these diagnoses could help clarifying whether AdLS have diagnostic or prognostic value in the context of exacerbations.

2.- This work mainly describes the associations with lung diseases. Other studies have found associations between AdLS and heart diseases. Surprisingly, we there were no significant associations between AdLS and heart diseases such as heart failure in our data. Notably, we used self-reported data on heart failure. Further studies could use objective measures of heart function such as echocardiography and NT pro-BNP.

3.- Less than half of the participants with COPD had AdLS. However, COPD seems to be a complex diagnosis with many subtypes of patients. It could be of interest to study the presence of AdLS in relationship to proposed phenotypes of COPD. AdLS could perhaps play a role in further defining different subtypes of patients that might require different types of treatment.

4.- Even though we found associations with the presence of AdLS and diminished lung function, we do not know if further screening such as spirometry or other pulmonary function tests would be beneficial for patients that present AdLS in primary care. Spirometry is a simple diagnostic test and many general practices have access to it. However is not used routinely in every patient with risk factors for obstructive lung disease. The finding of AdLS could be a clinical criterion to trigger a spirometry examination and other tests (for example, SpO_2). However, further studies are necessary to clarify whether this is cost/effective and beneficial for the patient.

5. The use of electronic stethoscopes is increasingly common. Some of these devices offer additional possibilities for interpretation of lung sounds such as frequency and spectral analyses. Also, it is possible to use artificial intelligence to predict a defined outcome (category of AdLS or even a clinical diagnosis). The use of these tools in diagnosis and monitoring lung and heart disease can offer more data points than traditional analog auscultation. Therefore, it would be interesting to investigate whether the use of these tools can provide the physician with a diagnostic or therapeutic advantage compared to traditional auscultation methods.

8 References

1. Merriam-Webster [Internet]. Springfield (MA): Merriam-Webster Incorporated; c2018. "diagnosis" [cited 2018 Oct 16]]. Available from: https://www.merriam-webster.com/dictionary/diagnosis.
2. Jones DS, Podolsky SH, Greene JA. The burden of disease and the changing task of medicine. N Engl J Med. 2012;366(25):2333-8.
3. Borghi L. Tapping on the chest of history. Lost and found memories of Leopold Auenbrugger, inventor of percussion, in Austria and beyond. Acta Med Hist Adriat. 2018;16(1):127-44.
4. Hale-White W. The history of percussion and auscultation. The Lancet. 1924;203(5240):263-6.
5. Weinberg F. The history of the stethoscope. Can Fam Physician. 1993;39:2223-4.
6. Roguin A. Rene Theophile Hyacinthe Laennec (1781-1826): the man behind the stethoscope. Clin Med Res. 2006;4(3):230-5.
7. 3M. Stethoscope History c2018 [Available from: https://www.littmann.com/3M/en_US/littmann-stethoscopes/education-center/history/.
8. Gottlieb ER, Aliotta JM, Tammaro D. Comparison of analogue and electronic stethoscopes for pulmonary auscultation by internal medicine residents. Postgrad Med J. 2018.
9. Gage AJ, Ilombu O, Akinyemi AI. Service readiness, health facility management practices, and delivery care utilization in five states of Nigeria: a cross-sectional analysis. BMC Pregnancy Childbirth. 2016;16(1):297.
10. Edelman ER, Weber BN. Tenuous Tether. N Engl J Med. 2015;373(23):2199-201.
11. Frishman WH. Is the Stethoscope Becoming an Outdated Diagnostic Tool? Am J Med. 2015;128(7):668-9.
12. English T. The Stethoscope: Timeless Tool Or Outdated Relic? 2016 [Available from: http://www.npr.org/sections/health-shots/2016/02/26/467212821/the-stethoscope-timeless-tool-or-outdated-relic.
13. Andres E, Gass R. Stethoscope: A Still-Relevant Tool and Medical Companion. Am J Med. 2016;129(5):e37-8.
14. Research GV. Stethoscope Market Size, Share, Industry Research Report, 2018-2025 2017 [Available from: https://www.grandviewresearch.com/industry-analysis/stethoscopes-market.
15. Robertson AJ, Coope R. Rales, rhonchi, and Laennec. Lancet. 1957;273(6992):417-23.
16. Pasterkamp H, Brand PL, Everard M, Garcia-Marcos L, Melbye H, Priftis KN. Towards the standardisation of lung sound nomenclature. Eur Respir J. 2016;47(3):724-32.
17. Mikami R, Murao M, Cugell DW, Chretien J, Cole P, Meier-Sydow J, et al. International Symposium on Lung Sounds. Synopsis of proceedings. Chest. 1987;92(2):342-5.
18. Bohadana A, Izbicki G, Kraman SS. Fundamentals of lung auscultation. N Engl J Med. 2014;370(8):744-51.
19. Forgacs P. The functional basis of pulmonary sounds. Chest. 1978;73(3):399-405.
20. Pasterkamp H. The highs and lows of wheezing: A review of the most popular adventitious lung sound. Pediatr Pulmonol. 2018;53(2):243-54.
21. Gavriely N, Shee TR, Cugell DW, Grotberg JB. Flutter in flow-limited collapsible tubes: a mechanism for generation of wheezes. J Appl Physiol (1985). 1989;66(5):2251-61.
22. Dines DE, DeRemee RA. Meaningful clues and physical signs in chest disease. Mod Treat. 1970;7(4):821-39.
23. Melbye H, Garcia-Marcos L, Brand P, Everard M, Priftis K, Pasterkamp H. Wheezes, crackles and rhonchi: simplifying description of lung sounds increases the agreement on their classification: a study of 12 physicians' classification of lung sounds from video recordings. BMJ Open Respir Res. 2016;3(1):e000136.

24. Vyshedskiy A, Alhashem RM, Paciej R, Ebril M, Rudman I, Fredberg JJ, et al. Mechanism of inspiratory and expiratory crackles. Chest. 2009;135(1):156-64.
25. Forgacs P. Lung Sounds. London: Baillière Tindall; 1978.
26. Baughman RP, Loudon RG. Stridor: differentiation from asthma or upper airway noise. Am Rev Respir Dis. 1989;139(6):1407-9.
27. Kumar V, Abbas A, Fausto N. Robbins and Cotran Pathologic Basis of Disease. 7th Edition ed. Philadelphia, Pennsylvania, USA. : Elsevier Saunders; 2005.
28. van Vugt SF, Broekhuizen BD, Lammens C, Zuithoff NP, de Jong PA, Coenen S, et al. Use of serum C reactive protein and procalcitonin concentrations in addition to symptoms and signs to predict pneumonia in patients presenting to primary care with acute cough: diagnostic study. BMJ. 2013;346:f2450.
29. Melbye H, Straume B, Aasebo U, Dale K. Diagnosis of pneumonia in adults in general practice. Relative importance of typical symptoms and abnormal chest signs evaluated against a radiographic reference standard. Scand J Prim Health Care. 1992;10(3):226-33.
30. Wipf JE, Lipsky BA, Hirschmann JV, Boyko EJ, Takasugi J, Peugeot RL, et al. Diagnosing pneumonia by physical examination: relevant or relic? Arch Intern Med. 1999;159(10):1082-7.
31. Singal BM, Hedges JR, Radack KL. Decision rules and clinical prediction of pneumonia: evaluation of low-yield criteria. Ann Emerg Med. 1989;18(1):13-20.
32. Schierenberg A, Minnaard MC, Hopstaken RM, van de Pol AC, Broekhuizen BD, de Wit NJ, et al. External Validation of Prediction Models for Pneumonia in Primary Care Patients with Lower Respiratory Tract Infection: An Individual Patient Data Meta-Analysis. PLoS One. 2016;11(2):e0149895.
33. Turnbull S, Lucas PJ, Redmond NM, Christensen H, Thornton H, Cabral C, et al. What gives rise to clinician gut feeling, its influence on management decisions and its prognostic value for children with RTI in primary care: a prospective cohort study. BMC Fam Pract. 2018;19(1):25.
34. Sgalla G, Walsh SLF, Sverzellati N, Fletcher S, Cerri S, Dimitrov B, et al. "Velcro-type" crackles predict specific radiologic features of fibrotic interstitial lung disease. BMC Pulm Med. 2018;18(1):103.
35. Fredberg JJ, Holford SK. Discrete lung sounds: crackles (rales) as stress-relaxation quadrupoles. J Acoust Soc Am. 1983;73(3):1036-46.
36. Cottin V, Cordier JF. Velcro crackles: the key for early diagnosis of idiopathic pulmonary fibrosis? Eur Respir J. 2012;40(3):519-21.
37. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med. 2011;183(6):788-824.
38. Shirai F, Kudoh S, Shibuya A, Sada K, Mikami R. Crackles in asbestos workers: auscultation and lung sound analysis. Br J Dis Chest. 1981;75(4):386-96.
39. Epler GR, Carrington CB, Gaensler EA. Crackles (rales) in the interstitial pulmonary diseases. Chest. 1978;73(3):333-9.
40. Melbye H. [The usefulness of clinical pulmonary examination in the diagnosis of bronchial obstruction]. Tidsskr Nor Laegeforen. 1998;118(13):1999-2003.
41. van Schayck CP, van Weel C, Harbers HJ, van Herwaarden CL. Do physical signs reflect the degree of airflow obstruction in patients with asthma or chronic obstructive pulmonary disease? Scand J Prim Health Care. 1991;9(4):232-8.
42. Leuppi JD, Dieterle T, Wildeisen I, Martina B, Tamm M, Koch G, et al. Can airway obstruction be estimated by lung auscultation in an emergency room setting? Respir Med. 2006;100(2):279-85.
43. MacNee W. ABC of Chronic Obstructive Pulmonary Disease: Pathology, pathogenesis, and pathophysiology. BMJ. 2006;332(7551):1202-4.
44. Murphy RL. In defense of the stethoscope. Respir Care. 2008;53(3):355-69.
45. Simel DL, D. R. The Rational Clinical Examination, Evidence-Based Clinical Diagnosis. New York: McGraw-Hill; 2009. 149-57 p.

46. Holleman DR, Jr., Simel DL, Goldberg JS. Diagnosis of obstructive airways disease from the clinical examination. J Gen Intern Med. 1993;8(2):63-8.
47. Jacome C, Oliveira A, Marques A. Computerized respiratory sounds: a comparison between patients with stable and exacerbated COPD. Clin Respir J. 2017;11(5):612-20.
48. 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):26978-96.
49. Shim CS, Williams MH, Jr. Relationship of wheezing to the severity of obstruction in asthma. Arch Intern Med. 1983;143(5):890-2.
50. Baumann UA, Haerdi E, Keller R. Relations between clinical signs and lung function in bronchial asthma: how is acute bronchial obstruction reflected in dyspnoea and wheezing? Respiration. 1986;50(4):294-300.
51. (2018) GIfA. Global Strategy for Asthma Management and Prevention [Available from: http://ginasthma.org.
52. Oshaug K, Halvorsen PA, Melbye H. Should chest examination be reinstated in the early diagnosis of chronic obstructive pulmonary disease? Int J Chron Obstruct Pulmon Dis. 2013;8:369-77.
53. Earis JE, Marsh K, Pearson MG, Ogilvie CM. The inspiratory "squawk" in extrinsic allergic alveolitis and other pulmonary fibroses. Thorax. 1982;37(12):923-6.
54. Paciej R, Vyshedskiy A, Bana D, Murphy R. Squawks in pneumonia. Thorax. 2004;59(2):177-8.
55. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail. 2016;18(8):891-975.
56. Hobbs FD, Korewicki J, Cleland JG, Eastaugh J, Freemantle N, Investigators I. The diagnosis of heart failure in European primary care: The IMPROVEMENT Programme survey of perception and practice. Eur J Heart Fail. 2005;7(5):768-79.
57. Damy T, Kallvikbacka-Bennett A, Zhang J, Goode K, Buga L, Hobkirk J, et al. Does the physical examination still have a role in patients with suspected heart failure? Eur J Heart Fail. 2011;13(12):1340-8.
58. Rihal CS, Davis KB, Kennedy JW, Gersh BJ. The utility of clinical, electrocardiographic, and roentgenographic variables in the prediction of left ventricular function. Am J Cardiol. 1995;75(4):220-3.
59. Thomas JT, Kelly RF, Thomas SJ, Stamos TD, Albasha K, Parrillo JE, et al. Utility of history, physical examination, electrocardiogram, and chest radiograph for differentiating normal from decreased systolic function in patients with heart failure. Am J Med. 2002;112(6):437-45.
60. Martin-Pfitzenmeyer I, Gauthier S, Bailly M, Loi N, Popitean L, d'Athis P, et al. Prognostic factors in stage D heart failure in the very elderly. Gerontology. 2009;55(6):719-26.
61. Guo L, Li G, Wang Y, Liang H, Shan X, Zhang N, et al. Diagnostic utility of N-terminal-proBNP in differentiating acute pulmonary embolism from heart failure in patients with acute dyspnea. Chin Med J (Engl). 2014;127(16):2888-93.
62. Jimeno Sainz A, Gil V, Merino J, Garcia M, Jordan A, Guerrero L. [Validity of Framingham criteria as a clinical test for systolic heart failure]. Rev Clin Esp. 2006;206(10):495-8.
63. Cowie MR, Wood DA, Coats AJ, Thompson SG, Suresh V, Poole-Wilson PA, et al. Survival of patients with a new diagnosis of heart failure: a population based study. Heart. 2000;83(5):505-10.
64. Anderson KM. Discharge clinical characteristics and 60-day readmission in patients hospitalized with heart failure. J Cardiovasc Nurs. 2014;29(3):232-41.
65. Oliveira A, Marques A. Respiratory sounds in healthy people: a systematic review. Respir Med. 2014;108(4):550-70.
66. Thacker RE, Kraman SS. The prevalence of auscultatory crackles in subjects without lung disease. Chest. 1982;81(6):672-4.

67. Workum P, Holford SK, Delbono EA, Murphy RL. The prevalence and character of crackles (rales) in young women without significant lung disease. Am Rev Respir Dis. 1982;126(5):921-3.
68. Spiteri MA, Cook DG, Clarke SW. Reliability of eliciting physical signs in examination of the chest. Lancet. 1988;1(8590):873-5.
69. Mulrow CD, Dolmatch BL, Delong ER, Feussner JR, Benyunes MC, Dietz JL, et al. Observer variability in the pulmonary examination. J Gen Intern Med. 1986;1(6):364-7.
70. Margolis PA, Ferkol TW, Marsocci S, Super DM, Keyes LL, McNutt R, et al. Accuracy of the clinical examination in detecting hypoxemia in infants with respiratory illness. J Pediatr. 1994;124(4):552-60.
71. Elphick HE, Lancaster GA, Solis A, Majumdar A, Gupta R, Smyth RL. Validity and reliability of acoustic analysis of respiratory sounds in infants. Arch Dis Child. 2004;89(11):1059-63.
72. Brooks D, Thomas J. Interrater reliability of auscultation of breath sounds among physical therapists. Phys Ther. 1995;75(12):1082-8.
73. Badgett RG, Tanaka DJ, Hunt DK, Jelley MJ, Feinberg LE, Steiner JF, et al. The clinical evaluation for diagnosing obstructive airways disease in high-risk patients. Chest. 1994;106(5):1427-31.
74. Badgett RG, Tanaka DJ, Hunt DK, Jelley MJ, Feinberg LE, Steiner JF, et al. Can moderate chronic obstructive pulmonary disease be diagnosed by historical and physical findings alone? Am J Med. 1993;94(2):188-96.
75. Anderson K, Qiu Y, Whittaker AR, Lucas M. Breath sounds, asthma, and the mobile phone. Lancet. 2001;358(9290):1343-4.
76. Spectrogram: Wikipedia; [Available from: https://en.wikipedia.org/wiki/Spectrogram.
77. Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njolstad I. Cohort profile: The Tromso Study. Int J Epidemiol. 2012;41(4):961-7.
78. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. Eur Respir J. 2012;40(6):1324-43.
79. G. D, F. N, C. R. Macleod's Clinical Examination. 13th edition ed. London: Churchill-Livingston; 2013.
80. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.
81. Cohen J. A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas. 1960;20(1):37-46.
82. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.
83. Vanbelle S. Agreement between raters and groups of raters. Liege, Belgium: Universite de Liege; 2009.
84. Kraemer HC, Periyakoil VS, Noda A. Kappa coefficients in medical research. Stat Med. 2002;21(14):2109-29.
85. Vanbelle S. Asymptotic variability of (multilevel) multirater kappa coefficients. Statistical Methods in Medical Research.0(0):0962280218794733.
86. Taiyun W, Simko V. R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from https://github.com/taiyun/corrplot. 2017.
87. Vanbelle S. Comparing dependent kappa coefficients obtained on multilevel data. Biom J. 2017.
88. Vanbelle S. R package "multiagree": Comparison of dependent kappa coefficients. 2017.
89. Fleiss J. Design and analysis of clinical experiments. New York: John Wiley & Sons; 1986.
90. Statistics Norway. Tromsø municipality facts (Age distribution per 1. januar 2018) [Internet]. Statistics Norway; 2018 [Available from: https://www.ssb.no/kommunefakta/tromso.
91. Dalbak LG, Straand J, Melbye H. Should pulse oximetry be included in GPs' assessment of patients with obstructive lung disease? Scand J Prim Health Care. 2015;33(4):305-10.

92. Global Initiative for Prevention Diagnosis and Management of COPD (2018). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease [Available from: https://goldcopd.org/wp-content/uploads/2017/11/GOLD-2018-v6.0-FINAL-revised-20-Nov_WMS.pdf.
93. Xavier Robin NT, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez and Markus Müller. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
94. Fox J, Weisberg S. An {R} Companion to Applied Regression. 2nd ed. Thousand Oaks, California, USA: SAGE Publications; 2011.
95. Harris AD, McGregor JC, Perencevich EN, Furuno JP, Zhu J, Peterson DE, et al. The use and interpretation of quasi-experimental studies in medical informatics. J Am Med Inform Assoc. 2006;13(1):16-23.
96. Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. Occup Environ Med. 1998;55(10):651-6.
97. Lam HC, Chan EY, Goggins WB, 3rd. Comparison of short-term associations with meteorological variables between COPD and pneumonia hospitalization among the elderly in Hong Kong-a time-series study. Int J Biometeorol. 2018;62(8):1447-60.
98. Lam HC, Li AM, Chan EY, Goggins WB, 3rd. The short-term association between asthma hospitalisations, ambient temperature, other meteorological factors and air pollutants in Hong Kong: a time-series study. Thorax. 2016;71(12):1097-109.
99. Osborne NJ, Alcock I, Wheeler BW, Hajat S, Sarran C, Clewlow Y, et al. Pollen exposure and hospitalization due to asthma exacerbations: daily time series in a European city. Int J Biometeorol. 2017;61(10):1837-48.
100. DeVries R, Kriebel D, Sama S. Low level air pollution and exacerbation of existing copd: a case crossover analysis. Environ Health. 2016;15(1):98.
101. DeVries R, Kriebel D, Sama S. Outdoor Air Pollution and COPD-Related Emergency Department Visits, Hospital Admissions, and Mortality: A Meta-Analysis. COPD. 2017;14(1):113-21.
102. Metersky ML, Masterton RG, Lode H, File TM, Jr., Babinchak T. Epidemiology, microbiology, and treatment considerations for bacterial pneumonia complicating influenza. Int J Infect Dis. 2012;16(5):e321-31.
103. Wesseling G. Occasional review: influenza in COPD: pathogenesis, prevention, and treatment. Int J Chron Obstruct Pulmon Dis. 2007;2(1):5-10.
104. Veloso SG, Lima MF, Salles PG, Berenstein CK, Scalon JD, Bambirra EA. Interobserver agreement of Gleason score and modified Gleason score in needle biopsy and in surgical specimen of prostate cancer. Int Braz J Urol. 2007;33(5):639-46; discussion 47-51.
105. Timmers JM, van Doorne-Nagtegaal HJ, Verbeek AL, den Heeten GJ, Broeders MJ. A dedicated BI-RADS training programme: effect on the inter-observer variation among screening radiologists. Eur J Radiol. 2012;81(9):2184-8.
106. Brosnan M, La Gerche A, Kumar S, Lo W, Kalman J, Prior D. Modest agreement in ECG interpretation limits the application of ECG screening in young athletes. Heart Rhythm. 2015;12(1):130-6.
107. Wang EE, Milner RA, Navas L, Maj H. Observer agreement for respiratory signs and oximetry in infants hospitalized with lower respiratory infections. Am Rev Respir Dis. 1992;145(1):106-9.
108. Halford JJ, Shiau D, Desrochers JA, Kolls BJ, Dean BC, Waters CG, et al. Inter-rater agreement on identification of electrographic seizures and periodic discharges in ICU EEG recordings. Clin Neurophysiol. 2015;126(9):1661-9.
109. Zhou S, Zha Y, Wang C, Wu J, Liu W, Liu B. [The clinical value of bedside lung ultrasound in the diagnosis of chronic obstructive pulmonary disease and cardiac pulmonary edema]. Zhonghua Wei Zhong Bing Ji Jiu Yi Xue. 2014;26(8):558-62.
110. Morris R, MacNeela P, Scott A, Treacy P, Hyde A, O'Brien J, et al. Ambiguities and conflicting results: The limitations of the kappa statistic in establishing the interrater reliability of the Irish nursing minimum data set for mental health: A discussion paper. Int J Nurs Stud. 2008;45(4):645-7.

111. Shrout PE. Measurement reliability and agreement in psychiatry. Stat Methods Med Res. 1998;7(3):301-17.
112. Bland JM, Altman DG. A Note on the Use of the Intraclass Correlation-Coefficient in the Evaluation of Agreement between 2 Methods of Measurement. Computers in Biology and Medicine. 1990;20(5):337-40.
113. de Vet HC, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's kappa. BMJ. 2013;346:f2125.
114. Jacome C, Marques A. Computerized Respiratory Sounds Are a Reliable Marker in Subjects With COPD. Respir Care. 2015;60(9):1264-75.
115. Janssens JP, Pache JC, Nicod LP. Physiological changes in respiratory function associated with ageing. Eur Respir J. 1999;13(1):197-205.
116. Bergeron C, Tulic MK, Hamid Q. Airway remodelling in asthma: from benchside to clinical practice. Can Respir J. 2010;17(4):e85-93.
117. Jakobsen KA, Melbye H, Kelly MJ, Ceynowa C, Molstad S, Hood K, et al. Influence of CRP testing and clinical findings on antibiotic prescribing in adults presenting with acute cough in primary care. Scand J Prim Health Care. 2010;28(4):229-36.
118. Bongartz T, Nannini C, Medina-Velasquez YF, Achenbach SJ, Crowson CS, Ryu JH, et al. Incidence and mortality of interstitial lung disease in rheumatoid arthritis: a population-based study. Arthritis Rheum. 2010;62(6):1583-91.
119. Nannini C, Medina-Velasquez YF, Achenbach SJ, Crowson CS, Ryu JH, Vassallo R, et al. Incidence and mortality of obstructive lung disease in rheumatoid arthritis: a population-based study. Arthritis Care Res (Hoboken). 2013;65(8):1243-50.
120. Shen TC, Lin CL, Chen CH, Tu CY, Hsia TC, Shih CM, et al. Increased risk of chronic obstructive pulmonary disease in patients with rheumatoid arthritis: a population-based cohort study. QJM. 2014;107(7):537-43.
121. Wilsher M, Voight L, Milne D, Teh M, Good N, Kolbe J, et al. Prevalence of airway and parenchymal abnormalities in newly diagnosed rheumatoid arthritis. Respir Med. 2012;106(10):1441-6.
122. Global Initiative for Prevention Diagnosis and Management of COPD (2018). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. 2018.
123. Chang K, Yang SM, Kim SH, Han KH, Park SJ, Shin JI. Smoking and rheumatoid arthritis. Int J Mol Sci. 2014;15(12):22279-95.
124. Karnani NG, Reisfield GM, Wilson GR. Evaluation of chronic dyspnea. Am Fam Physician. 2005;71(8):1529-37.
125. Santos M, Kitzman DW, Matsushita K, Loehr L, Sueta CA, Shah AM. Prognostic Importance of Dyspnea for Cardiovascular Outcomes and Mortality in Persons without Prevalent Cardiopulmonary Disease: The Atherosclerosis Risk in Communities Study. PLoS One. 2016;11(10):e0165111.
126. van Riet EE, Hoes AW, Limburg A, Landman MA, van der Hoeven H, Rutten FH. Prevalence of unrecognized heart failure in older persons with shortness of breath on exertion. Eur J Heart Fail. 2014;16(7):772-7.
127. Centers for Disease Control and Prevention (US); National Center for Chronic Disease Prevention and Health Promotion (US); Office on Smoking and Health (US). How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General. Atlanta (GA): Centers for Disease Control and Prevention (US); 2010 [29 Dec 2018]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK53021/.
128. Bhalla DK, Hirata F, Rishi AK, Gairola CG. Cigarette smoke, inflammation, and lung injury: a mechanistic perspective. J Toxicol Environ Health B Crit Rev. 2009;12(1):45-64.
129. Chiba Y, Murata M, Ushikubo H, Yoshikawa Y, Saitoh A, Sakai H, et al. Effect of cigarette smoke exposure in vivo on bronchial smooth muscle contractility in vitro in rats. Am J Respir Cell Mol Biol. 2005;33(6):574-81.
130. Mirza S, Benzo R. Chronic Obstructive Pulmonary Disease Phenotypes: Implications for Care. Mayo Clin Proc. 2017;92(7):1104-12.

131. Huang WC, Tsai YH, Wei YF, Kuo PH, Tao CW, Cheng SL, et al. Wheezing, a significant clinical phenotype of COPD: experience from the Taiwan Obstructive Lung Disease Study. Int J Chron Obstruct Pulmon Dis. 2015;10:2121-6.
132. Rasanen J, Nemergut ME, Gavriely N. Effect of PEEP on breath sound power spectra in experimental lung injury. Intensive Care Med Exp. 2014;2(1):25.

Paper I

BMJ Open Respiratory Research

International perception of lung sounds: a comparison of classification across some European borders

Juan Carlos Aviles-Solis,[1] Sophie Vanbelle,[2] Peder A Halvorsen,[1] Nick Francis,[3] Jochen W L Cals,[4] Elena A Andreeva,[5] Alda Marques,[6] Päivi Piirilä,[7] Hans Pasterkamp,[8] Hasse Melbye[1]

To cite: Aviles-Solis JC, Vanbelle S, Halvorsen PA, *et al*. International perception of lung sounds: a comparison of classification across some European borders. *BMJ Open Resp Res* 2017;**4**:e000250. doi:10.1136/bmjresp-2017-000250

Received 2 October 2017
Revised 20 November 2017
Accepted 22 November 2017

ABSTRACT

Introduction Lung auscultation is helpful in the diagnosis of lung and heart diseases; however, the diagnostic value of lung sounds may be questioned due to interobserver variation. This situation may also impair clinical research in this area to generate evidence-based knowledge about the role that chest auscultation has in a modern clinical setting. The recording and visual display of lung sounds is a method that is both repeatable and feasible to use in large samples, and the aim of this study was to evaluate interobserver agreement using this method.

Methods With a microphone in a stethoscope tube, we collected digital recordings of lung sounds from six sites on the chest surface in 20 subjects aged 40 years or older with and without lung and heart diseases. A total of 120 recordings and their spectrograms were independently classified by 28 observers from seven different countries. We employed absolute agreement and kappa coefficients to explore interobserver agreement in classifying crackles and wheezes within and between subgroups of four observers.

Results When evaluating agreement on crackles (inspiratory or expiratory) in each subgroup, observers agreed on between 65% and 87% of the cases. Conger's kappa ranged from 0.20 to 0.58 and four out of seven groups reached a kappa of ≥0.49. In the classification of wheezes, we observed a probability of agreement between 69% and 99.6% and kappa values from 0.09 to 0.97. Four out of seven groups reached a kappa ≥0.62.

Conclusions The kappa values we observed in our study ranged widely but, when addressing its limitations, we find the method of recording and presenting lung sounds with spectrograms sufficient for both clinic and research. Standardisation of terminology across countries would improve international communication on lung auscultation findings.

Key messages

- We found variation in the level of agreement when clinicians classify lung sounds.
- Digital recordings with the use of spectrograms is a method suitable for research of lung sounds.
- Standardisation of the terminology of lung sounds would improve internatonal communication on the subject.

INTRODUCTION

Lung auscultation is an old and well-known technique in clinical medicine. Adventitious lung sounds, such as wheezes and crackles, are helpful in the diagnosis of several lung and heart-related conditions.[1–5] However, the diagnostic value of chest auscultation may be questioned due to variability in recognising lung sounds.[6–8] In a scale from 0 to 1, a study by Spiteri *et al* found a kappa of κ=0.41 for crackles and κ=0.51 for wheezes when clinicians classified lung sounds.[9] Similar results have been found in other studies.[10–14] Lower agreement levels have also been found.[7 15]

However, most of these agreement measures were based on clinicians sequentially listening to patients with a stethoscope. Clinicians working in the same hospital department have rated the sounds in these studies making the sample homogeneous and applicability of the results may be questioned.[1 11–14] In addition, the use of such methods would be difficult to implement in large epidemiological studies due to logistical challenges. New methods are needed for clinical research in this area to generate evidence-based knowledge about the role that lung sounds have in a modern clinical setting.

Studies of interobserver agreement using lung sound recordings, rather than traditional auscultation, may be a good alternative.[15–17] Recorded sounds may be presented with a visual display, and creating spectrograms of lung sounds is already an option in the software of electronic stethoscopes. Recording and visual display of lung sounds may be applied in large samples and classifications of the sounds may be repeated. However, we still do not know the reliability of such classifications.

The aim of the present study was to describe the interobserver agreement among an international sample of raters, including general

For numbered affiliations see end of article.

Correspondence to
Dr Juan Carlos Aviles-Solis; juan.c.solis@uit.no

practitioners (GP), pulmonologists and medical students, when classifying lung sounds in adults aged 40 years or older using audio recordings with display of spectrograms.

METHODS

In August to October 2014 we conducted a cross-sectional study to explore agreement in the classification of lung sounds. In order to obtain material to classify, we recruited a convenience sample of 20 subjects aged 40 years or older. We took contact with a rehabilitation programme in northern Norway for patients with heart and lung-related diseases (lung cancer, chronic obstructive pulmonary disease, heart failure, and so on). We got permission to hold a presentation about lung sounds and at the end of the presentation we invited the patients to be part of our research project as subjects. Fourteen patients attending the rehabilitation programme agreed to participate and we recorded the lung sounds that same evening. The patients were 67.43 years old on average (44–84) and nine were female. To hold a balanced sample (concerning the prevalence of wheezes, crackles and normal lung sounds), we obtained the rest of our recordings from six self-reported healthy employees at our university aged 51.83 years old on average (46–67) and five were female. We registered the following information about the subjects: age, gender and self-reported history of heart or lung disease. No personal information was registered that could link the sound recordings to the individual subjects.

Recording of lung sounds

To record the lung sounds, we used a microphone MKE 2-EW with a wireless system EW 112-P G3-G (Sennheiser electronic, Wedemark, Germany) placed in the tube of a Littmann Master Classic II stethoscope (3M, Maplewood, MN, USA) at a distance of 10 cm from the headpiece. The microphone was connected to a digital sound Handy recorder H4n (Zoom, Tokyo, Japan).

We placed the membrane of the stethoscope against the naked thorax of the subjects. We asked the subjects to breathe deeply while keeping their mouth open. We started the recording with an inspiration and continued for approximately 20 s trying to capture three full respiratory cycles with good quality sound. We performed this same procedure at six different locations (figure 1). The researcher collecting recordings used a headphone as an audio monitor to evaluate the quality of the recording. When too much noise or cough was heard during the recording, a second attempt was performed.

We obtained a total of 120 audio files. The audio files were in '.wav' format and recorded at a sample rate of 44 100 Hz and 16 bit depth in a single monophonic channel. We did not perform postprocessing of the sound files or implement filters.

Presentation of the sounds

One researcher (HM) selected the sections with less noise according to his acoustic perception. Breathing phases were determined by listening to the recordings (which usually started with inspiration) and visual analysis of the spectrograms. A spectrogram for each of these recordings was created using Adobe Audition V.5.0 (Adobe Systems, San Jose, CA, USA) (figure 1). The spectrograms showed time on the x-axis, frequency on the y-axis and intensity by colour saturation. Videos of the selected spectrograms, where an indicator bar follows the sound, were made from the computer screen using Camtasia Studio V.8 software (TechSmith, Okemos, MI, USA). We compiled these 120 videos of lung sounds in a PowerPoint presentation (Microsoft, Redmond, WA, USA). Age, gender and recording location, but no clinical information, were presented about the subjects. The majority of the recordings started during inspiration and if that was not the case, this was specified.

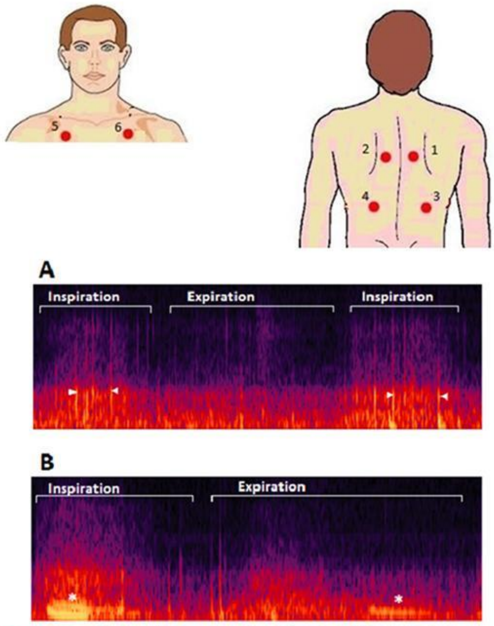


Figure 1 Upper: illustration showing the different places where lung sounds were recorded. (1_2) Between the spine and the medial border of the scapula at the level of T4–T5; (3_4) at the middle point between the spine and the mid-axillary line at the level of T9–T10; (5_6) at the intersection of the mid-clavicular line and second intercostal space. Lower: image showing two different spectrograms containing crackles (A) and wheezes (B). Crackles appear as vertical lines (arrowheads) and wheezes as horizontal lines (*).

Recruitment of the raters and classification of the files

We recruited seven groups of four raters to classify the 120 recordings: We wanted a heterogeneous sample,

Table 1 Prevalence, probability of agreement, Conger's kappa (SE) and 95% CI for the seven groups of observers when classifying 120 sound files for the presence of crackles and wheezes

	Prevalence	P (agree)	Kappa	SE (kappa)	95% CI
Crackles					
Experts	0.21	0.86	0.56	0.080	0.40 to 0.72
GP Norway	0.23	0.85	0.58	0.083	0.42 to 0.74
GP Russia	0.31	0.65	0.20	0.051	0.10 to 0.30
GP UK	0.17	0.87	0.53	0.089	0.36 to 0.70
GP Netherlands	0.17	0.86	0.49	0.105	0.28 to 0.70
Students	0.27	0.76	0.40	0.086	0.23 to 0.57
Pulmonologists	0.29	0.74	0.37	0.082	0.21 to 0.53
Wheezes					
Experts	0.079	0.96	0.75	0.125	0.51 to 1
GP Norway	0.083	0.94	0.62	0.163	0.30 to 0.94
GP Russia	0.22	0.69	0.09	0.076	–0.06 to 0.24
GP UK	0.065	0.99	0.97	0.024	0.92 to 1.00
GP Netherlands	0.050	0.94	0.39	0.087	0.22 to 0.56
Students	0.073	0.95	0.66	0.042	0.58 to 0.74
Pulmonologists	0.14	0.82	0.27	0.102	0.07 to 0.47

GP, general practitioner.

therefore we included GPs from the Netherlands, Wales, Russia, and Norway, pulmonologists working at the University Hospital of North Norway, an international group of experts (researchers) in the field of lung sounds (Pasterkamp H, Piirila P, Sovijärvi A, Marques A) and sixth year medical students at the Faculty of Health Sciences at UiT, The Arctic University of Norway. We chose to have four raters in each group for pairwise comparisons. The mean age of the groups of raters varied between 25 (the students) and 59 years (the lung sound researchers), and years of experience from 0 (the students) to 28.5 (the lung sound researchers).

All the 28 observers independently classified the 120 recordings. We first asked the observers to classify the lung sounds as normal or abnormal. If abnormal, they had to further classify them as containing crackles, wheezes or other abnormal sounds. It was possible to mark more than one option. The observers specified whether the abnormalities occurred in inspiration or expiration. In addition, they could mark if there was noise present in the recording. We offered two options for answering the survey: an electronic form in Microsoft Access (Microsoft), and a printed version of the questionnaire. We did not perform training of the raters. To make the raters familiar with sounds and spectrograms, the PowerPoint presentation with the 120 recordings started with a demonstration of the three examples, one with normal lung sounds, one with crackles and one with wheezes. The raters were free to play the videos (containing the sound recording and the spectrogram simultaneously) several times and to go back and forth through the cases ad libitum. We used English language in the presentation of the videos and the survey forms. In Russia and the Netherlands, observers were offered translations of the terms included in the survey. These translations were taken from previous studies using lung sound terminology.[18 19]

Statistical analysis

We calculated the probability of agreement and multirater Conger's kappa using the delta method for the analysis of multilevel data.[20] Conger's kappa coefficient was chosen over Fleiss' kappa because the observers classifying the sounds were the same for all sounds. We analysed the intragroup agreement in each of the seven groups of observers when classifying the recordings for the presence of wheezes and crackles disregarding the breathing phase. We used the statistical software 'R' V.3.2.1 together with the package 'multiagree' for the statistical analysis of kappa statistics.[21]

In order to permit the comparison of the agreement levels between and within groups, within and between-group agreement levels were summarised in a matrix, where the diagonal elements represent the mean agreement level between all possible pairs formed by two observers in the same group, and the off-diagonal elements represent the mean agreement level between all possible pairs with one observer in one group and the second observer of the pair in another group. This information was summarised in correlograms using the R package 'Corrplot'.[22]

This study has been reported according to the Guidelines for Reporting Reliability and Agreement Studies.[23]

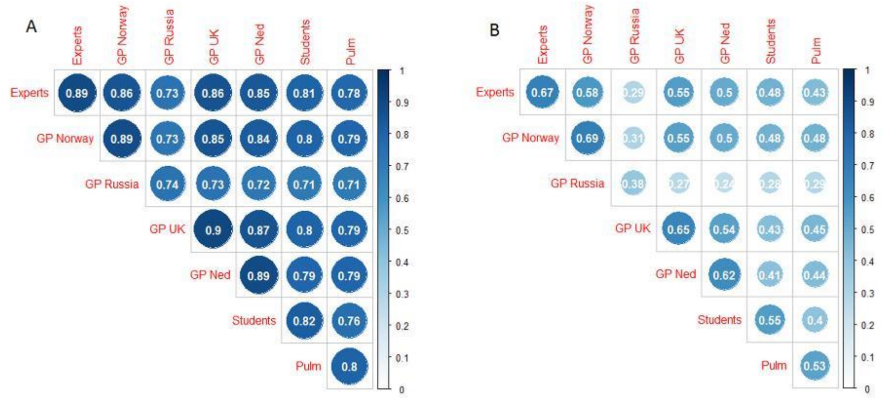


Figure 2 Average proportion of agreement (A) and kappa (B) between pairs of raters from the same (diagonal) and different (off-diagonal) groups when classifying the sounds for the presence of crackles. GP, general practitioner.

RESULTS

Prevalence of wheezes and crackles

All the 28 observers independently classified the 120 recordings. According to the experts' classification, crackles were present in 21% of the 120 recordings and wheezes in 7.9%. Per case (n=20), 15% of the individuals had wheezes, and 50% had crackles in one or more recordings. The prevalence of crackles and wheezes in the 120 recordings varied between groups with mean values among the four observers of 17.0%–29% for crackles and 5.0%–22% for wheezes (table 1). The group average noise reporting ranged from 1.46% to 17.70% (mean=7.5%) of the recordings. There was no significant correlation between the use of this variable and agreement or kappa coefficients. The groups with the highest level of agreement tended to use this variable more often.

Interobserver agreement within the same group

When evaluating interobserver agreement on crackles (inspiratory or expiratory) in each subgroup, observers agreed on between 65% and 87% of the cases. Conger's kappa ranged from 0.20 to 0.58 (table 1) and four out of seven groups reached a kappa of ≥0.49 (median). In the classification of wheezes, we observed a probability of agreement between 69% and 99.6% and kappa values from 0.09 to 0.97 (table 1). Four out of seven groups reached a kappa ≥0.62 (median).

Interobserver agreement between different groups

Lower range probability agreement (<0.8 for crackles and <0.9 for wheezes) within a group was associated with a lower range probability agreement with members of

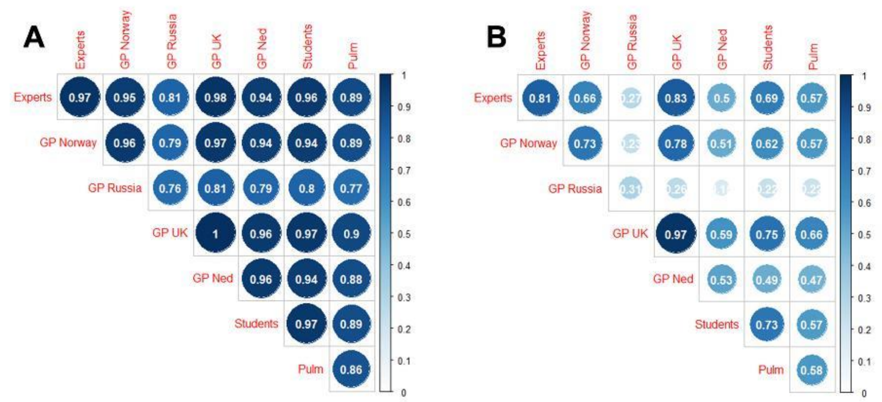


Figure 3 Average proportion of agreement (A) and kappa (B) between pairs of raters from the same (diagonal) and different (off-diagonal) groups when classifying the sounds for the presence of wheezes. GP, general practitioner.

other groups. Correspondingly, high agreement within a group was associated with high agreement with members of other groups (figures 2 and 3).

In particular, the probability of agreement between GPs and the experts was very similar to the probability of agreement within the group of experts (0.86 for crackles and 0.96 for wheezes), except for the group of Russian GPs. Students agreed slightly less with the experts (0.81 for crackles and 0.96 for wheeze) while pulmonologists showed even lower agreement levels with the experts (0.78 for crackles and 0.89 for wheeze). Similar conclusions can be drawn according to Cohen's kappa coefficient values (figures 2 and 3).

DISCUSSION

This study showed a median kappa agreement of 0.49 for crackles and 0.62 for wheezes in the observer groups. Even though kappa coefficients are not directly comparable, our results are similar to those found in other studies analysing interobserver agreement when classifying for wheezes,[10–12] crackles[16] or for both.[6 7 9 13–15 24] The kappa agreements we found were not inferior to those found for other widely accepted clinical examinations.[2 25–29]

In our study, when the agreement levels between clinicians from the same country were in a higher range, we also found a higher level of agreement with members of other groups and vice versa. This finding argues for a general understanding across groups about how to classify crackles and wheezes with some groups encountering greater difficulty in uniform classification.

We found the highest levels of agreement within the experts and some groups of GPs. GPs might be more familiar with the use of lung auscultation, since information from chest imaging, advanced lung function testing or blood gas analysis is not available. Also, GPs are more used to listening to normal lung sounds and sounds with discrete abnormalities. This may have been reflected in the similar levels of agreement between GPs from UK and Norway and the experts in this study.

Strengths and limitations

It was a strength of our study that we included a group of experienced lung sound researchers. They represent recommended use of terminology, and comparison with their classifications may be enlightening, although they were not used as a reference standard.

A strength of our study was also the heterogeneity of the observers in terms of clinical background, experience and country of residency. We believe this gives us a better external validity than if we had included a homogenous sample. However, this factor also presented some challenges concerning language and terminology, which was a weakness of the study.

Different use of lung sound terminology may influence the interobserver agreement.[24 30] The group of Russian GPs had a lower intragroup and intergroup agreement. We think this situation might be partly explained by confusion around the terminology. Anecdotally, we note that the Russian GPs were familiar with a terminology for lung sounds similar to the classic terminology of Laennec, which offers more options than the simple distinction between wheezes and crackles.[31] A higher agreement within the group and with the experts would probably be found if the study had been based on their own terminology. A similar problem was present in the Dutch sample, where the observers found it difficult to classify what they call 'rhonchi' as wheezes or crackles and used the variable 'other abnormal sounds' more frequently than the other groups. In contrast, a terminology restricted to wheezes and crackles is used in UK and Norway, and this has probably made it easier to obtain higher agreements in these countries.

We did not present audiological definitions of crackles and wheezes.[32] As indicated by the Russian and Dutch classifications, the example sounds and the translations to own language did not quite remove the terminology problems. However, clinicians are not familiar with audiological definitions, and we do not think such definitions would have been helpful.

Implications for research

For future research, it is important to be aware that it might be difficult to reach high kappa values when the prevalence of the trait of study is very low or very high, even though absolute agreement may be high.[33 34] This has probably had little impact on the kappa coefficients we observed, since the prevalence of crackles and wheezes was 21% and 7.9%, respectively. However, much lower prevalence of adventitious lung sounds could be found in real epidemiological data. Accordingly, specific measures should be implemented when using this method in epidemiological studies in order to improve its reliability such as training of raters, consensus agreement, multiple independent observations and standardisation of the terminology.[35]

Conclusion

The strength of agreement and correspondingly kappa values were wide ranging, and some groups found it more challenging to produce uniformity in breath sound classification than others. Although the technology was through our experience found to be quite suitable for research, standardisation of terminology across countries with supportive training could improve international communication on lung auscultation findings.

Author affiliations
[1]General Practice Research Unit, Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway
[2]Department of Methodology and Statistics, University of Maastricht, Maastricht, The Netherlands
[3]Department of Primary Care and Public Health, Cardiff University, Cardiff, UK
[4]Department of Family Medicine, CAPHRI Care and Public Health Research Institute, Maastricht University, Maastricht, The Netherlands

[5]Department of Family Medicine, Northern State Medical University (NSMU), Arkhangelsk, Russia
[6]Lab 3R–Respiratory Research and Rehabilitation Laboratory, School of Health Sciences (ESSUA) and Institute for Research in Biomedicine (iBiMED), University of Aveiro, Aveiro, Portugal
[7]Unit of Clinical Physiology, HUS Medical Imaging Center, Helsinki University Hospital and University of Helsinki, Helsinki, Finland
[8]Department of Pediatrics and Child Health, University of Manitoba College of Medicine, Winnipeg, Manitoba, Canada

Acknowledgements The authors thank Professor A Sovijärvi as well as all the other raters involved in this study for their help in the classifications. The authors also thank LHL rehabilitation centre at Skibotn, Norway, for its cooperation to perform the recordings.

Contributors JCAS: analysis of data, data gathering, main responsibility for writing the manuscript. SV: design of the data analysis, analysis of data, substantial contributions to the final manuscript. PAH, EAA, NF, JWLC: data gathering, classification of sounds, substantial contributions to the final manuscript. AM, PP, HP: classification of sounds, substantial contributions to the final manuscript. HM: data collection, study design, substantial contributions to the final manuscript.

Funding General Practice Research Unit, Department of Community Medicine, UiT The Arctic University of Norway. The publication charges for this article have been funded by a grant from the publication fund of UiT The Arctic University of Norway.

Competing interests None declared.

Ethics approval The project was presented to the Regional Committee for Medical and Health Research Ethics, and it was considered to be outside the remit of the Act on Medical and Health Research.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data of the study is available.

REFERENCES

1. Badgett RG, Lucey CR, Mulrow CD. Can the clinical examination diagnose left-sided heart failure in adults? *JAMA* 1997;277:1712–9.
2. Wang CS, FitzGerald JM, Schulzer M, *et al*. Does this dyspneic patient in the emergency department have congestive heart failure? *JAMA* 2005;294:1944–56.
3. Metlay JP, Kapoor WN, Fine MJ. Does this patient have community-acquired pneumonia? Diagnosing pneumonia by history and physical examination. *JAMA* 1997;278:1440–5.
4. Holleman DR, Simel DL. Does the clinical examination predict airflow limitation? *JAMA* 1995;273:313–9.
5. Ponikowski P, Voors AA, Anker SD, *et al*. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: the task force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur J Heart Fail* 2016;18:891–975.
6. Mulrow CD, Dolmatch BL, Delong ER, *et al*. Observer variability in the pulmonary examination. *J Gen Intern Med* 1986;1:364–7.
7. Elphick HE, Lancaster GA, Solis A, *et al*. Validity and reliability of acoustic analysis of respiratory sounds in infants. *Arch Dis Child* 2004;89:1059–63.
8. Jauhar S. The demise of the physical exam. *N Engl J Med* 2006;354:548–51.
9. Spiteri MA, Cook DG, Clarke SW. Reliability of eliciting physical signs in examination of the chest. *Lancet* 1988;1:873–5.
10. Badgett RG, Tanaka DJ, Hunt DK, *et al*. Can moderate chronic obstructive pulmonary disease be diagnosed by historical and physical findings alone? *Am J Med* 1993;94:188–96.
11. Badgett RG, Tanaka DJ, Hunt DK, *et al*. The clinical evaluation for diagnosing obstructive airways disease in high-risk patients. *Chest* 1994;106:1427–31.
12. Margolis PA, Ferkol TW, Marsocci S, *et al*. Accuracy of the clinical examination in detecting hypoxemia in infants with respiratory illness. *J Pediatr* 1994;124:552–60.
13. Wipf JE, Lipsky BA, Hirschmann JV, *et al*. Diagnosing pneumonia by physical examination: relevant or relic? *Arch Intern Med* 1999;159:1082–7.
14. Holleman DR, Simel DL, Goldberg JS. Diagnosis of obstructive airways disease from the clinical examination. *J Gen Intern Med* 1993;8:63–8.
15. Brooks D, Thomas J. Interrater reliability of auscultation of breath sounds among physical therapists. *Phys Ther* 1995;75:1082–8.
16. Workum P, DelBono EA, Holford SK, *et al*. Observer agreement, chest auscultation, and crackles in asbestos-exposed workers. *Chest* 1986;89:27–9.
17. McCollum ED, Park DE, Watson NL, *et al*. Listening panel agreement and characteristics of lung sounds digitally recorded from children aged 1-59 months enrolled in the Pneumonia Etiology Research for Child Health (PERCH) case-control study. *BMJ Open Respir Res* 2017;4:e000193.
18. Andreeva E, Melbye H. Usefulness of C-reactive protein testing in acute cough/respiratory tract infection: an open cluster-randomized clinical trial with C-reactive protein testing in the intervention group. *BMC Fam Pract* 2014;15:80.
19. Francis NA, Melbye H, Kelly MJ, *et al*. Variation in family physicians' recording of auscultation abnormalities in patients with acute cough is not explained by case mix. A study from 12 European networks. *Eur J Gen Pract* 2013;19:77–84.
20. Vanbelle S. Comparing dependent kappa coefficients obtained on multilevel data. *Biom J* 2017;59:1016–34.
21. Vanbelle S. *R package "multiagree": comparison of dependent kappa coefficients*, 2017.
22. Wei T, Simko V. *R package "corrplot": visualization of a correlation matrix*, 2017.
23. Kottner J, Audigé L, Brorson S, *et al*. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96–106.
24. Melbye H, Garcia-Marcos L, Brand P, *et al*. Wheezes, crackles and rhonchi: simplifying description of lung sounds increases the agreement on their classification: a study of 12 physicians' classification of lung sounds from video recordings. *BMJ Open Respir Res* 2016;3:e000136.
25. Veloso SG, Lima MF, Salles PG, *et al*. Interobserver agreement of gleason score and modified gleason score in needle biopsy and in surgical specimen of prostate cancer. *Int Braz J Urol* 2007;33:639–51.
26. Timmers JM, van Doorne-Nagtegaal HJ, Verbeek AL, *et al*. A dedicated BI-RADS training programme: effect on the inter-observer variation among screening radiologists. *Eur J Radiol* 2012;81:2184–8.
27. Brosnan M, La Gerche A, Kumar S, *et al*. Modest agreement in ECG interpretation limits the application of ECG screening in young athletes. *Heart Rhythm* 2015;12:130–6.
28. Halford JJ, Shiau D, Desrochers JA, *et al*. Inter-rater agreement on identification of electrographic seizures and periodic discharges in ICU EEG recordings. *Clin Neurophysiol* 2015;126:1661–9.
29. Zhou S, Zha Y, Wang C, *et al*. [The clinical value of bedside lung ultrasound in the diagnosis of chronic obstructive pulmonary disease and cardiac pulmonary edema]. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue* 2014;26:558–62.
30. Pasterkamp H, Brand PL, Everard M, *et al*. Towards the standardisation of lung sound nomenclature. *Eur Respir J* 2016;47:724–32.
31. Robertson AJ, COOPE R. Rales, rhonchi, and laennec. *Lancet* 1957;273:417–23.
32. MLP SARA, Charbonneau G, Vanderschoot J, *et al*. Characteristics of breath sounds and adventitious respiratory sounds. *Eur Respir Rev* 2000;10:591–6.
33. Guggenmoos-Holzmann I. The meaning of kappa: probabilistic concepts of reliability and validity revisited. *J Clin Epidemiol* 1996;49:775–82.
34. Chmura Kraemer H, Periyakoil VS, Noda A. Kappa coefficients in medical research. *Stat Med* 2002;21:2109–29.
35. Kraemer HC, Bloch DA. Kappa coefficients in epidemiology: an appraisal of a reappraisal. *J Clin Epidemiol* 1988;41:959–68.

Paper II

Original Research Article

Impact of spectrograms on the classification of wheezes and crackles in an educational setting. An interrater study.

Aviles-Solis JC.[1], Storvoll I[2], Vanbelle Sophie[3], Melbye H.[1]

1) General Practice Research Unit, Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway.
2) Faculty of Health Sciences, UiT, The Arctic University of Norway, Tromsø, Norway.
3) Department of methodology and statistics, University of Maastricht, Maastricht, The Netherlands.

Corresponding author: Juan Carlos Aviles Solis, UiT The Arctic University of Norway, Post-box 6050 Langnes, 9037 Tromsø. E-mail: juan.c.solis@uit.no

Word count: 2182

Abstract

Background: Chest auscultation is a widely used method in the diagnosis of lung diseases. However, the interpretation of lung sounds is a subjective task and disagreements arise. New technological developments like the use of visual representation of sounds through spectrograms could improve the agreement when classifying lung sounds, but this is not yet known.

Aims: To test if the use of spectrograms improves the agreement when classifying wheezes and crackles.

Methods: We used 30 lung sounds recordings. The sample contained 15 normal recordings and 15 with wheezes or crackles. We produced spectrograms of the recordings. Twenty-three third to fifth-year medical students at UiT the Arctic University of Norway classified the recordings using an online questionnaire. We first showed the students examples of how wheezes and crackles looked in the spectrogram. Then, we played the recordings in a random order two times, first without the spectrogram, then with live spectrograms displayed. We asked them to classify the sounds for the presence of wheezes and crackles. We calculated kappa values for the agreement between each student and the expert classification with and without display of spectrograms and tested for significant improvement. We also calculated Fleiss kappa for the 23 observers with and without the spectrogram.

Results: When classifying wheezes 13/23 (1 with p<.05) students had a positive change in k, and 16/23 (2 with p<.05). All the statistically significant changes were in the direction of improved kappa values (.52 - .75). Fleiss kappa values were k=.51 and k=.56 (p=.63) for wheezes without and with spectrograms. For crackles, these values were k=.22 and k=.40 (p=<0.01) in the same order.

Conclusions: The use of spectrograms had a positive impact on the inter-rater agreement and the agreement with experts. We observed a higher improvement in the classification of crackles compared to wheezes.

Introduction

Chest auscultation is a widely used method in the diagnosis and follow up of several diseases. Medical doctors use it to guide clinical decisions and treatment strategies. (1) However, the identification and interpretation of the sounds remains a subjective task. Generally, auscultation occurs in a "solitary" fashion since normal stethoscopes are designed for individual listening. Often, disagreement between health professionals arises when classifying lung sounds and the reliability in the classification of wheezes and crackles has been found to be moderate at best. (2-6) This variation is caused by not only diverse identification, but also due to different labeling of the sounds. (7, 8) These limitations of auscultation make the training of new health professionals a challenging task. It has been suggested that difficulties in teaching and learning auscultation also contribute to the demise of this technique. (9)

However, new electronic stethoscopes can capture and store sounds in a digital form. Assisted by the processing capacity of personal computers and/or mobile phones it is possible to generate visual representations of sound in the form of spectrograms. In spectrograms, the common adventitious lung sounds, wheezes and crackles, show recognizable patterns, (figure 1) which may be of help in the identification of these sounds. (10) Some of the electronic stethoscopes available on the market offer this solution. The use of spectrograms also gives the possibility for group analysis and discussion. In addition, the use of recordings with spectrograms could avoid the objectification of patients used in the teaching of auscultation. (11)

It would be natural to think that the use of a visual support could help to improve the classification of lung sounds since there are two sensory inputs instead of one. The use of spectrograms could improve the teaching of lung sounds by giving an aid to the listening abilities in training. Andrés et al (12) found that the spectrograms do have a positive impact on how medical students assign a diagnosis with the help of lung sounds. However, the design of this study did not isolate the effect of spectrograms and therefore this affirmation is inconclusive. There are very few studies on this subject. Therefore, we do not know whether the use of spectrograms would help to medical students to better classify lung sounds. If spectrograms have a positive effect in the classification of lung sounds it would make it a valuable tool to make auscultation training more simple and effective.

The aim of this study is to explore how the use of spectrograms affects the agreement between medical students and a panel of experts and the agreement within a group of students in classifying lung sounds.

Methods

Data for classification

We recorded lung sounds from 20 adults. They were 67.4 years old on average (44–84) and nine were female. We registered the following information about the subjects: age, gender and self-reported history of heart or lung disease. No personal information was registered that could link the sound recordings to the individual subjects. The project was presented for the Regional Committee for Medical and Health Research Ethics, and it was considered to be outside the remit of the Act on Medical and Health Research.

To record the lung sounds, we used a microphone MKE 2-EW with a wireless system EW 112-P G3-G (Sennheiser electronic, Wedemark, Germany) placed in the tube of a Littmann Master Classic

II stethoscope (3M, Maplewood, MN, USA) at a distance of 10 cm from the headpiece. The microphone was connected to a digital sound Handy recorder H4n (Zoom, Tokyo, Japan).

We placed the membrane of the stethoscope against the naked thorax of the subjects. We asked the subjects to breathe deeply while keeping their mouth open. We started the recording with an inspiration and continued for approximately 20 seconds trying to capture three full respiratory cycles with good quality sound. We performed this same procedure at six different locations (figure 2). The researcher recording sounds used a headphone as an audio monitor to evaluate the quality. When too much noise or cough was heard during the recording, a second attempt was performed.

We obtained 120 audio files in '.wav' format and recorded at a sample rate of 44 100 Hz and 16 bit depth in a single monophonic channel. We did not perform post-processing of the sound files or implement filters. We chose 30 recordings for this study from 18 different subjects. This selection contained 15 normal and 15 abnormal sounds, from which nine were classified as containing crackles and six containing wheezes.

A panel of four experts in the field of lung sound research classified the recordings according to the presence of wheezes and crackles by a majority criterion.

Observers

We invited medical students from third to fifth year at UiT, The Arctic University of Norway. The students had received the standard curricular training in physical examination provided in the medicine school. The training in lung auscultation included a lecture on lung sounds which included the demonstration of some spectrograms. Beyond this, the students had no previous experience with the use of spectrograms.

Presentation and classification of the lung sounds

First, we presented a couple of recordings with wheezes and crackles and showed the students how these sounds looked in the spectrograms. Then we played the 30 recordings in a random order in two sessions. In the first session, we presented sound only. In the second session, the sound and the spectrograms were simultaneously displayed in the classroom screen. There was a pause of 20 minutes in between the two sessions. We presented no additional information beyond the sound and the spectrograms. The observers were not aware that the same sounds were played in both sections.

In both sessions, each recording was played two times and the students had up to 30 seconds to classify it before moving to the next recording. The observers used their personal computers and an online classification scheme (Questback AS, Norway). In this scheme, the observers had to specify if the recording contained only normal respiratory sounds. If this was not the case, the observers had to further specify if the recording contained wheezes, crackles or other sounds and if they appeared during inspiration or expiration. It was also possible to mark the recording as containing too much noise to be classified. At the end of the classifications session we obtained a report in an excel document. (Microsoft,Redmond,WA,USA)

Statistical analysis

We calculated Cohen kappa for the agreement between the observers and the experts, and Fleiss kappa for all the observers as a group. We then compared the kappa values obtained in the sections with and without the use of spectrograms and calculated p values to explore for statistically significant differences using an adaption of Hotelling's T2 test described by Vanbelle, S. (13) In this analysis, the recordings were clustered by the individual they were recorded from. We used Holm's correction procedure to adjust p values for multiple hypothesis testing. We used R version 3.2.1 and the package "magree" to perform all the calculations. Significance level was set at $p < 0.05$. In addition, we calculated sensitivity and specificity of each participant using the experts' classification as the gold standard. We tested for significant differences in sensitivity and specificity with and without the intervention using paired Wilcoxon signed rank test with continuity correction.

The results of this study are reported according to the Guidelines for Reporting Reliability and Agreement Studies (GRRAS). (14)

Results

Observers

We included 23 observers in the study. At the beginning, 30 students accepted to participate in the study. From them, two withdrew before the start of the study, two did not show up and three did not complete the classification session due to lack of time. Eight participants were third year students, fourteen participants were from the fourth year and one was from the fifth year. There were 19 women and four men.

Agreement

The students observed a mean prevalence of wheezes of 9.7 (6 – 15) without spectrograms and 8.3 (5 – 12) with spectrograms. In the case of crackles the students observed a mean prevalence of 11.5 (4 – 22) and 10.9 (5 – 18) in the same order. The mean proportion of agreement (%) and Cohen kappa (k) with the experts for all 23 participants classifying wheezes without spectrograms was 82 % and k=.56. We observed 88 % and k=.68 with the use of spectrograms. In the case of crackles we observed a proportion of agreement of 72 % and k=.38 without spectrograms. With the use of spectrograms 80 % and k=.56. Fleiss kappa values for the multirater agreement were k=.51 and k=.56 (p=.63) for wheezes without and with spectrogram, respectively. For crackles, we observed k=.22 and k=.40 (p=<0.01) in the same order. (figure 3) Compared to the expert panel's classification, 13/23 students had a positive change in kappa when classifying wheezes (one with p<.05), and 16/23 (two with p<.05) when classifying crackles. (figure 4 and figure 5) All the statistically significant changes were in the direction of improved kappa values (.52 - .75).

When looking at the classification of normal vs abnormal sounds (wheezes or crackles) we observed a mean prevalence of abnormal sounds of 18.7 and 18 with and without spectrogram (experts 15). The mean absolute agreement was 72 % with a mean kappa of k= 0.44 without spectrograms and 80% and k=0.60 with spectrograms. Only one participant had a significant improvement in this analysis.

The median sensitivity for wheezes did not present a significant change but the specificity was higher in the classification with the use of spectrograms (p=0.002). In the case of crackles, there was a significant increase in sensitivity (p=0.03) when using spectrograms but without significant change in specificity. (figure 6)

Discussion

We found improved agreement with the experts in the classification of lung sounds with the use of spectrograms. However, most of the improvements were not statistically significant. We did observe a significant improvement in the agreement within the group (Fleiss kappa) when classifying for crackles when the sounds were presented with spectrograms.

The levels of individual and group agreement observed in this study corresponds with that reported in previous studies. (15) We generally observed a higher agreement for wheezes than for crackles, in accordance with what is described in the literature. (3) It is interesting that the impact of spectrograms was different for wheezes and crackles. It might be that wheezes are easier to recognize without spectrograms due to its relatively long duration and its musical quality, which make them more familiar to the human ear. On the contrary, crackles are short explosive sounds that could easily be missed by ear appreciation or perceived as noise. (16) For this reason, having a visual aid could be an advantage to identify them.

Andrés et al (12) found that the spectrograms do have a positive impact on how medical students assign a diagnosis to a patient with the help of lung sounds. Even though the observer populations are similar, the results are not comparable since the outcome in their test was a diagnosis and not the classification of the sounds. Since the participants in Andrés' study got clinical information together with the sound this could have influenced how they classified lung sounds. Nguyen et al observed that the addition of clinical information has an effect on the classification of lung sounds and this effect is experience-dependent. In Nguyen's study, the group classifying lung sounds without clinical information achieved similar scores classifying lung sounds regardless of clinical experience. When clinical information was provided, then the more experienced raters achieved higher scores. (17)

Our exploratory study suggests that the use of spectrograms might be helpful to improve the teaching of auscultation, mostly in the cases of crackles. This by enhancing interrater agreement and facilitating discussion of the sounds in a teaching arena.

Strengths and limitations

We think the methodology employed allowed us to analyze the isolated effect of the spectrograms in the classifications.

The students had free seating. Even though absolute silence was required during the classification, we cannot fully rule out the possibility that the students could have influenced each other. This could have influenced the estimation of agreement in the group (multi-rater) agreement in a positive direction.

We have tested the same hypothesis 23 times. This increases the chance of making a type I error. We have taken this situation into account by correcting the p-values with Holm's procedure.

Some limitations in our study concern the sample size of sounds to classify. Due to the exploratory nature of this study, we chose a number of sounds to be classified which could allow us to perform the whole procedure in a time window of two hours. This, to avoid dropouts from the study. It is possible that we could have had observed more significant changes if the sound sample would have been larger. Future studies looking at the effect of spectrograms in agreement should take this into account in its design.

Conclusion

The use of spectrograms had a positive impact on the inter-rater agreement and the agreement compared to experts. We observed a higher improvement in the classification of crackles compared to wheezes.

Akcnowledgements: We would like to thank Professor A. Marques, Professor P. Piirilä, Professor H. Pasterkamp and Professor A. Sovijarvi for their help in classifying the lung sound recordings used in this study. We would like to thank Dr. Cristina Jácome for its valuable insight. The authors also would like to thank LHL rehabilitation centre at Skibotn, Norway, for its cooperation in performing the recordings.

References

1. Turnbull S, Lucas PJ, Redmond NM, Christensen H, Thornton H, Cabral C, et al. What gives rise to clinician gut feeling, its influence on management decisions and its prognostic value for children with RTI in primary care: a prospective cohort study. BMC Fam Pract. 2018;19(1):25.
2. Badgett RG, Lucey CR, Mulrow CD. Can the clinical examination diagnose left-sided heart failure in adults? JAMA. 1997;277(21):1712-9.
3. Benbassat J, Baumal R. Narrative review: should teaching of the respiratory physical examination be restricted only to signs with proven reliability and validity? J Gen Intern Med. 2010;25(8):865-72.
4. Brooks D, Thomas J. Interrater reliability of auscultation of breath sounds among physical therapists. Phys Ther. 1995;75(12):1082-8.
5. Margolis P, Gadomski A. The rational clinical examination. Does this infant have pneumonia? JAMA. 1998;279(4):308-13.
6. Metlay JP, Kapoor WN, Fine MJ. Does this patient have community-acquired pneumonia? Diagnosing pneumonia by history and physical examination. JAMA. 1997;278(17):1440-5.
7. Spiteri MA, Cook DG, Clarke SW. Reliability of eliciting physical signs in examination of the chest. Lancet. 1988;1(8590):873-5.
8. Melbye H, Garcia-Marcos L, Brand P, Everard M, Priftis K, Pasterkamp H. Wheezes, crackles and rhonchi: simplifying description of lung sounds increases the agreement on their classification: a study of 12 physicians' classification of lung sounds from video recordings. BMJ Open Respir Res. 2016;3(1):e000136.
9. Rice T. Learning to listen: auscultation and the transmission of auditory knowledge. J Roy Anthropol Inst. 2010;16:S41-S61.
10. Riella RJ, Nohama P, Maia JM. Method for automatic detection of wheezing in lung sounds. Braz J Med Biol Res [Internet]. 2009;42(7):674-84.
11. Rice T. “Beautiful Murmurs”: Stethoscopic Listening and Acoustic Objectification. The Senses and Society. 2008;3(3):293-306.
12. Andres E, Gass R. Stethoscope: A Still-Relevant Tool and Medical Companion. Am J Med. 2016;129(5):e37-8.
13. Vanbelle S. Comparing dependent kappa coefficients obtained on multilevel data. Biom J. 2017;59(5):1016-34.
14. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64(1):96-106.
15. Aviles-Solis JC, Vanbelle S, Halvorsen PA, Francis N, Cals JWL, Andreeva EA, et al. International perception of lung sounds: a comparison of classification across some European borders. BMJ Open Respir Res. 2017;4(1):e000250.
16. Kiyokawa H, Greenberg M, Shirota K, Pasterkamp H. Auditory detection of simulated crackles in breath sounds. Chest. 2001;119(6):1886-92.
17. Nguyen DQ, Patenaude JV, Gagnon R, Deligne B, Bouthillier I. Simulation-based multiple-choice test assessment of clinical competence for large groups of medical students: a comparison of auscultation sound identification either with or without clinical context. Can Med Educ J. 2015;6(1):e4-e13.

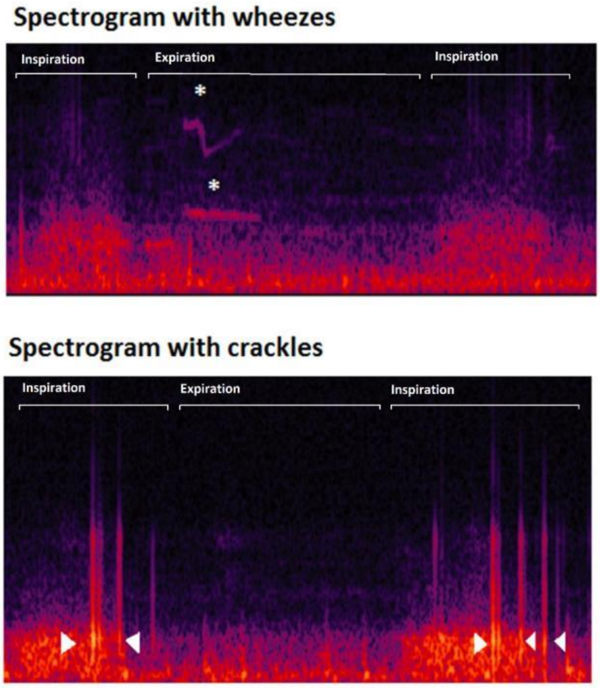


Figure 1.- Examples of spectrograms of lung sound recordings showing the presence of wheezes (stars) and crackles (arrowheads.

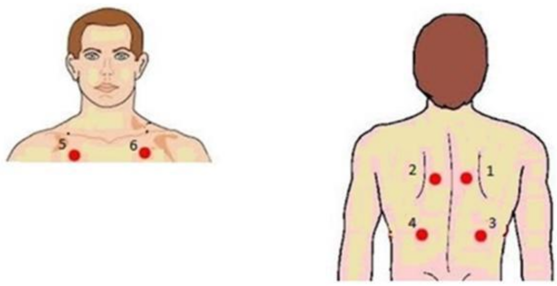


Figure 2.-Illustration showing the different places where lung sounds were recorded. (1_2) Between the spine and the medial border of the scapula at the level of T4–T5; (3_4) at the middle point between the spine and the mid-axillary line at the level of T9–T10; (5_6) at the intersection of the mid-clavicular line and second intercostal space.

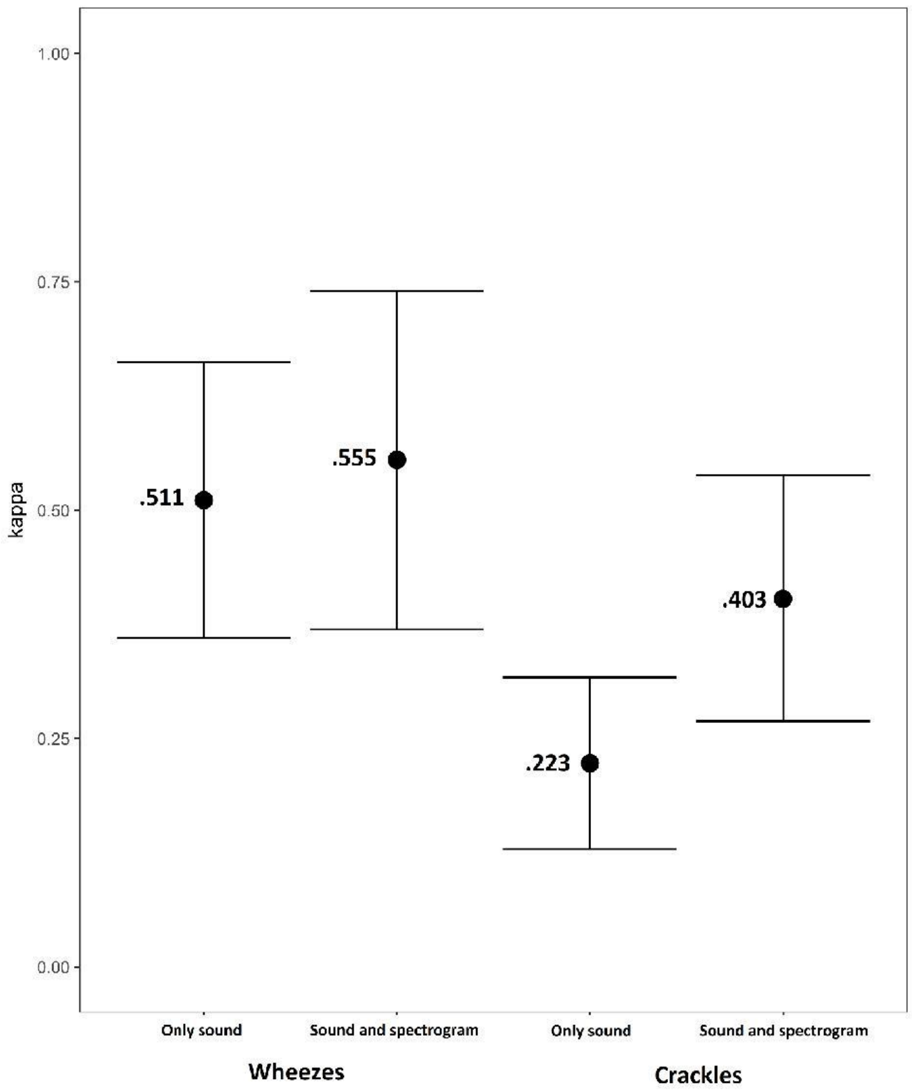


Figure 3.- Fleiss kappa for the group of 23 participants when classifying wheezes and crackles with only sound and sound plus spectrogram.

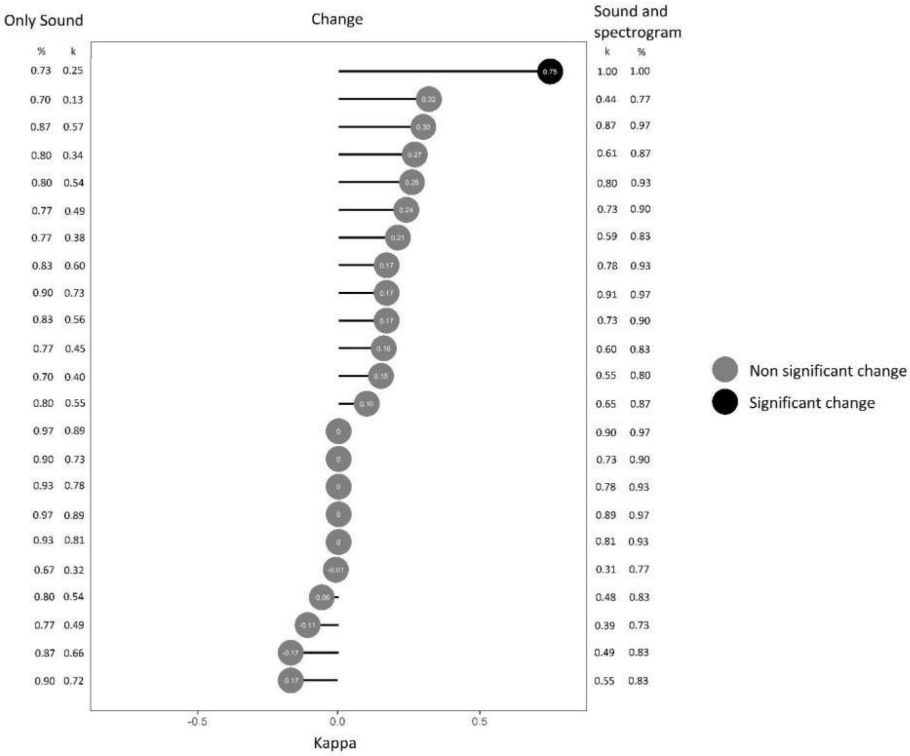


Figure 4.- Cohen's kappa of each participant when classifying wheezes with only sound (left) and with sound and spectrograms (right) compared to the reference standard. The change in kappa between the two classifications and its statistical significance is illustrated at the center. Proportion of agreement (%) is presented on the lateral columns.

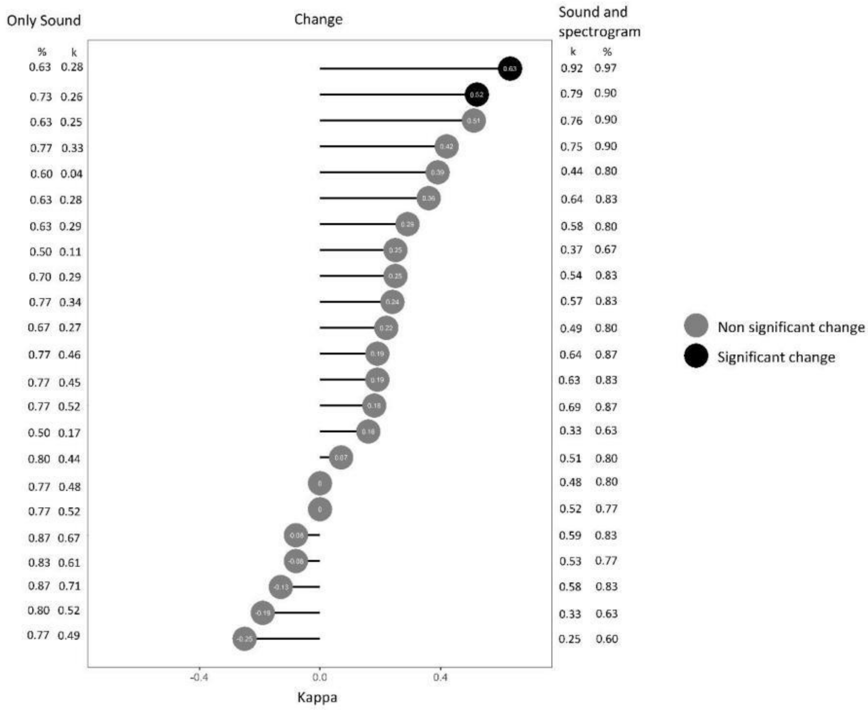


Figure 5.- Cohen's kappa of each participant when classifying crackles with only sound (left) and with sound and spectrograms (right) compared to the reference standard. The change in kappa between the two classifications and its statistical significance is illustrated at the center. Proportion of agreement (%) is presented on the lateral columns.

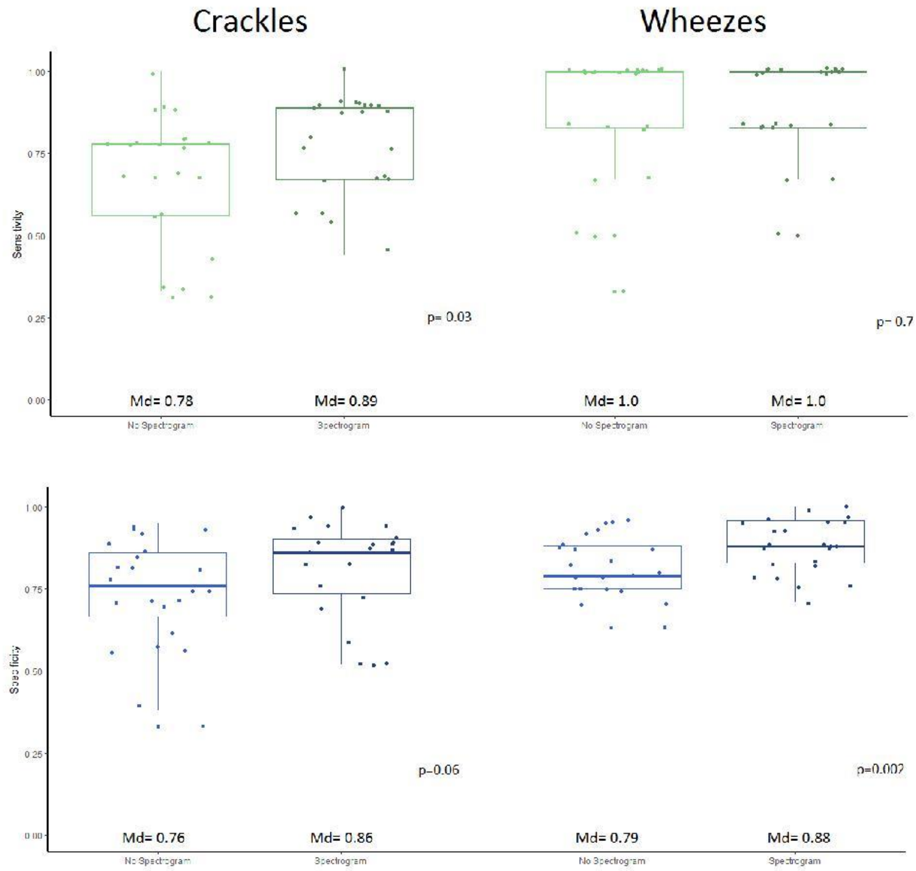


Figure 6.- Box and whiskers diagrams showing the change in sensitivity and specificity of the students classifying wheezes and crackles with and without spectrograms. The answers of the experts were considered as the reference standard. P values shown were obtained from the test of difference between means using paired Wilcoxon signed rank test with continuity correction. Md= Median.

Paper III

Paper IV

Aviles-Solis *et al. BMC Pulmonary Medicine* (2019) 19:173
https://doi.org/10.1186/s12890-019-0928-1

BMC Pulmonary Medicine

RESEARCH ARTICLE Open Access

Prevalence and clinical associations of wheezes and crackles in the general population: the Tromsø study

J. C. Aviles-Solis[1*], C. Jácome[2], A. Davidsen[1], R. Einarsen[1], S. Vanbelle[3], H. Pasterkamp[4] and H. Melbye[1]

Abstract

Background: Wheezes and crackles are well-known signs of lung diseases, but can also be heard in apparently healthy adults. However, their prevalence in a general population has been sparsely described. The objective of this study was to determine the prevalence of wheezes and crackles in a large general adult population and explore associations with self-reported disease, smoking status and lung function.

Methods: We recorded lung sounds in 4033 individuals 40 years or older and collected information on self-reported disease. Pulse oximetry and spirometry were carried out. We estimated age-standardized prevalence of wheezes and crackles and associations between wheezes and crackles and variables of interest were analyzed with univariable and multivariable logistic regressions.

Results: Twenty-eight percent of individuals had wheezes or crackles. The age-standardized prevalence of wheezes was 18.6% in women and 15.3% in men, and of crackles, 10.8 and 9.4%, respectively. Wheezes were mostly found during expiration and crackles during inspiration. Significant predictors of expiratory wheezes in multivariable analyses were age (10 years increase - OR 1.18, 95%CI 1.09–1.30), female gender (1.45, 1.2–1.8), self-reported asthma (1.36, 1.00–1.83), and current smoking (1.70, 1.28–2.23). The most important predictors of inspiratory crackles were age (1.76, 1.57–1.99), current smoking, (1.94, 1.40–2.69), mMRC ≥2 (1.79, 1.18–2.65), SpO_2 (0.88, 0.81–0.96), and FEV_1 Z-score (0.86, 0.77–0.95).

Conclusions: Nearly over a quarter of adults present adventitious lung sounds on auscultation. Age was the most important predictor of adventitious sounds, particularly crackles. The adventitious sounds were also associated with self-reported disease, current smoking and measures of lung function. The presence of findings in two or more auscultation sites was associated with a higher risk of decreased lung function than solitary findings.

Keywords: Wheezes, Crackles, Auscultation, Population

Background

Two hundred years after its invention, the relevance of the stethoscope in modern medical practice has become a topic of debate [1, 2]. There are some obvious advantages of lung auscultation, such as availability, low cost and non-invasiveness. Lung auscultation remains thus an important part of the respiratory examination, mainly in primary care and in resource-constrained settings.

Lung auscultation has shown to be useful in diagnosing various respiratory disorders. Adventitious lung sounds (ALS) such as wheezes and crackles are associated with common diseases like asthma [3], chronic obstructive pulmonary disease (COPD) [4, 5], interstitial lung disease [6], bronchiectasis [7], heart failure [8] and pneumonia [9–11]. Positive findings during auscultation influence clinical decisions such as the rate of antibiotic prescriptions [12, 13] and referrals to specialist care [14].

Presence of ALS alone, however, only show moderate sensitivities and specificities, limiting their diagnostic utility [15–17]. This modest accuracy is mainly related

* Correspondence: juan.c.solis@uit.no
[1]General Practice Research Unit, Department of Community Medicine, UIT the Arctic University of Norway, Tromsø, Norway
Full list of author information is available at the end of the article

to the fact that both wheezes and crackles can also be present in apparently healthy adults [10, 18–20]. To determine the real usefulness of ALS it is crucial to define first their behavior, presence and characteristics, in apparently healthy people. Most studies to date, however, have investigated how ALS relate to specific diagnostic categories without considering their distribution across the whole spectrum from health to disease. Moreover, the few existent studies investigating ALS in apparently healthy people used small samples [19], failing to be representative of the general population. The prevalence of wheezes and crackles in a general population has never been reported. [21].

With this study, we aimed to estimate the prevalence of wheezes and crackles in a large general adult population. We also explored to which degree ALS are associated with self-reported disease, smoking status and clinical measures of lung function.

Methods

Design and participants

The Tromsø Study is an epidemiological survey that started in 1976 with the main goal to determine the reasons for the high cardiovascular mortality in the municipality of Tromsø, Norway. The study has been periodically repeated with the last survey (7th) taking place in 2015–16.Details of the Tromsø Study can be consulted elsewhere [22, 23].

In this cross-sectional study, our sample consisted of randomly selected participants attending the second visit of the seventh survey of the Tromsø study (Tromsø 7), between May 2015 and October 2016. All Tromsø residents 40 years and older ($n = 32{,}591$) received a postal invitation to participate in the first visit of Tromsø 7. A random sample was selected for the second visit including 20% of those aged 40–59 years and 60% of those aged 60–84 years, and those attending the first visit were invited. In addition, individuals who had participated in previous surveys of the study were invited to obtain repeated measurements. The mean time between the visits was of 52 days (± 32). All study participants provided written consent. The Regional Committee for Medical and Health Research Ethics in North Norway approved the study.

Questionnaires and examinations

In the first visit, the participants filled a questionnaire that included questions on medical conditions such as arterial hypertension, heart failure, atrial fibrillation, COPD, asthma, among others. For each condition, the participants were asked to specify if it was a current diagnosis, if they had that diagnosis at some point in the past or if they never had that diagnosis. They also responded questions about smoking habits. The full questionnaires employed at the Tromsø Study can be consulted in English elsewhere [22].

At the second visit, the participants answered the modified Medical Research Council questionnaire (mMRC) on dyspnea [24]. Dyspnea was further characterized using the question: "How is your breathing today compared to normal?". To better characterize the respiratory status, participants were also asked if they had respiratory infection in the previous week ("Have you had symptoms of common cold, bronchitis or other airway infection the last 7 days?").

Spirometry was performed using SensorMedics Vmax 20c Encore (VIASYS Healthcare Respiratory Technologies, Yorba Linda, CA, USA). Calibration was done daily. We followed the standards of the American Thoracic Society (ATS)/ European Respiratory Society (ERS) [25]. Tests with FEV1 < 0.3 l and with expiration lasting for less than 3 s were regarded invalid. We did not perform post-bronchodilator measurements. We used the Global Lung Function Initiative (GLI 2012) as a reference [26]. We registered arterial oxygen saturation (SpO_2) with a pulse oximeter Onyx II model 9550 (Nonin Medical, Inc., Plymouth, MN, USA) after resting 15 min. The highest value after three measurements was registered. We accepted only $SpO_2 \geq 80\%$ due to uncertain validity of lower values ($n = 1$). At the end of this second visit we recorded lung sounds.

Recording of the lung sounds

We used a microphone MKE 2-EW with a wireless system EW 112-P G3-G (Sennheiser electronic GmbH, Wedemark, Germany), placed in the tube of a Littmann Classic II stethoscope (3 M, Maplewood MN, USA) at 10 cm from the headpiece. The signal went to an external sound card (Scarlett 2i2, Focusrite Audio Engineering Ltd., High Wycombe UK) which connected to a computer's audio input. The computer used custom developed software to label the sounds (participant ID, recording site) and allowed us to start the recording with a wireless control (R700, Logitech Europe S.A., Lausanne Switzerland).

We recorded in a quiet room with the participants sitting and the thorax exposed. They were asked to breathe deeper than normal with an open mouth. We started the recordings on inspiration and recorded for 15 s. We performed the same procedure subsequently at six different locations (Fig. 1). The quality of the recordings was monitored using a wireless headset (SDR 160, Sennheiser electronic GmbH, Wedemark, Germany). If the health professional deemed the quality to be unsatisfactory, a second attempt was performed.

We obtained audio files in ".wav" format at a sample rate of 44,100 Hz and 16-bit depth in a single (monophonic) channel. We did not implement audio filters or other digital pre or post-processing techniques.

Fig. 1 Recording sites and prevalence of findings. (1 and 2) Between the spine and the medial border of the scapula at the level of T4–T5; (3 and 4) at the middle point between the spine and the mid-axillary line at the level of T9–T10; (5 and 6) at the intersection of the mid-clavicular line and second intercostal space. Cr = crackles, Wh = wheezes, Insp = inspiratory, Exp = expiratory

Classification of the recordings

The classification process consisted of three steps.

At the first step, two observers independently listened to all the recordings with a headset and simultaneously viewed the sound spectrograms using Adobe Audition 5.0 (Adobe Systems, San Jose, CA, USA). J.C.A. was observer 1 and either R.E., A.D. or C.J. were observer 2. They evaluated if the recording contained wheezes (including rhonchi), crackles or other ALS and whether these were heard in inspiration or expiration. They entered their findings in an electronic form (Access, Microsoft Corporation, Redmond WA, USA) and registered if artefactual noise made the classification difficult. The observers could listen to the recordings with freedom to stop or repeat parts or the whole recording if necessary. They were blinded to any information about the participant. Agreement and kappa statistics were calculated accounting for the clustered structure of the data using the R package "magree". [27].

At the second step, all disagreements were evaluated with the two initial observers and a third experienced observer (H.M.). The three observers listened to the sounds and solved disagreements through consensus. If consensus was difficult to reach at this point, the sounds were submitted for classification at the third step.

At the third step, all recordings classified as containing ALS were re-classified by two pairs of observers consisting of one junior (J.C.A. and C.J.) and senior (H.P. and H.M.) lung sound researcher each. These observers had the possibility to mark the findings as "certain", "possible" or "absent". A finding was changed into absent when wheezes or crackles were marked as "absent" or "possible" by both observers. Findings classified as "present" by one observer and "absent" by the other were discussed in a face-to-face meeting with all the four observers. Agreement between at least three out of four observers was required to classify an ALS as "present". At the same session, difficult sounds from step two and sounds categorized as "other sounds" were classified.

All observers performed an audiometry at the time of involvement in the project. All observers had normal hearing.

Statistical methods

We calculated age-standardized prevalence of wheezes and crackles in men and women using the population distribution from the municipality of Tromsø per January 2018 [28]. The ALS were divided into three categories: wheezes and no crackles, crackles and no wheezes, and both wheezes and crackles, irrespective of respiratory phase. We calculated prevalence by participant characteristics and used linear models to explore statistically significant differences among the groups. Tukey's procedure was used to account for multiple testing. The continuous variables were dichotomized with cutoff values for age $\geq$ 65 years, for oxygen saturation $\leq$ 95% [29], mMRC score $\geq$ 2 [30], Body Mass Index (BMI) $\geq$30 (obesity threshold) and FEV_1 below the lower limit of normal (LLN), according to the Global Lung Initiative reference [26].

We used univariable logistic regression to study wheezes and crackles in relation to the variables of interest. In this analysis, wheezes were counted as present also when accompanied by crackles and vice versa. The following outcome variables were considered separately: (1) any wheeze, (2) wheezes only during the inspiratory phase, (3) wheezes during the expiratory phase and (4) wheezes during the expiratory phase at two or more recording sites. For crackles, the outcomes were (1) any crackle, (2) inspiratory crackles, (3) inspiratory crackles at two or more locations, (4) only expiratory crackles. The categorical variables of $FEV_1 <$ LLN and $SpO_2 \leq 95\%$ were substituted by continuous data (FEV_1 Z-score and SpO_2%) to avoid loss of information. We divided age per decades and kept it as a continuous variable. The variables of self-reported disease were dichotomized as present (which included both present or past diagnosis) and absent (never diagnosed).

All statistically significant variables for each outcome in the univariable analyses were entered into multivariable logistic regression models. We performed a backward elimination procedure with a threshold of $p < .05$ to obtain the best fitting models for each outcome. We plotted Receiver Operator Characteristics (ROC) curves for all the final models and calculated the area under the curve (AUC) with the r package "pROC" [31]. Multicollinearity in the final models was assessed using variance inflation factor with the statistical package "car" [32]. We used R statistical computing version 3.2.1 package to perform all the calculations [33]. Results were considered significant at 5% level.

Results

Participants

Tromsø 7 had an attendance of 21,083 (65%) in the first visit [22]. Of these, 9253 had been selected in advance to be invited to the second visit, and 90% ($n = 8346$) took part. Limited by absences of the staff, we recorded lung sounds in 6035 (72.3%). Restricted by human resources and time constraints, only 4033 participants were included in the classification procedure. Our final number of participants represents 19.1% of the participation in Tromsø 7 and 48.3% of those attending the second visit. A comparison of the main characteristics between all the participants of Tromsø 7 and the final study sample and the flow diagram of the participants included in our analyses are available online. (Additional file 1: Table S1, Additional file 2: Figure S1).

General characteristics of the groups

The mean age of all 4033 participants was 63.5 years, and 2159 (53.5%) were female. (Table 1). There were 477 (11.0%) and 2372 (47%) current and previous smokers, respectively. We found an $FEV_1 < LLN$ in 286 (7.1%) participants and 182 (4.5%) had oxygen saturation ≤ 95% (Table 1). We observed that women had lower proportion of myocardial infarction, heart failure and past smokers, but they presented a higher proportion of self-reported asthma, dyspnea (mMRC) and oxygen saturation ≤ 95%.

Classification agreement

We included 24,198 (4033 × 6 recording sites) recordings for classification. At the first step the observers agreed on inspiratory wheezes in 98.7% of the recordings (kappa (k) = 0.43; 95%CI 0.37–0.49), on expiratory wheezes 96.2% (k = 0.56; 0.53–0.59), on inspiratory crackles in 96.5% (k = 0.46; 0.42–0.49), and on expiratory crackles in 98.5% (k = 0.20; 0.15–0.25). Examples of the recordings can be consulted online (Additional file 3: Figure S3).

At the second step, 1257 recordings were marked as containing wheezes and 894 containing crackles. At the third step we discarded wheezes in 224 of these recordings and crackles in 174.

The presented prevalence of ALS are based on six recordings in 3771 (93.5%) participants. However, in 262 (6.5%) of the participants included in the analysis there was noise in one or more recordings. Five recording sites were considered in 223 (5.5%) participants and four or less recording sites in 39 (1%) participants.

Prevalence of wheezes and crackles

We found 28% ($n = 1131$) of individuals with ALS at least at one recording site. Of these, 599 (14.9%) had only wheezes, 402 (10.0%) had only crackles and 130 (3.2%) had both wheezes and crackles (Table 2). Expiratory wheezes and inspiratory crackles were the most common findings (Fig. 1). Of the 729 participants with wheezes, 534 (73.3%) had wheezes at one location, 132 (18.1%) at two locations, 63 (8.6%) at three or more locations. Of the 532 participants with crackles, 381 (71.6%) had crackles at one recording site, 127 (23.9%) at two recording sites, 24 (4.5%) at three or more recording sites. Inspiratory crackles were more frequent at the bases (Fig. 1).

The age-standardized prevalence of wheezes was 18.6% for women and 15.3% for men and of crackles, 10.8 and 9.4%, respectively. The prevalence of ALS increased significantly with age in both men and women ($p < .001$). This was particularly the case for crackles (Fig. 2). Pleural rub and bronchial breathing were rarely noticed, each in only two participants.

Wheezes or crackles were found in more than 40% of participants with the following characteristics: self-reported COPD, mMRC≥2, $FEV_1 < LLN$ and $SpO_2 \leq 95\%$ (Table 2). These characteristics were also associated with the highest prevalence of having both wheezes and crackles, 6.6–8.8% (Table 2).

Predictors of wheezes

In the univariable analysis, we found that wheezes were associated with age (10 years increase), female gender, self-reported asthma, current smoking, mMRC≥2, and a reduction in FEV_1 Z-score, (Table 3). The associations with mMRC≥2, current smoking and FEV_1 Z-score were stronger for inspiratory than for expiratory wheezes. In the multivariable analysis age, female gender, self-reported asthma, and current smoking predicted the occurrence of expiratory wheezes (Table 4). FEV_1-Z score was a significant predictor for the occurrence of inspiratory wheezes. The AUC for all the multivariable models were similar (0.59–0.60, Table 4). Multicollinearity was not problematic since the maximum variance inflation factor was < 1.07.

Table 1 Characteristics of the study population

	Male (n = 1874) n (%)	Female (n = 2159) n (%)	Missing (n = 4033) n (%)
Age	63.7 (±10.5)	63.4 (±10.7)	
< 65 years	908 (48.5%)	1071 (49.6%)	
≥ 65 years	966 (51.5%)	1088 (50.4%)	
Body-mass index			12 (0.3%)
< 30	1425 (76.0%)	1683 (78.0%)	
≥ 30	445 (23.7%)	468 (21.7%)	
Smoking status			59 (1.5%)
Never smoker	686 (36.6%)	916 (42.4%)***	
Current smoker	208 (11.1%)	269 (12.5%)	
Previous smoker	954 (50.9%)	941 (43.6%)***	
Self-reported disease			
Hypertension	473 (25.2%)	557 (25.8%)	119 (3.0%)
Myocardial Infarction	141 (7.5%)	57 (2.6%)***	171 (4.2%)
Heart failure	33 (1.8%)	16 (0.7%)**	175 (4.3%)
Atrial Fibrillation	92 (4.9%)	80 (3.7%)	178 (4.4%)
COPD	74 (3.9%)	87 (4.0%)	157 (3.9%)
Asthma	128 (6.8%)	196 (9.1%)**	254 (6.3%)
Rheumatoid arthritis	83 (4.4%)	117 (5.4%)	227 (5.6%)
Airways infection last week §	278 (14.8%)	303 (14.0%)	165 (4.1%)
Dyspnea			
mMRC			165 (4.1%)
mMRC 0	1323 (70.6%)	1368 (63.4%)***	
mMRC 1	412 (22.0%)	575 (26.6%)***	
mMRC 2–4	70 (3.7%)	120 (5.6%)**	
Breathing worse than usual §	210 (11.2%)	242 (11.2%)	156 (3.9%)
Oxygen saturation, SpO_2			161 (4.0%)
≤ 95%	110 (5.9%)	72 (3.3%)***	
Spirometry			
FEV_1 < LLN †	150 (8.0%)	136 (6.3%)	234 (5.8%)

Abbreviations: *mMRC* = Modified Medical Research Council questionnaire, FEV_1 = Forced Expiratory Volume in one second, *LLN* = Lower Limit of Normal
§On examination day
***p value <.001, **p value <.01, *p value <.05 as compared to male by X^2 test

Predictors of crackles

The explanatory variables were stronger predictors of crackles than of wheezes (Table 3). Age and gender were the only variables associated with expiratory crackles. For inspiratory crackles, the effect of age, self-reported COPD, asthma, current and previous smoking, mMRC ≥2, oxygen saturation and FEV_1 Z-score was stronger when inspiratory crackles were found at two or more recording sites than for inspiratory crackles at one site only. Similarly, in the multivariable analysis the strongest associations were found in the model with inspiratory crackles heard at two or more sites as outcome. This was the model with the highest area under the curve (AUC = 0.79). Inspiratory crackles appeared more often and at more locations in individuals with a negative FEV_1 Z-score and low oxygen saturation (Fig. 3). Multicollinearity was not problematic since the maximum variance inflation factor was < 1.01.

Predictors of wheezes and crackles in the same subject

In the multivariable analysis with both wheezes and crackles as outcome, age, female gender and FEV_1 Z-score were the significant predictors (data not shown). The AUC of the model was 0.7.

The variables “respiratory infection previous week” and “more short of breath than usual” predicted neither wheezes nor crackles.

Table 2 Frequency of wheezes, crackles and both by characteristics of the study population

	Normal n (%)	Wheezes, no crackles n (%)	Crackles, no wheezes n (%)	Both crackles and wheezes n (%)
All (*n* = 4033)	2902 (72.0%)	599 (14.9%)	402 (10.0%)	130 (3.2%)
Age				
< 65 years	1539 (77.8%)	277 (14.0%)**	123 (6.2%)***	40 (2.0%)***
≥ 65 years	1363 (66.4%)	322 (15.7%)	279 (13.6%)	90 (4.4%)
Gender				
Male	1389 (74.1%)	190 (13.4%)	251 (10.1%)	44 (2.3%)
Female	1513 (70.1%)	348 (16.1%) *	212 (9.8%)	86 (4.0%)**
Body-mass index				
< 30	2232 (71.8%)	483 (15.5%)*	292 (9.4%)	101 (3.2%)
≥ 30	663 (72.6%)	114 (12.5%)	107 (11.7%)	29 (3.2%)
Smoking status				
Never smoker	1209 (75.5%)	227 (14.2%)	133 (8.3%)**	33 (2.1%)**
Current smoker	299 (62.7%)	89 (18.7%) **	64 (13.4%) **	25 (5.2%) **
Previous smoker	1349 (71.2%)	278 (14.7%)	198 (10.4%)	70 (3.7%)
Self-reported disease				
Healthy †	1177 (76.8%)	218 (14.0%)	120 (7.0%)***	38 (2.4%)*
Hypertension	722 (70.1%)	151 (14.7%)	120 (11.7%)	37 (3.6%)
Myocardial Infarction	123 (62.1%)	35 (17.7%)	30 (15.2%) *	10 (5.1%)
Heart failure	32 (65.3%)	11 (22.4%)	5 (10.2%)	1 (2.0%)
Atrial Fibrillation	114 (66.3%)	29 (16.9%)	25 (14.5%)	4 (2.3%)
COPD	95 (59.0%)	27 (16.8%)	30 (18.6%) ***	9 (5.6%)
Asthma	209 (64.5%)	55 (17.0%)	38 (11.7%)	22 (6.8%) **
Rheumatoid arthritis	129 (64.5%)	29 (14.5%)	30 (15.0%) *	12 (6.0%)
Airways infection last week §	427 (73.5%)	89 (15.3%)	44 (7.6%)*	21 (3.6%)
Dyspnea				
mMRC				
mMRC 0	1996 (74.2%)	388 (14.4%)	239 (8.9%)**	68 (3.7%)**
mMRC 1	688 (69.7%)	148 (15.0%)	111 (11.2%)	40 (4.1%)
mMRC 2–4	109 (57.4%)	30 (15.8%)	36 (18.9%) ***	15 (7.9%) ***
Breathing worse than usual §	321 (71.0%)	70 (15.5%)	47 (10.4%)	14 (3.1%)
Oxygen saturation SpO_2				
≤ 95%	106 (58.2%)	25 (13.7%)	35 (19.1%)***	16 (8.8%)***
Spirometry				
FEV_1 < LLN ‡	175 (61.2%)	52 (18.2%)	40 (14.0%)*	19 (6.6%)***

Abbreviations: *mMRC* = Modified Medical Research Council questionnaire, *FEV_1* = Forced Expiratory Volume in one second, *LLN* = Lower limit of Normal
Plus-minus values are means + − SD
Percentages (%) represent the distribution of each variable between the different groups
****p* value <.001, ***p* value <.01, **p* value <.05 as compared to normal
† Not current smokers who stated not to have any of the diseases considered for this analysis
‡ Calculated from Global Lung Function Initiative reference (GLI)
§ On examination day

Discussion

Wheezes and crackles were common findings. Any of these sounds were found in almost one third of our sample. Wheezes and crackles were associated with increasing age. The sounds were not always related to clinically diagnosed disease, but their prevalence increased in the presence of decreased lung function or chronic shortness of breath.

We are not aware of any comparable study carried out in a general population. An investigation with 700

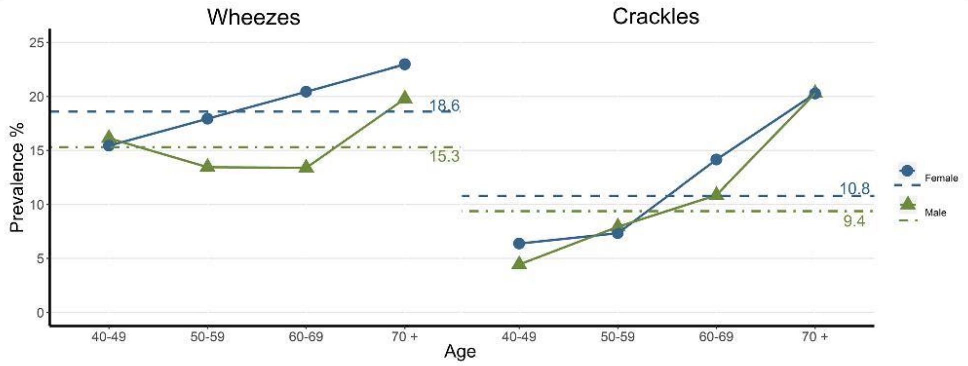


Fig. 2 Prevalence of wheezes and crackles by age (in years). Dotted lines represent standardized prevalence rates

participants conducted by Murphy et al., found wheezes and rhonchi in 4 and 4%, respectively, in a subgroup of 334 apparently healthy adults. Wheezes were heard in 59% of patients with asthma. For crackles, the prevalence was 21% in the apparently healthy group and 71% in patients with COPD. [10] These prevalences were higher than what we observed, except for wheezes in the apparently healthy group. In their study, the age of the participants was not taken into account. They included more recording sites than in ours and used a computerized classification of the sounds. Different sensitivities of the classification methods may partly explain the discordance in prevalence. Crackles detected by a computer algorithm may be inaudible with a stethoscope since crackles may be masked by normal sounds. [34].

In most participants with ALS in our study, these were heard at only one of the six recording sites. The number of sites with positive findings had an impact on the associations. The model with inspiratory crackles at two or more locations as outcome performed better than the model predicting any crackles, reaching an AUC of 0.79. No similar effect of increasing number of sites was found regarding wheezes.

The importance of age was described by Kataoka et al., who observed a rising prevalence of crackles from 11% in cardiovascular asymptomatic adults 40–65 years to 70% in participants 80–95 years old. [35] Murphy et al. found an association with age among asbestos exposed workers. [36] Age relates to a reduction of supporting tissue around the airways causing a premature closure of the airways. [37]. The influence of lung and heart disease associated with ageing might have contributed to the strong association between crackles and age in our study, an influence beyond what indicated by self-reported diseases, spirometry and pulse oximetry.

Self-reported asthma was associated with wheezes, which was in line with previous studies. [3] Hypertension, self-reported asthma, myocardial infarction, self-reported COPD, and RA were associated with crackles in the univariable analysis, but only the latter two remained statistically significant in the multivariable models. The association of crackles with RA could be explained by the presence of parenchymal lung abnormalities in patients with this diagnosis. [38] However, we did not have an independent confirmation of the diagnosis. Self-reported heart failure was not associated with crackles probably due to underdiagnosed heart failure [39]. Interestingly, symptoms suggestive of airway infection the week before the examination was not an independent predictor of crackles or wheezes. In a European study from primary care of 2810 adults with acute cough, crackles were registered in 31% of patients in the pneumonia group. [40] Since the prevalence of pneumonia in this study was only 5%, and probably far less in our study, it is likely that crackles represented chronic rather than acute changes in the lungs in most cases.

Women had a higher prevalence of wheezes than men. Considering subcategories of wheezes, this observation was valid for expiratory but not for inspiratory wheezes. The same gender disparity has been reported in epidemiological studies on self-reported wheeze [41, 42]. Although wheezing is more common in male newborns and infants, this gender difference seems to change sometime during adolescence when females start to show a higher risk of wheezing. [43] Our findings indicate that this may persist into later adulthood.

Oxygen saturation was significantly associated with the presence of inspiratory crackles. Crackles are related to the sudden opening of closed airways or to air movement through obstructed airways. [44] These conditions may impair ventilation/perfusion matching, the most

Table 3 Odds ratio for the occurrence of crackles and wheezes in univariable regression models

	Wheezes OR (95% CI)				Crackles OR (95% CI)			
	Any Wheezes (n = 729)	InspiratoryWheezes (n = 151)	Expiratory wheezes (n = 649)	Expiratory wheezes at 2 or more locations (n = 167)	Any crackles (n = 532)	Inspiratory crackles (n = 495)	Expiratory crackles (n = 79)	Inspiratory crackles at 2 or more locations (n = 135)
Age (10 years)	1.2*** (1.1–1.2)	1.1 (0.9–1.2)	1.2*** (1.1–1.3)	1.2** (1.1–1.4)	1.7*** (1.6–1.9)	1.8*** (1.6–2.0)	1.4** (1.1–1.8)	2.2*** (1.8–2.8)
Female	1.3*** (1.1–1.6)	1.2 (0.8–1.6)	1.4*** (1.2–1.7)	1.5* (1.1–2.1)	1.1 (0.9–1.3)	1.1 (0.9–1.4)	1.7* (1.1–2.7)	1.1 (0.8–1.5)
BMI > 30	0.8* (0.7–1.0)	0.8 (0.5–1.1)	0.8 (0.7–1.0)	0.9 (0.6–1.3)	1.2 (1.0–1.5)	1.2* (1.0–1.5)	0.6 (0.3–1.1)	1.5* (1.0–2.2)
Hypertension	1.0 (0.8–1.2)	1.2 (0.8–1.7)	1.0 (0.8–1.2)	1.2 (0.8–1.6)	1.3* (1.0–1.6)	1.3* (1.1–1.6)	0.9 (0.5–1.5)	1.2 (0.8–1.7)
Myocardial Infarction	1.3 (0.9–1.9)	1.5 (0.8–2.7)	1.3 (0.9–1.9)	1.4 (0.7–2.4)	1.7** (1.2–2.5)	1.7** (1.2–2.5)	1.0 (0.3–2.5)	1.8 (0.9–3.3)
Heart Failure	1.5 (0.7–2.8)	1.7 (0.4–4.7)	1.3 (0.6–2.6)	0.5 (0.0–2.2)	0.9 (0.4–2.0)	1.0 (0.4–2.2)	< 0.1 (< 0.1 - > 100)	0.6 (0.0–2.9)
Atrial fibrillation	1.1 (0.7–1.6)	1.3 (0.6–2.4)	0.9 (0.6–1.4)	1.1 (0.5–2.2)	1.4 (0.9–2.0)	1.4 (0.9–2.2)	0.3 (< 0.1–1.2)	1.7 (0.8–3.2)
COPD	1.3 (0.9–1.9)	1.5 (0.7–2.9)	1.3 (0.8–1.9)	1.6 (0.8–2.9)	2.3*** (1.5–3.2)	2.4*** (1.6–3.4)	1.6 (0.6–3.7)	5.6*** (3.3–8.9)
Asthma	1.5** (1.1–1.9)	1.5 (0.8–2.4)	1.4* (1.0–1.8)	1.8* (1.1–2.8)	1.6** (1.1–2.1)	1.4* (1.0–2.0)	1.8 (0.9–3.4)	1.7* (1.0–2.8)
Rheumatoid Arthritis	1.2 (0.8–1.7)	1.2 (0.6–2.2)	1.1 (0.7–1.5)	0.7 (0.3–1.5)	1.8** (1.3–2.6)	1.9** (1.3–2.7)	1.6 (0.6–3.5)	1.6 (0.8–3.6)
Current smoker	1.6*** (1.3–2.1)	2.1** (1.3–3.3)	1.6** (1.2–1.8)	1.7* (1.0–2.6)	2.0*** (1.5–2.6)	2.0*** (1.5–2.7)	1.3 (0.6–2.5)	2.5** (1.5–2.6)
Previous smoker	1.2 (1.0–1.4)	1.1 (0.8–1.6)	1.1 (1.0–1.2)	1.2 (0.8–1.6)	1.4** (1.2–1.8)	1.5*** (1.2–1.9)	0.9 (0.6–1.5)	2.0** (1.3–2.1)
mMRC ≥ 2	1.5* (1.0–2.0)	2.0* (1.1–3.5)	1.4 (0.9–1.9)	1.5 (0.7–2.6)	2.6*** (1.8–3.6)	2.7*** (1.9–3.7)	1.7 (0.7–3.7)	3.4*** (1.9–5.5)
Breathing worse than usual	1.1 (0.8–1.4)	1.4 (0.9–2.2)	1.0 (0.7–1.3)	1.0 (0.6–1.5)	1.0 (0.8–1.4)	1.1 (0.8–1.4)	0.8 (0.3–1.6)	1.7* (1.0–2.6)
Symptoms of airways infection last week	1.1 (0.9–1.4)	1.4 (0.9–2.1)	1.0 (0.8–1.3)	1.4 (0.9–2.0)	0.8 (0.6–1.1)	0.9 (0.6–1.1)	0.8 (0.4–1.5)	1.2 (0.8–1.9)
Oxygen Saturation SpO_2 (1%)	1.0 (0.9–1.0)	0.9 (0.8–1.0)	1.0 (0.9–1.0)	0.9 (0.9–1.1)	0.8*** (0.7–0.9)	0.8*** (0.7–0.8)	1.0 (0.9–1.3)	0.7*** (0.6–0.8)
FEV_1 Z-score (1 unit)	0.9** (0.8–1.0)	0.7*** (0.6–0.8)	0.9* (0.8–1.0)	0.8* (0.7–1.0)	0.8*** (0.7–0.9)	0.8*** (0.7–0.9)	0.9 (0.7–1.1)	0.6*** (0.5–0.7)

Confidence intervals shown in brackets. mMRC = Modified Medical Research council questionnaire. FEV1 = Forced expiratory volume in 1 s. LLN = Lower Limit of Normal
****p value < .001, **p value < .01, *p value < .05*

Table 4 Odds ratio for the occurrence of crackles and wheezes in multivariable regression models

	Inspiratory wheezes (n = 130)	Expiratory wheezes (n = 587)	Expiratory wheezes 2+ (n = 151)	Inspiratory crackles (n = 445)	Expiratory crackles (n = 70)	Inspiratory crackles 2+ (n = 118)
Age (×0.1)	–	1.2*** (1.1–1.3)	1.2** (1.1–1.5)	1.8*** (1.6–2.0)	1.5** (1.2–1.9)	2.2*** (1.8–2.9)
Female Gender	–	1.5*** (1.2–1.8)	1.5* (1.1–2.1)	–	–	–
COPD	–	–	–	–	–	2.3** (1.3–4.1)
Asthma	–	1.4* (1.0–1.8)	1.9** (1.1–3.0)	–	–	–
Rheumatoid Arthritis	–	–	–	1.6* (1.1–2.3)	–	–
Current smoker	–	1.7*** (1.3–2.2)	–	1.9*** (1.4–2.7)	–	–
Previous smoker	–	1.1 (0.9–1.3)	–	1.3* (1.0–1.6)	–	–
mMRC ≥ 2	–	–	–	1.8** (1.2–2.6)	–	–
Oxygen saturation SpO_2 (1%)	–	–	–	0.9** (0.8–1.0)	–	0.7*** (0.6–0.8)
FEV_1 Z-score (1 unit)	0.7*** (0.6–0.8)	–	–	0.9** (0.8–1.0)	–	0.7*** (0.6–0.8)
AUC	.59 (0.54–0.64)	.59 (0.56–0.62)	.60 (0.55–0.64)	.69 (0.67–0.72)	.62 (0.56–0.69)	.79 (0.75–0.84)

Confidence intervals shown in brackets. 2+ = presence of the adventitious sounds in more than two locations. *mMRC* = Modified Medical Research council questionnaire. *FEV1* = Forced expiratory volume in 1 s, *AUC* = Area under the curve
***p value < .001, **p value < .01, *p value <.05

common cause of hypoxemia, which could explain the relationship in our study. [37].

Strengths and limitations

To our knowledge, this is the largest sample characterizing the occurrence of wheezes and crackles to date. Tromsø 7 had a high response rate (65%). The study has a high external validity for the Norwegian population. [23] Nevertheless, our results might not be valid in other populations, for instance in those with poorer health. Unhealthy people may be underrepresented since some might have chosen to refrain from participation or were not able to attend and complete the survey.

The questionnaires employed at Tromsø 7 did not ask about the presence of interstitial lung disease and bronchiectasis. Both conditions have an increased prevalence with age [45, 46] and are associated with the presence of crackles [6, 7]. It is possible that participants with these conditions were categorized as apparently healthy and this constitutes a limitation of our study.

All the selection processes were randomized and took place prior to the classification of the recordings without

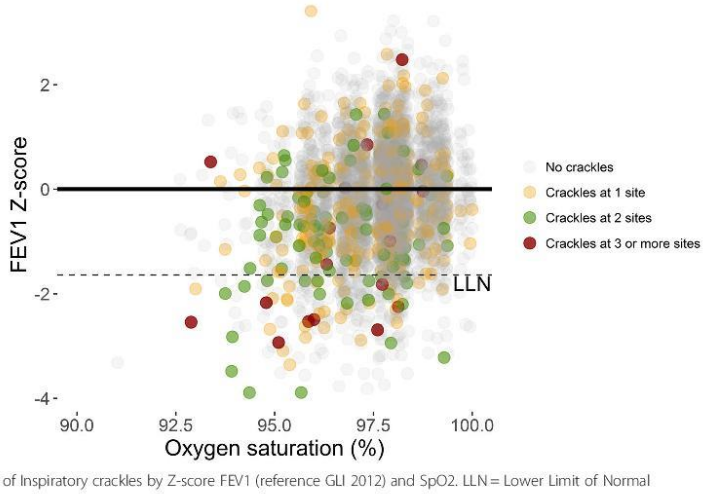


Fig. 3 Occurrence of Inspiratory crackles by Z-score FEV1 (reference GLI 2012) and SpO2. LLN = Lower Limit of Normal

having any knowledge of the health status of the participants. When randomizing for the second visit a higher representation was chosen among those aged 60 years or more, and the subjects invited due to participation in previous surveys of the Tromsø study were usually 60 years or older. In terms of prevalence, we have taken care of this selection bias by age standardization, but some influence on associations with self-reported diseases and lung function cannot be excluded.

The inter-observer agreement at the first step of our classification compares to that found among general practitioners. [47] The repeated independent classifications have without doubt increased the reliability. [48] A lack of reliability could have influenced our results by diluting the strength of the estimates. At the third step of the classification, we chose to discard as positive findings the recordings marked as "possible" by two observers. In a sensitivity analysis, they were classified as "present" but the results were similar, and our conclusions unchanged.

Conclusion

Our findings support a cautious attitude when using ALS to diagnose lung disease in elderly patients. The presence of wheezes or crackles in one lung location did not strongly predicted the outcomes analyzed. Nonetheless, it is possible that these solitary findings are a manifestation of lung senescence and/or represent subclinical disease in apparently healthy subjects. However, when inspiratory crackles at two or more locations or both wheezes and crackles are heard, risk of decreased lung function increases considerably. Such findings, particularly when unexpected in a patient, should lead to further investigation regarding possible heart or lung disease.

Additional files

Additional file 1: Table S1. Comparison between all the participants attending the 7th Survey of the Tromsø Study and the final sample included in this article. (DOCX 12 kb)

Additional file 2: Figure S1. Flow diagram of the participants included in our analyses. (JPG 292 kb)

Additional file 3: Figure S2. Spirogram and clinical information of four participants with presence of adventitious lung sounds in the recordings. (PPTX 2605 kb)

Abbreviations

ALS: Adventitious Lung Sounds; AUC: Area Under the Curve; COPD: Chronic Obstructive Pulmonary Disease; FEV_1: Forced Expiratory Volume in 1 s; GLI: Global Lung Initiative; LLN: Lower Limit of Normal; mMRC: Modified Medical Research Council dyspnea scale; OR: Odds Ratio; RA: Rheumatoid Arthritis; SpO_2: Arterial oxygen saturation

Acknowledgements

We would like to thank Katrine Wang, Gøril Henriksen and Lars Tunby for their help with data collection. We would also like to thank all participants in the Tromsø 7th study, as well as the staff, to make this research possible. The publication charges for this article have been funded by a grant from the publication fund of UiT The Arctic University of Norway.

Ethical approval and consent to participate

All study participants provided written consent. The Regional Committee for Medical and Health Research Ethics in North Norway approved the study.

Author contributions

MH conceived the research idea. ASJC, DA and ER had responsibility for data collection. ASJC, JC, DA, ER, PH and MH participated in the classification of lung sounds. ASJC, MH and SV had responsibility for the data analysis. ASJC and MH had the responsibility for the main drafting of the manuscript. All authors contributed substantially to the interpretation of the results, and completion of the manuscript. All authors approved the final manuscript.

Funding

UiT, The Arctic University of Norway. The funding source had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The data that support the findings of this study are available from The Tromsø Study but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Information and guidelines to apply for access to data from The Tromsø study are available at https://en.uit.no/prosjekter/prosjekt?p_document_id=71247.

Consent for publication

Written informed consent was obtained from the patient for publication of Fig. 1. A copy of the written consent is available for review by the Editor of this journal.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]General Practice Research Unit, Department of Community Medicine, UIT the Arctic University of Norway, Tromsø, Norway. [2]CINTESIS - Center for Health Technology and Services Research, Faculty of Medicine, University of Porto, Porto, Portugal. [3]Department of methodology and statistics, University of Maastricht, Maastricht, The Netherlands. [4]Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Canada.

Received: 27 February 2019 Accepted: 26 August 2019
Published online: 11 September 2019

References

1. van der Wall EE. The stethoscope: celebration or cremation after 200 years? Neth Heart J. 2016;24(5):303–5.
2. Edelman ER, Weber BN. Tenuous tether. N Engl J Med. 2015;373(23): 2199–201.
3. Pasterkamp H. The highs and lows of wheezing: a review of the most popular adventitious lung sound. Pediatr Pulmonol. 2018;53(2):243–54.
4. Jacome C, Marques A. Computerized respiratory sounds in patients with COPD: a systematic review. COPD. 2015;12(1):104–12.
5. Bennett S, Bruton A, Barney A, Havelock T, Bennett M. The relationship between crackle characteristics and airway morphology in COPD. Respir Care. 2015;60(3):412–21.
6. Bohadana A, Izbicki G, Kraman SS. Fundamentals of lung auscultation. N Engl J Med. 2014;370(21):2053.
7. Barker AF. Bronchiectasis. N Engl J Med. 2002;346(18):1383–93.
8. 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(3):1468–74.
9. Moore M, Stuart B, Little P, Smith S, Thompson MJ, Knox K, et al. Predictors of pneumonia in lower respiratory tract infections: 3C prospective cough complication cohort study. Eur Respir J. 2017;50(5).
10. Murphy RL. In defense of the stethoscope. Respir Care. 2008;53(3):355–69.
11. Norisue Y, Tokuda Y, Koizumi M, Kishaba T, Miyagi S. Phasic characteristics of inspiratory crackles of bacterial and atypical pneumonia. Postgrad Med J. 2008;84(994):432–6.
12. Jakobsen KA, Melbye H, Kelly MJ, Ceynowa C, Molstad S, Hood K, et al. Influence of CRP testing and clinical findings on antibiotic prescribing in

adults presenting with acute cough in primary care. Scand J Prim Health Care. 2010;28(4):229–36.
13. Strumilo J, Chlabicz S, Pytel-Krolczuk B, Marcinowicz L, Rogowska-Szadkowska D, Milewska AJ. Combined assessment of clinical and patient factors on doctors' decisions to prescribe antibiotics. BMC Fam Pract. 2016;17:63.
14. Turnbull S, Lucas PJ, Redmond NM, Christensen H, Thornton H, Cabral C, et al. What gives rise to clinician gut feeling, its influence on management decisions and its prognostic value for children with RTI in primary care: a prospective cohort study. BMC Fam Pract. 2018;19(1):25.
15. Badgett RG, Lucey CR, Mulrow CD. Can the clinical examination diagnose left-sided heart failure in adults? JAMA. 1997;277(21):1712–9.
16. Holleman DR Jr, Simel DL. Does the clinical examination predict airflow limitation? JAMA. 1995;273(4):313–9.
17. Metlay JP, Kapoor WN, Fine MJ. Does this patient have community-acquired pneumonia? Diagnosing pneumonia by history and physical examination. JAMA. 1997;278(17):1440–5.
18. Thacker RE, Kraman SS. The prevalence of auscultatory crackles in subjects without lung disease. Chest. 1982;81(6):672–4.
19. Oliveira A, Marques A. Respiratory sounds in healthy people: a systematic review. Respir Med. 2014;108(4):550–70.
20. Workum P, Holford SK, Delbono EA, Murphy RL. The prevalence and character of crackles (rales) in young women without significant lung disease. Am Rev Respir Dis. 1982;126(5):921–3.
21. Kostas NP, Leontios JH, Mark LE. Breath sounds from basic science to clinical Pratice. Cham, Switzerland: Springer; 2018.
22. UiT The Arctic University of Norway [Internet]. The Tromsø Study. 2019. [Cited 03 Aug 2019] Available from: http://tromsoundersokelsen.uit.no/tromso/.
23. Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njolstad I. Cohort profile: the Tromso study. Int J Epidemiol. 2012;41(4):961–7.
24. Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. Chest. 1988;93(3):580–6.
25. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. Eur Respir J. 2005;26(2):319–38.
26. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. Eur Respir J. 2012;40(6):1324–43.
27. Vanbelle S. Asymptotic variability of (multilevel) multirater kappa coefficients. Stat Methods Med Res. 2018;[Epub ahead of print. https://doi.org/10.1177/962280218794733.
28. Statistics Norway [Internet]. Tromsø municipality facts (Age distribution per 1. januar 2018). 2018. [Cited 12 Jan 2018]. Available from: https://www.ssb.no/kommunefakta/tromso.
29. Dalbak LG, Straand J, Melbye H. Should pulse oximetry be included in GPs' assessment of patients with obstructive lung disease? Scand J Prim Health Care. 2015;33(4):305–10.
30. Global Initiative for Prevention Diagnosis and Management of COPD (2018). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. 2018 [Online] Available from: https://goldcopd.org/gold-reports/ . [Accessed 03 Aug. 2019].
31. Xavier Robin NT, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
32. Fox J, Weisberg S. An {R} companion to applied regression. 2nd ed. Thousand Oaks, California, USA: SAGE Publications; 2011.
33. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.
34. Kiyokawa H, Greenberg M, Shirota K, Pasterkamp H. Auditory detection of simulated crackles in breath sounds. Chest. 2001;119(6):1886–92.
35. Kataoka H, Matsuno O. Age-related pulmonary crackles (rales) in asymptomatic cardiovascular patients. Ann Fam Med. 2008;6(3):239–45.
36. Murphy RL Jr, Gaensler EA, Holford SK, Del Bono EA, Epler G. Crackles in the early detection of asbestosis. Am Rev Respir Dis. 1984;129(3):375–9.
37. Janssens JP, Pache JC, Nicod LP. Physiological changes in respiratory function associated with ageing. Eur Respir J. 1999;13(1):197–205.
38. Wilsher M, Voight L, Milne D, Teh M, Good N, Kolbe J, et al. Prevalence of airway and parenchymal abnormalities in newly diagnosed rheumatoid arthritis. Respir Med. 2012;106(10):1441–6.
39. van Riet EE, Hoes AW, Limburg A, Landman MA, van der Hoeven H, Rutten FH. Prevalence of unrecognized heart failure in older persons with shortness of breath on exertion. Eur J Heart Fail 2014;16(7):772–7.
40. van Vugt SF, Verheij TJ, de Jong PA, Butler CC, Hood K, Coenen S, et al. Diagnosing pneumonia in patients with acute cough: clinical judgment compared to chest radiography. Eur Respir J. 2013;42(4):1076–82.
41. Arathimos R, Granell R, Henderson J, Relton CL, Tilling K. Sex discordance in asthma and wheeze prevalence in two longitudinal cohorts. PLoS One. 2017;12(4):e0176293.
42. Mandhane PJ, Greene JM, Cowan JO, Taylor DR, Sears MR. Sex differences in factors associated with childhood- and adolescent-onset wheeze. Am J Respir Crit Care Med. 2005;172(1):45–54.
43. Fu L, Freishtat RJ, Gordish-Dressman H, Teach SJ, Resca L, Hoffman EP, et al. Natural progression of childhood asthma symptoms and strong influence of sex and puberty. Ann Am Thorac Soc. 2014;11(6):939–44.
44. Vyshedskiy A, Alhashem RM, Paciej R, Ebril M, Rudman I, Fredberg JJ, et al. Mechanism of inspiratory and expiratory crackles. Chest. 2009;135(1):156–64.
45. Olson AL, Gifford AH, Inase N, Fernandez Perez ER, Suda T. The epidemiology of idiopathic pulmonary fibrosis and interstitial lung diseases at risk of a progressive-fibrosing phenotype. Eur Respir Rev. 2018;27(150).
46. O'Donnell AE. Bronchiectasis Chest. 2008;134(4):815–23.
47. Aviles-Solis JC, Vanbelle S, Halvorsen PA, Francis N, Cals JWL, Andreeva EA, et al. International perception of lung sounds: a comparison of classification across some European borders. BMJ Open Respir Res. 2017;4(1):e000250.
48. Kraemer HC, Bloch DA. Kappa coefficients in epidemiology: an appraisal of a reappraisal. J Clin Epidemiol. 1988;41(10):959–68.

Publisher's Note

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