A STUDY OF ADVENTITIOUS BREATH SOUNDS:
INTERRATER AGREEMENT

by

Julie R. Wellman

A Thesis

Presented to

The Oregon Health Sciences University
School of Nursing
in partial fulfillment
of the requirements for the degree of
Master of Nursing

June 1985

APPROVED:

[redacted]

Mary McFarland, R.N., M.S.N., Associate Dean for Undergraduate Studies, Thesis Advisor

[redacted]

Carol Burckhardt, R.N., PH.D., Assistant Professor, Adult Health and Illness, First Reader

[redacted]

Linda K. Fahr, R.N., M.N., Pulmonary Clinical Nurse Specialist, Visiting Nurses Association and Rehabilitation Service, Second Reader

[redacted]

Carol A. Lindeman, R.N., Ph.D., Dean, School of Nursing

ACKNOWLEDGEMENTS

This Thesis is dedicated to Justin, Meredith and Janelle. May you also find the love of learning and the support and encouragement of your teachers. I am indebted to Christine Tanner, Mary McFarland, Carol Burckhardt and Linda Fahr for the completion of this Thesis.

I appreciate the interest and time of the nurse rater's who worked together on this project. Finally, I thank Carol Duncan for the generous sharing of her materials.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

In 1975 the American College of Chest Physicians and American Thoracic Society Joint Committee on Pulmonary Nomenclature recommended standardized terms for adventitious pulmonary lung sounds auscultated during the physical examination of the chest. Terminology confusion had been traced to Laennec, the inventor of the stethoscope, and to his many translations of the term "rale". First, rale was translated into Latin as rhonchus and later Laennec translated rhonchus back into English as wheeze. Laennec's use of terms was general; however, some clinicians chose to use rhonchus to mean wheeze, while other clinicians used the traditional general meaning. Clinicians still vary on term usage, and terms for pulmonary sounds lack a physiologic theoretical base.

To add precision to the communication of clinical findings, the terms crackle (rale) and wheeze (rhonchus) were recommended by the Joint Committee on Pulmonary Nomenclature. These terms are based upon the works of Forgacs, who provided a more sound physiological basis for the terms (Forgacs & Lond, 1967). With more descriptive terminology it was hoped more valid indices for auscultatory pulmonary findings would be found.

A nursing trend toward the use of adventitious breath sounds to determine patient lung changes and appropriate nursing interventions is underway. In 1981, Schare and

Stehlin wrote of interventions nurses should use when certain adventitious sounds are heard in patients. Nursing research in the use of pulmonary sounds is occurring in the midst of the process of changing from old terminology to the new standardized terminology. Unless the nursing profession uses the new terminology, the results of nursing research based on inexact terms and definitions may not be widely accepted or applied.

To facilitate this trend in nursing research the new terminology needs to be firmly established in the nursing profession and used for the basis of communication in this area. The first step is to show that with adequate nomenclature instruction and uniform reporting methods, different nurses listening to the same adventitious sounds will report the same results. Only from this starting point will research into the use of adventitious sounds as predictors for the need for nursing action have the reliability and the validity on which any widespread application is dependent.

Review of the Literature

The review of the literature focuses on the following areas of study:

1. the mechanism and production processes that affect transmission of normal, abnormal and adventitious lung sounds;
2. the observer variability in detecting and discriminating lung sounds; and

3. teaching methods that facilitate content learning.

This review relies on medical literature as the primary source for consideration due to the paucity of nursing studies reported in the nursing literature.

Mechanism and Production Processes of Lung Sounds

The following section relates current theory of the genesis of pulmonary breath sounds and conditions that affect respiratory sound transmission.

All breath sounds may be classified as either normal, abnormal or adventitious sounds. Though there are different mechanisms which produce these sounds, the following anatomical structures and composition of the pulmonary system contribute to sound production. The tree-like pulmonary conducting system branches approximately twenty-three times before reaching the bronchioles. The trachea is made of oval cartilagenous ringlike structures which surround the anterior surface. However, the posterior surface is made of membrane, and under certain conditions this membrane may collapse inward. The larger airways are made of cartilagenous material. As the airways become smaller the structures become supplanted by elastic connective tissue and disappear at the level of the terminal bronchioles. Under certain conditions the lung airways deflate (Price & Wilson, 1978).

It is thought that turbulence is responsible for sound generation of normal and abnormal breath sounds from the respiratory system. Laminar flow of gas through airway passages produces the turbulence. Turbulence is generated by a critical minimal velocity at which the gas molecules of respiration (at a certain density) flow across pulmonary structures. Sound generation of normal and abnormal breath sounds are modified by the flow rate, volume and density of the gas and by the patency, caliber and length of the airway (Forgacs & Lond, 1967).

Characteristically, normal breathing is never completely silent. It is audible only with the ear close to the patient's mouth or with the aid of the stethoscope over the trachea. The sound is like a hiss of no definite pitch and its energy is evenly distributed over a wide range of frequencies. This sound is called white noise (Forgacs, 1969). The source of turbulence creating white noise is generated at least as far peripherally as the main bronchi and possibly in more peripheral branches as well. Respiratory white noise becomes loud enough to be heard at a distance during high flow rates. The sound frequency of normal respiration ranges between 200 and 2000 cycles/second.

At the mouth, sounds from the proximal airways are not filtered. However, transmission of lung sounds through the chest wall and surrounding tissue affects both filtering and sound intensity. Sound frequency over the right base ranges from 200 to 400 cycles/second with higher frequencies being filtered out.

Abnormal breath sounds occur when there are changes affecting the lung structures or when there is alteration of the transmitting properties of the lung tissue or chest wall. They are caused by changes in the intensity of the source of the sound. For example, bronchitis increases the sound intensity due to narrowing of the proximal airways; lung tissue diseased with emphysema will cause poor transmission of sound due to the dampening effect of air trapping (Forgacs & Lond, 1967).

Another property affecting abnormal lung sounds is gas density. A reduction of the density of the gas molecules will reduce the intensity of the sounds made from the origin of the turbulence (Forgacs & Lond, 1967). Auscultation of breath sound intensity has also been tested as a predictor of airflow obstruction. Bohadana, Preslin, and Uffholtz (1978), using a grade-rating scale on 34 inpatients with varying degrees of obstruction, found a positive correlation between breath sound intensity and obstruction during forced expirations.

According to Forgacs (1969), the genesis of adventitious lung sounds differs from normal and abnormal breath sounds. Adventitious lung sounds are generally considered to be abnormal sounds produced by movement of air in the lung. Adventitious sounds are divided into two categories each having different origins. The wheeze is an example of a continuous adventitious lung sound. One theory of production is that the wheeze is produced in a simple uncoupled reed-like system (e.g. toy trumpet). An airway narrowed to the point of closure, perhaps by intralumen edema or loss of elasticity, produces this musical sound. Providing the air flow pressure exceeds a critical minimum value, the walls of the airway oscillate between the closed and barely open positions (Forgacs, 1969; 1978).

Wheezes have been characterized as monophonic (consisting of a single note or several notes starting and ending at different times) and polyphonic (containing several notes starting and ending simultaneously, like a chord). These sounds may be classified as high or low pitched, inspiratory or expiratory, short or long (Forgacs, 1978). Wheezes tend to be transmitted throughout the thorax. For example, single or multiple monophonic wheezes are clinical signs of asthma. Though the number of wheezes may not be large, the illusion of innumerable wheezes are heard wherever the stethoscope is applied to the chest wall (Forgacs, 1978).

The clinical significance of wheezing is uncertain since healthy individuals can produce a wheeze upon forced expiration. Though unreliable, a wheeze is characteristic of airway obstruction. Marini, Pierson, Hudson, and Laksmhminarayan (1979) studied expiratory wheezing in patients with airway obstruction (N=83). Using a grade-rated scale for reporting auscultated findings, they reported a high correlation between wheezing scores and degree of obstruction documented by spirometry.

Another adventitious sound generated is a discontinuous sound of short duration called a crackle (rale). Forgacs and Lond, (1967) suggest that crackling is due to explosive equalization of pressure that follows the sudden removal of a barrier separating two peripheral airway compartments containing gas at widely different pressures. Crackles may be described as high or low in pitch, loud or faint in sound intensity, scanty or profuse in number and located in the inspiratory or expiratory cycle or both. The timing characteristics of crackle occurrence, such as random or regular spacing in the respiratory cycle, have been studied for clinical significance (Nath & Capel, 1974; 1980).

Forgacs (1978) classified the timing of the crackles in the respiratory cycle as follows:

1. Early inspiratory crackles; the crackles that occur early in inspiration are thought to be due to the sudden opening of proximal airways. The sudden equalization of pressure possibly causes the crackle sound.
2. Both inspiratory and expiratory crackles; crackles that are found in a random pattern are attributed to secretions in the airways. With sequential respiratory cycles the secretions may change the order of the opening of the airways due to transpulmonary pressure causing the random production of crackle sounds.
3. Late inspiratory crackles; the crackles

heard in late inspiration are thought to be due to the opening of peripheral airways caused by the changes in transpulmonary pressure. In deflated areas of the lung, peripheral airways remain closed until late inspiration. By then the pressure in the alveoli supplied by these airways is below atmospheric pressure. The sudden equalization of pressure in the airway produces a crackling sound.

The following studies support Forgacs theory of the genesis of adventitious lung sounds. In 1974, Nath ans Capel studied 2 groups of patients. The first group was known to have inspiratory crackles (n=56) and the second group (n=44) had airway obstruction as documented by spirometry. The purpose of the study was to find if the timing of inspiratory crackles had clinical and physiological significance. The subject's inspiratory crackles and airflow rate were measured simultaneously by waveform analysis and spirometry. The findings associated early inspiratory crackles with severe airway obstruction and late inspiratory crackles with obstructive disease. The important clinical distinctions reported from their findings follow.

Early crackles are generally scanty in number, located in lower zones, heard at one or both lung bases and are usually transmitted to the mouth. They are not silenced by cough or by change in posture. Crackles associated with secretions in the larger airways are modified or silenced by coughing. They are present in both inspiration and expiration and occur in a more random sequence within the respiratory cycles. In contrast, late inspiratory crackles are usually more profuse. They are heard best over the lower lung areas but in advanced disease they may be heard in the middle and upper lung areas. They are rarely transmitted to the mouth. Depending upon the stage of disease, posture may have an effect upon them. In late stages of disease they are independent of posture.

This important study outlines the clinical helpfulness of timing crackles in the respiratory cycle. Information such as numbers of crackles, distribution and transmission, as well as the effect of cough and change in posture may provide information regarding the functional state of the lung airways. The use of validating instruments to measure the crackles, however, does not suggest the degree of skill a clinician must have to be able to make valid auscultative findings with a stethoscope.

In 1980, Nath and Capel further studied the clinical relevance of the timing of crackles as a physical sign. Three diagnostic groups (N=38) composed respectively of chronic bronchitis, alveolitis and bronchiectasis were studied to compare crackle distribution in the respiratory cycle pattern among the groups. Crackle distribution occurred in different and distinct patterns among the groups. The crackle patterns of chronic bronchitis are typically confined to early inspiration, while in alveolitis the lung crackles continue to the end of inspiration but may begin in the early or the mid-phase of inspiration. It was suggested that the crackles of bronchiectasis relate to retained secretions in damaged and dilated bronchi. These crackles are present in both inspiratory and expiratory phases of the respiratory cycle and tend to be random in occurrence. They are less abundant after coughing. Finally, they are well transmitted to the mouth.

In 1982, Worklum, Holford, Delbono, and Murphy, studied 56 women who did not have significant lung disease. The purpose of the study was to ascertain if crackles were present in normal persons, as diagnostic and therapeutic decisions may be made on the basis of the presence of crackles. They found mid-inspiratory fine crackles at the lung bases in 35 subjects. The pulmonary specialist's findings were validated by a bioengineer using an 800 Hertz high pass filtered stethoscope, who heard crackles in 53 basilar areas of the 35 subjects. All subjects were asked to take slow inspirations from a residual volume. It was thought that the crackles were nonpathologic and occurred when basilar airways, which close at the end of a forced expiration, suddenly open during inspiration. The examination of the quality, the timing and the anatomical distribution of the crackles in apparently normal subjects suggest that they can be distinguished from those of patients with diseases by a pulmonary specialist. The amount of instruction required to be a pulmonary specialist was not discussed.

In 1984, Murphy, Gaensler, Holford, Del Bono and Epler studied crackles as early physical signs of asbestosis. Extensive auscultation instruction for a technician, with the use of waveform analysis and simultaneous recordings of crackles was done. The results suggested that auscultation by a trained technician could be effective in screening large industrial populations at risk for the disease. The subjects' basilar lung regions were auscultated and the subjects were asked to open mouth breathe slightly more deeply than normal. Of the sample tested (N=386), 6.2% were found to have crackles; these findings were validated by waveform analysis.

Currently, nursing actions based on clinical findings of adventitious sounds are largely dependent upon the amount of skill and experience a nurse brings to the patient. Schare and Stehlin (1981) write of respiratory nursing care interventions based on the respiratory patterns of patients with bronchitis, pneumonia and atelectasis. The article uses old descriptive terminology in an attempt to help define for the reader what sounds will be heard in patients with certain underlying pathology. When moisture is heard in the smaller airways, it is suggested that postural drainage and clapping be used to move the secretions to larger airways, prior to suctioning. If suctioning is the appropriate method for removal of the secretions, its effectiveness may depend upon the accessibility of the secretions. The nurse is to encourage the patient to deep breathe. If moisture is present in the larger airways, the nurse should encourage the patient to perform deep breathing exercises and coughing; suctioning may also be needed. The authors caution against using suctioning when musical wheezes are present as they do not signify the presence of secretions. However, when low-pitched wheezes are heard it is assumed secretions, or swelling in larger airways may be the cause. According to Bates (1983) and Forgacs (1978) the pitch of a wheeze cannot indicate the type of airway affected. Schare and Stehlin make a strong case for nurses to be aware of the sounds they are auscultating and to infer from these sounds what underlying physiological mechanisms may be associated with them. Since their assumptions may be incorrect, relating this information to patient care without further study to establish validity and reliability of these inferred relationships is insupportable.

Variables that Affect Transmission of Lung Sounds.

The pulmonary sounds can be magnified or changed by the density of surrounding tissue (Forgacs, Nathoo & Richardson, 1971). It is important to remember that in the transmission of sound through the chest wall, the surrounding tissue affects sound intensity (Ploy-Song-Sang, Martin & Ross, 1977). Another variable is the velocity of the airflow. As the velocity of airflow increases, the turbulence in the lung increases, resulting in greater intensity

(Banazek, Kory & Snider, 1973). The lung tissue acts as a selective filter and the human ear also compounds filtering of high frequency and low intensity sound.

The literature suggests that varying air flow rates, lung sizes, chest wall thicknesses, air passage constriction, pathology and interobserver differences influence auscultatory findings. The use of standard terminology and standard procedures for pulmonary assessment in correlating disease processes to auscultatory clinical findings is discussed below.

Observer Variability of Lung Sounds

Reliability of pulmonary clinical findings may suggest the presence or absence of disease or disease processes. However, Koran (1975) reminds clinicians of the difference between reliable findings and findings that are accurate. For example, clinicians have reported the same subjective findings, yet have been wrong when a different validating measure was used. Therefore, an objective perspective must be maintained when evolving clinical subjective judgments. Considering this perspective, studies showing agreement and reliability of pulmonary sounds will be summarized for similar method characteristics.

Hudson, Conn, Matsubara and Pribble (1978) used a standardized procedure, a qualitative description and the timing of the respiratory cycle for the basis of reporting rales (crackles). The purpose was to study both the reliability of the observers and to see if the measures were valid indices of disease categories. Three physician observers who had no prior knowledge of the admission diagnosis auscultated patients who were admitted to the hospital. All patients (N=100) had rales. Excluded from this study were patients in the coronary care unit and infectious isolation patients. Results showed that agreement between the observers using qualitative adjectives (fine to coarse, moist to dry) occurred only in 47%, while agreement between the observers in the timing of the rales occurred 85% of the time. These findings suggested the timing of discontinuous inspiratory rales (crackles) is more reliable when observers correlate the rales with the respiratory cycle and that disagreement occurs more often when descriptive terms are used.

To assess the validity of clinical signs of diminished or absent breath sounds as an indicator of poorly ventilated lungs, Nairn and Turner-Warwick (1969) compared the auscultatory findings of two physicians and correlated the results to radioactive Xenon 133 pulmonary tests. The observers graded the intensity of breath sounds of patients (N=34) with emphysema. The examined standardized chest areas with the bell or diaphragm of the stethoscope, and rated the auscultated breath sounds using a predetermined graded scale. The total scores were then correlated to the results of pulmonary tests. The correlation of both anterior and posterior lung sounds to the pulmonary tests was significant. Findings indicated interrater reliability.

One observer did tend to rate consistently lower and bias was suggested. However, results suggest that diminished or absent breath sounds were indicators of a poorly ventilated lung.

The Pardee, Martin and Morgan (1976) study compared the standardized physical examinations of four examiners on a group of patients who had undergone ventilatory tests. Patients were instructed to take deep inspirations. The diaphragm of the stethoscope was used to auscultate the lungs. Findings from 184 patients were graded and then correlated with their ventilatory function. Correlation of breath-sound scores with percentage of predicted forced expiratory volume in one second (FEV_1, percent predicted) was positive for all of the examiners (ranging from $r = +.57$ to $r = +.68$). The positive correlations suggested that the observer's gradings were valid.

The above findings suggest that clinicians using a standardized procedure and conforming to a nomenclature reporting system can independently examine patients (before the patients condition changes) with concurrent validating measures and obtain a degree of agreement and reliability.

Most older studies have shown disagreement. In a 1955 study, Schilling, Hughes and Dingwall-Fordyce examined 187 male patients with respiratory disease and a control group consisting of 88 men. While there was a 93% agreement between 2 physician examiners on the diagnosis based on the history alone, there was disagreement over clinical data including abnormal chest sounds. Even though the study took place over four months, the examination findings of each patient were compared. The two physicians were unaware of each other's judgments; however, the researchers built a case for systematic bias between the two raters, as the A rater consistently rated in favor of abnormal findings. It was also possible that observer error was increased due to physician inexperience. However, skilled physicians have been reported to make observation errors (Fletcher, 1952). Another possible influencing factor was thought to be observer fatigue. As the study progressed over time fewer abnormal findings were reported.

Fatigue was thought to be a contributing factor in the study by Schneider and Anderson (1965). In an effort to correlate clinical signs with ventilatory function in obstructive lung disease, nine physicians independently auscultated thirteen patients; two had emphysema and eleven had chronic irreversible obstructive lung diseases. To control for variation in the subjects, patients who had reversible obstruction and congestive heart failure were excluded. Diagnosis was made from forced expiratory volume and maximum voluntary ventilation. Each physician examined as many patients possible in a two hour period. They noted the presence or absence of seventeen signs. No prior agreement was given regarding observational criteria for each sign, nor were the physicians given the patients physiologic test results. Physicians were unable to examine all of the patients and it is possible that fatigue may have occurred. A striking feature of the findings was the amount of variation of physician agreement not only in diagnosing emphysema but in their abilities to detect specific signs. There was overall disagreement about coarse rales and decreased breath sounds findings in 29% and 32% of the total 84 examinations, respectively. The variation of disagreement of rales found for individual patients was as high as 50%. The small sample size makes it difficult to generalize findings.

The above studies suggest that an agreed upon uniform rating system and a standard procedure used within a limited time frame may increase reliability in reporting subjective findings.

Method of Learning Breath Sounds

The literature does not document one best method for teaching breath sounds. However, in an attempt to improve instruction of pulmonary sounds and reduce the amount of time spent in the laboratory with students, Cugell (1971, 1978) suggested a technique borrowed from current heart sound instruction. Auditory lung sounds and coinciding visual cues of respiration were given students during a formal class. Although Cugell believed that complete skill learning requires the auscultation of lungs, the approach permitted a more objective definition of respiration in the classroom setting. Weiss and Carlson's (1972) descriptive report on making lung sound tape recordings for instruction support the use of this instructional aid.

It is not known at what point clinicians achieve auscultation skill mastery. However, there are methods to increase achievement by learning objectives. One method is mastery learning. Kulik, Kulik, and Hertzler (1977) suggest that an instructor set a passing criterion level on a test (e.g. 90%), and require the student to attain it. The level of achievement almost always exceeds the level attained without the mastery requirement.

Thorne (1983) describes the practical application of mastery learning theory by suggesting the following format:

1. Design the course based upon preferred outcomes. Ask, "what are the most important learning outcomes the student should know?";
2. Upon the initial meeting with students ask their expectations of the course;
3. After discussing the expectations set content priorities; and
4. The course content and course tests are based upon the criteria referenced in the objectives. All items should clearly measure the course objectives in proportion to the priority accorded the objectives.

Krumme (1975) discusses the evaluation of performance based upon criterion referenced measures. Because the performance score does not depend upon the person being compared and rated to others, it provides a better measure of whether the person has met performance criteria. Therefore, criterion referenced measures represent what an individual can perform. It may be a more effective way of appraising individual knowledge regarding the performance of a skill in clinical nursing practice.

Conceptual Framework

Clinicians listening to the same adventitious lung sounds may report different findings. The primary reasons for the lack of agreement and reliability among observers may be ambiguous terminology and a lack of uniform reporting methods. Other contributing factors may include the uncertain physiological nature of breath sounds, possible observer auditory differences and inadequate skill instruction and usage. The lack of reliability and agreement in reporting clinical observation may limit the use and effectiveness of auscultation in the clinical setting (Murphy, 1981; Koran, 1975).

To standardize and add precision to breath sound terminology, the Joint Ad Hoc Subcommittee of Pulmonary Nomenclature in America and the American College of Chest Physicians (1975) published a report entitled "Pulmonary Terms and Symbols" which included specific terminology for breath sounds. However, Bunin and Loudon (1979) while comparing 663 case reports from seven journals beginning in July of 1977, found continued use of inexact terms. Many physiologically ambiguous descriptive terms such as "dry", "wet", or "moist" rales were often used. Such terms provide an inadequate basis for another clinician to form any conclusions about the patient's lung sounds. More recently, Wilkens, Dexter and Smith (1984) surveyed eight British and American medical journals for adventitious lung sound terms from 1978 through 1982. Data taken from 590 case reports suggest an increased use of the adopted American Thoracic Society (ATS) terminology. The trends may only reflect editors' responses to the recent discussions of lung sound physiology and terminology and may or may not depict the current usage in the clinical setting. Though current ATS terminology is seen more frequently in the literature, confusion over terms remains part of the problem (Bates, 1983, p.152).

Inadequate instruction has contributed to inconsistent nomenclature and reporting variability among observers. Instruction in auscultation has traditionally been carried on at the bedside with the class listening and the instructor supplying the name of the sound which they heard. The instruction method assumes the student will remember the name for the sound when it is heard again (Louden, 1982). Without a uniform reporting method even greater variation in reported finding occurs (Murphy, 1981).

Another cause of inexact terminology usage is the incomplete understanding of the physiology underlying lung sounds (Forgacs, 1978). A more precise understanding about the nature of lung sounds would improve reporting accuracy and the reliability of observer findings. However inexact, the usefulness of breath sounds in the evaluation of regional lung physiology or pathophysiology is the reason they are used more frequently by clinicians than any other diagnostic device in respiratory care (Koran, 1975). Auscultation is readily accessible in a variety of clinical settings. It is relatively low in cost and of no risk to the patient (Louden, 1982). When nurses use this method of assessment it assists them in: 1) care planning; 2) intervention; 3) evaluation of care effectiveness; and 4) assessment of patient progress (Alexy, 1978).

The nurse is expected to observe for changes in patient signs, interpret and report the findings and use them for planning care. Routine cardiopulmonary assessment is one such example. The daily changes in lung sounds relate to the clinical significance of the disease process (Carrieri, et al., 1982). Adventitious breath sounds, have traditionally been used by nurses as an indicator of the need to suction a patient (Grossbach-Landis & McLane, 1979).

In an attempt to find clinical indicators of the need to suction patients, Amborn (1976) studied 35 intubated subjects requiring suctioning. The researcher correlated the amount of suctioned secretions to 22 clinical signs. The data showed that as the frequency of the number of relevant clinical signs changed from patient baseline measures, the amount of secretions obtained increased. These data suggest that increases in the change in the number of indicators from baseline data may predict the need to suction the patient. Fifteen signs were found to be possible predictor indicators. Coarse breath sounds and prolonged expiratory sounds were among them. The ability to distinguish changes in breath sounds from baseline measures may therefore have important nursing implications for treatment.

Since Laennec's invention of the monaural stethoscope, respiratory sounds have aided physicians with clues of underlying abnormalities (Robertson, 1957). Only recently have adventitious lung sounds been validated as possible indicators of specific disease processes (Nath & Capel, 1974).

The auscultation of one adventitious sound, the crackle, within the timing of the respiratory cycle has been correlated to restrictive or obstructive diseases. The diseases include asbestosis, bronchiectasis, alveolitis and bronchitis (Nath & Capel, 1974a, 1980b; Murphy et al. 1984). However, crackles have also been found in young women without significant lung disease after a forced expiration (Workum, Holford, Del Bono & Murphy, 1982).

Another adventitious sound, the wheeze, is commonly observed in patients with chronic airflow obstruction. However, wheezes may also be detected in patients with normal FEV_1/FVC (forced expiratory volume in one second/forced vital capacity, a measure of airflow obstruction), as a forced expiratory maneuver results in compression of the central airways and produces a wheeze

(Marini, Pierson, Hudson & Lakshminarayan, 1979). The literature supports the clinical relevance of lung sounds as indicators of abnormality and suggests that continued study is needed.

Despite the usefulness of pulmonary auscultation, Carrieri, Stotts, Murdaugh, Levinson, and Holzemer's (1982) survey of nurses trained in cardiopulmonary assessment found that nurses were not practicing the skill on their jobs. The study obtained self-report from nurses within a wide range of nursing occupations who had received instruction in cardiopulmonary assessment (N=626). The study found 78% of the population used the pulmonary exam; however, only 47% used it daily. The pulmonary exam was used most by the clinical nurse specialists, followed by the staff nurses and then the head nurses in acute health care settings. Although 72% believed that assessment skills had significantly increased the quality of nursing care, only 25% believed their assessments altered the medical regimen.

Even though the nurse population studied by Carrieri et al. (1982) had been taught current examining methods, they continued to use the classical pulmonary exam (auscultation for presence and absence of breath sounds, use of accessory muscles and retraction) rather than the current exam (auscultation for normal breath sounds and adventitious lung sounds, anterior-posterior diameter, inspiratory/expiratory ratio, lateral expansion, tracheal alignment, and diaphragmatic excursion). Several possible reasons were given for the lack of use of the current exam. Some nurses indicated that nursing administration did not value assessment as a nursing function, while others noted resistance from co-workers. A few nurses believed that physicians questioned the validity of their nursing assessments. However, the most frequent deterrents mentioned were insufficient time at work to use the skill and insufficient practice of the skill during the instruction.

The auscultation of lung sounds is used more frequently than any other diagnostic devise in respiratory care. Its use assumes that each clinician is able to hear, evaluate and report the sounds with sufficient accuracy and reliability to provide meaningful information for the clinical team. However, many nurses do not use auscultatory skills on a daily basis and some believe that the validity of their auscultatory findings is not readily accepted. The purpose of this study is to determine whether nurses who receive training in a standardized nomenclature and who use uniform reporting methods will report similar findings.

Assumptions of the Study

This clinical study made the following assumptions:

1. The raters' hearing are within normal limits;

2. Raters listening to the same anatomical area at different times hear the same sounds;

3. Raters report their subjective findings honestly;

4. Raters are able to achieve criterion-referenced mastery learning with instruction based upon the criteria;

5. The investigator maximizes the patients' efforts to conform with coached breathing instructions;

6. The effect of controlled breathing on the patient minimizes the response of a forced expiration;

7. Crackle sounds anatomically located on one side of the chest are not heard on the other side;

8. Wheeze sounds might travel through the thoracic cage and can be heard on either side; and

9. Two randomly chosen nurse raters and a convenience patient population reflect a typical nursing population and patient population.

Research Question

Will nurse-raters auscultating the same subjects and using the same instrumentation agree on their findings? The research question was tested by the following null hypothesis:

Hypothesis

There will be no statistically significant agreement among the auscultatory findings rated and reported by three nurse-raters who examine the same patient simultaneously and independently for adventitious breath sounds.

Operational Definitions

Crackle - short bubbling, crackling, popping or fizzing discontinuous sounds that may vary in pitch, intensity, number and location in the respiratory cycle. The grading scales are found in the Methods section.

Wheeze - a continuous, musical sound characterized by monophonic or polyphonic chords; high or low pitched, either inspiratory or expiratory, short or long, loud or faint. The grading scales are found in the Methods section.

Stethoscope - binaural earpieces and single tubing connecting to both bell and diaphragm. The diaphragm was used in this study.

Anatomical assessment sites - lung base and apex regions located in nine thoracic areas as stated in the Methods section.

CHAPTER II

METHOD

Preliminary Study

In 1981, this investigator conducted a pilot study to establish a level of interrater reliability and agreement on auscultated lung sounds using a grade-rated assessment tool (see Appendix A). Different anterior chest wall areas were auscultated using the same standardized procedure. The nomenclature of the American Thoracic Society was used to maintain a uniform description of any sounds obtained. A grade rated assessment scale was used to record and distinguish crackles and to record the absence or presence of wheezes.

Intensive Care Unit nurses received instruction in the use of the assessment tool and the nomenclature. The instruction included:

1. written materials explaining the use and the purpose of the tool and the nomenclature;
2. a lecture on the suggested underlying pathophysiology and on the data collection procedures; and
3. an auditory review of adventitious lung sounds.

After a patient was determined to have met the inclusion criteria, paired nurse-raters using stethoscopes with binaural earpieces and single tubing, simultaneously (or within one or two minutes of each other) gathered the patient data. Assessment tools were completed independently without first comparing results. The data sheets were collected in an envelope and not shared with or evaluated by the raters until the preliminary study was completed. After the data tool was placed in the envelope, the nurse responsible for that patient's care was consulted and information was shared by all raters. To maintain raters' anonymity the nurses were assigned a code letter for use on the form. At the termination of the study, 14 patients had been rated by three different A raters and by the investigator who was the B rater in each case. The data were analyzed for interrater reliability and agreement.

All patients included were at least 18 years of age and each had a properly placed cuffed orotracheal or tracheostomy tube as documented on the patient's chart by x-ray findings. The timing of each patient's examination was determined by the patient's primary nurse. A patient was not included in the study population more than once. The patient's hospital record number was used to protect anonymity. A table and scatterplot of the crackle data (see Appendix B) showed a positive correlation between A and B raters. The Wilcoxon T test (Downie & Heath, 1974) for the total scores indicated there was no significant difference between the A and B raters' grade-rated assessment data. High interrater reliability was suggested. Using the Wilcoxon T test each lung area was assessed for agreement. No significant difference, $T\ (N=13) = 4.0\ p < .01$ was found between the A and B raters except at the 7th right intercostal space, posterior-axillary line.

The same areas were also assessed for wheezes. The range of percent agreement was 57% to 100%.

This preliminary study supports the hypothesis that Intensive Care Nurses using a uniform assessment tool and instructed in the same nomenclature will report similar findings. The methods used in the present study differ from the preliminary study as follows:

1. The subject population was expanded to be a more heterogeneous population by including all patients on units who met the inclusion criteria;
2. The areas for auscultation assessment were expanded to include both anterior and posterior lung fields, including apices and bases;
3. The crackle and wheeze grading scale were made more detailed and were based upon the timing of the respiratory cycle;
4. The patients were coached to open mouth breathe at a normal respiratory rate but slightly deeper to control differences in upper respiratory airway openings and to eliminate forced expiration; and
5. The nurse-raters did not discuss any findings until after the completion of the study to reduce bias.

The above changes are reflected in the following research design.

Research Design

The descriptive design of this study reflects the early stages of research about the criterion variables and the clinical nature of the research question; "Will nurses who have been taught current ATS nomenclature and who have used a uniform assessment tool, report similar findings after auscultating the same patients?" Grade-rated adventitious breath sound scales for crackles and wheezes were further categorized by their timing and used in conjunction with a systematized auscultatory examination. Descriptive results of the examination were reported as quantified subjective data on the auscultation tool. Interrater agreement was demonstrated numerically.

Sample and Setting

The sample of 34 patients was obtained from consenting medical, cardiovascular and respiratory units at a metropolitan teaching hospital. The units were chosen for several reasons. There was a relative frequency of diagnoses that suggested the presence of adventitious breath sounds in the patient population. The units' patient capacity totaled 40 per day and nurses could perform frequent pulmonary auscultation. Finally, staffing patterns on the units allowed for the participation of two registered nurses.

Only those patients meeting the following subject inclusion criteria were examined:

1. eighteen years of age or older, and
2. agreeable and accessible for auscultation

(see Appendix C).

To ensure subject anonymity, the patient's hospital record number was used for identification.

Several environmental conditions on the unit were not controlled including noise, room temperature, and maintenance of privacy. The data were collected during June, July and August of 1984.

Raters

Two nurse-raters were randomly selected from two larger volunteer pools of medical unit registered nurses respectively. Day shift and evening shift personnel comprised the two groups. All were recruited during initial inservice classes on breath sounds. Each gave written consent to be in the study. The unit nurses did not follow a specific protocol when conducting pulmonary exams. However, a formal chart review of nurses' notes revealed an infrequent comment on respiratory evaluation.

Instruction

The two nurse-raters were given additional instruction by the investigator based upon criterion referenced objectives (see Appendix D). Instruction included the following areas: pulmonary anatomy, adventitious breath sounds (tape recordings), assessment tool usage, and application of the rating scale and the data collection procedure. In addition, the video-cassette "Lung Sounds" by Nathoo and Capel (1981) was viewed.

Next, a criterion referenced self-test (see Appendix E) was given to the raters to assess their understanding of and ability to apply the American Thoracic Society's Pulmonary Nomenclature. Content validity of the self-test was analyzed by a Pulmonary Clinical Nurse Specialist. The raters self-corrected the test items. All nurse-raters completed the self-test at a level of 100% accuracy on the first correcting.

Finally, the nurse raters participated in a practice auscultatory session on three patients and compared their respective findings. Upon checking the assessment tool results by inspection, the results were found to be similar, the procedure was judged workable and it seemed feasible to continue. The testing effect bias of the self-test and the practice session were expected to reinforce the nurse-raters' knowledge of the new nomenclature and to clarify any problems.

The three nurse raters determined convenient times to meet for data collection. On specified dates during shift overlap (2 p.m. to 3 p.m.) the three raters examined patients. Nurse anonymity was maintained on the data sheets by using assigned code letters.

Assessment Tool

The assessment tool was a single paper and pencil form (see Appendix F). The data included date, time, subject, primary diagnosis and rater identification codes. The auscultatory numerical rating scale and descriptive criteria for both crackle and wheeze values were provided. The auscultatory areas were referenced and corresponding boxes were included for reporting purposes.

Crackles. With respect to each chest area examined on each patient, the raters' assigned the following values to crackles:

	Value		Rater Assessment
Grade	0	=	no crackles;
Grade	1	=	crackles occurring only during early inspiration;
Grade	2	=	crackles occurring only during mid-inspiration;
Grade	3	=	crackles occurring only during late inspiration;
Grade	4	=	crackles occurring during more than one stage of inspiration but not during expiration;
Grade	5	=	crackles occurring only during expiration;
Grade	6	=	crackles occurring in both inspiration and during expiration.

Wheezes. With respect to each chest area examined on each patient, the raters' were assigned the following values to wheezes:

	Value		Rater Assessment
Grade	0	=	no wheezes;
Grade	1	=	wheezes occurring only during inspiration;
Grade	2	=	wheezes occurring only during expiration;
Grade	3	=	wheezes occurring during both inspiration and expiration.

Assessment areas. The following nine regional lung areas were auscultated to assess for adventitious lung sounds in lung bases and apices.

1. Above posterior right scapula at mid-scapular line;
2. Above posterior left scapula at mid-scapular line;
3. At the posterior tenth level of the thoracic process at the right mid-scapular line;
4. At the posterior tenth level of the thoracic process at the left mid-scapular line;
5. At the anterior right second intercostal space mid-clavicular line;
6. At the anterior left third intercostal space, mid-clavicular line;
7. At the anterior right fourth intercostal space, mid-clavicular line;
8. At the right seventh intercostal space, anterior axillary line;

9. At the left seventh intercostal space, anterior axillary line (Bates, 1983).

Data Collection Procedures

Prior to auscultation, the diagnosis or condition of the patients' lung areas were unknown to the nurse raters. After subject inclusion criteria were met, the nurse-raters used the same kind of stethoscope (binaural earpieces with single tubing) and the following auscultation procedure:

1. In a relaxed atmosphere patient privacy was established (the patient was undressed to the waist and examined with good lighting in a warm room);
2. The patient was sitting or if unable to sit, the position noted;
3. The patient was coached by the investigator to cough and to breathe deeply through his/her open mouth at a normal rate of respiration;
4. The raters simultaneously, yet independently, auscultated the lung fields with the diaphragm of the stethoscope and rated them (One rater worked on the anterior chest surface while the other two raters worked on the posterior chest. Raters then rotated until all areas were auscultated.);
5. After auscultation, the investigator informed each subject's nurse if adventitious

sounds were present. If the nurse was unavailable for verbal report, the sounds were documented in the nurses' notes of the patients' chart.

Analysis of the Data

To compare findings of the three nurse-raters for interrater agreement, tables were constructed using the scores from each lung area as graded by each of the three raters. The tables illustrate crackle data and wheeze data. Percent agreement was computed on both crackle data and wheeze data to determine rater agreement. Missing data were given the value 9.

To determine the relationship of the amount of rater agreement from that of chance occurrence; Cohen's Kappa as extended by Light (1971) was computed. The Cohen equation follows:

$$(1) \quad K = \frac{P_o - P_c}{1 - P_c} .$$

CHAPTER III

RESULTS

Three nurse raters each auscultated the same 34 patients for either the presence or the absence of adventitious breath sounds. Crackle and wheeze sounds in each patient were evaluated separately. The nurses independently rated and recorded their findings from each of the nine specified anatomical chest areas that were assessed. The purpose of this chapter is to describe and analyze the data collected. The text incorporates tables and figures to present general findings of the distribution of crackles and wheezes and descriptive results of the sample. The data most relevant to the research question regarding agreement among the raters' findings are compared and statistically analyzed. Finally, nurse interrater response patterns are outlined.

General Findings-Crackles and Wheezes

Each rater was required to report the phase in the respiratory cycle in which the sound was heard. For crackles the phases were divided into seven grades as follows: no crackles (Grade 0); early inspiration (Grade 1); mid-inspiration (Grade 2); late inspiration (Grade 3); more than one phase of inspiration (Grade 4); only expiration (Grade 5); and inspiration and expiration (Grade 6). Table 1 provides the frequency distributions of the crackle sounds found in the sample. Crackle sounds were reported 63 times in 16 different subjects. The following

Table 1

Crackle Grade-score Frequency Distribution by Lung Area

Lung Area	Grade[a]					
	1	2	3	4	5	6
CR_1[b]				4		2
CL_1[b]		2	1	2		
CR_2[b]	3	2	1	3	2	3
CL_2[b]	3	2		9		1
CR_3[c]	1			2		2
CR_4[c]				3		
CL_5[c]				3		
CR_6[d]	1	1	1	1	1	1
CL_6[d]				2		

[a] Grade scores follow: 1, early inspiration; 2, mid-inspiration; 3, late inspiration; 4, more than one phase of inspiration; 5, only expiration; and 6, both inspiration and expiration.

[b] Posterior lung areas: CR_1 = above right scapula at mid-scapular line; CL_1 = above left scapula at mid-scapular line; CR_2 = at the tenth level of the thoracic process at the right mid-scapular line; and, CL_2 = at the tenth level of the thoracic process at the left mid-scapular line. [c] Anterior lung areas: CR_3 = at the right second intercostal space, mid-clavicular line; CR_4 = at the right fourth intercostal space, mid-clavicular line; and, CL_5 = at the left third intercostal space, mid-clavicular line. [d] Lateral lung areas: CR_6 = at the right seventh intercostal space, anterior axillary line; and CL_6 = at the left seventh anterior axillary line.

Table 2 summarizes the reported crackles' grade frequency in percentages.

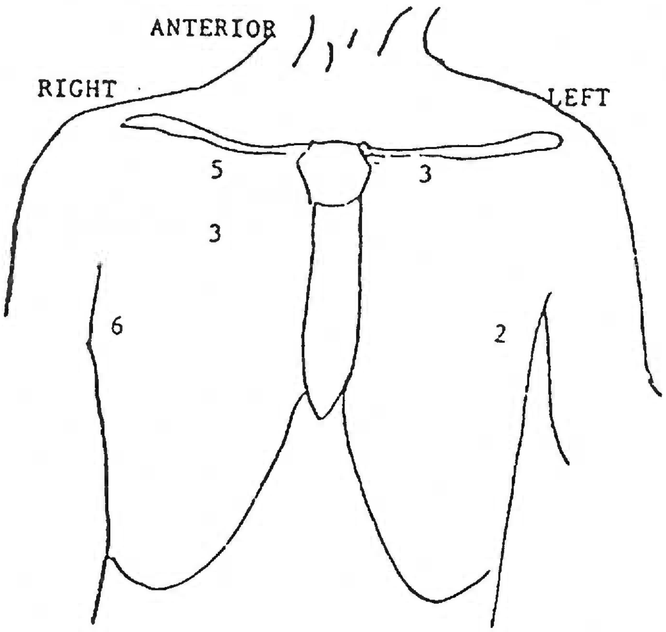
Table 2

Frequency of Crackle Grade Scores.

Grade	Frequency	Percentage
1	8	13%
2	7	11%
3	3	5%
4	29	46%
5	3	5%
6	13	20%

As shown in Figure 1, crackles reported in the sample were heard most frequently during auscultation at the mid-scapular line of the right and left posterior thorax at level T 10 (CR_2 and CL_2 in Table 1). Forty-nine percent of the reported crackles were heard in these bibasilar areas. The two next most frequent locations of reported crackles, each contributed 10% of the reported crackles, were the right anterior mid-axillary location (CR_6 in Table 1), and the right posterior, above the mid-scapular line (CR_1 in Table 1).

For wheezes the respiratory phases were divided into these different grades: no wheeze (Grade 0); only inspiration (Grade 1); only expiration (Grade 2); or both



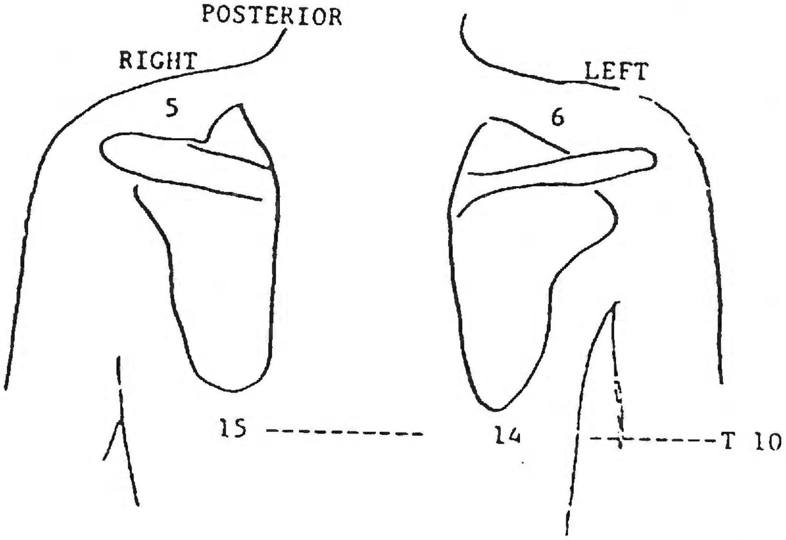


Figure 1. Relative Frequency of Crackle Distribution. Total frequency of crackles is 59.

inspiration and expiration (Grade 3). Table 3 provides the frequency distributions of the wheeze sounds found in the sample. Wheezes were reported 72 times in nine different patients. The grade most often reported was Grade 2 which accounted for 46% of the wheezes heard. Grade 3 accounted for 38% of the wheezes and Grade 1 accounted for the remaining 16%.

As shown in Figure 2, wheezes were reported most frequently (28%) at the combined right and left posterior thorax bibasilar level T10, midscapular line (WR_2 and WL_2 in Table 3). Wheeze distribution in all other areas ranged between 14% and 11% of the total determinations.

Table 3

Wheeze Grade-score Frequency Distribution by Lung Area

Lung Area	Grade[a] 1	2	3
WR_1[b]	1	2	6
WL_1[b]	2	3	3
WR_2[b]	1	5	4
WL_2[b]	2	4	4
WR_3[c]	2	3	4
WR_4[c]	1	4	2
WL_5[c]	1	4	1
WR_6[d]	1	4	2
WL_6[d]	0	4	2

[a] Grade scales follow: 1, inspiration; 2, expiration; and, 3, both inspiration and expiration. [b] Posterior lung areas follow: WR_1 = above right scapula at mid-scapular line; WL_1 = above left scapula at mid-scapular line; WR_2 = at the tenth level of the thoracic process at the right mid-scapular line; and, WL_2 = at the tenth level of the thoracic process at the left mid-scapular line. [c] Anterior lung areas follow: WR_3 = at the right second intercostal space, mid-clavicular line; WR_4 = at the right fourth intercostal space, mid-clavicular line; and, WL_5 = at the left third intercostal space, mid-clavicular line. [d] Lateral lung areas follow: WR_6 = at the right seventh intercostal space, anterior axillary line; and, WL_6 = at the left seventh intercostal space, anterior axillary line.

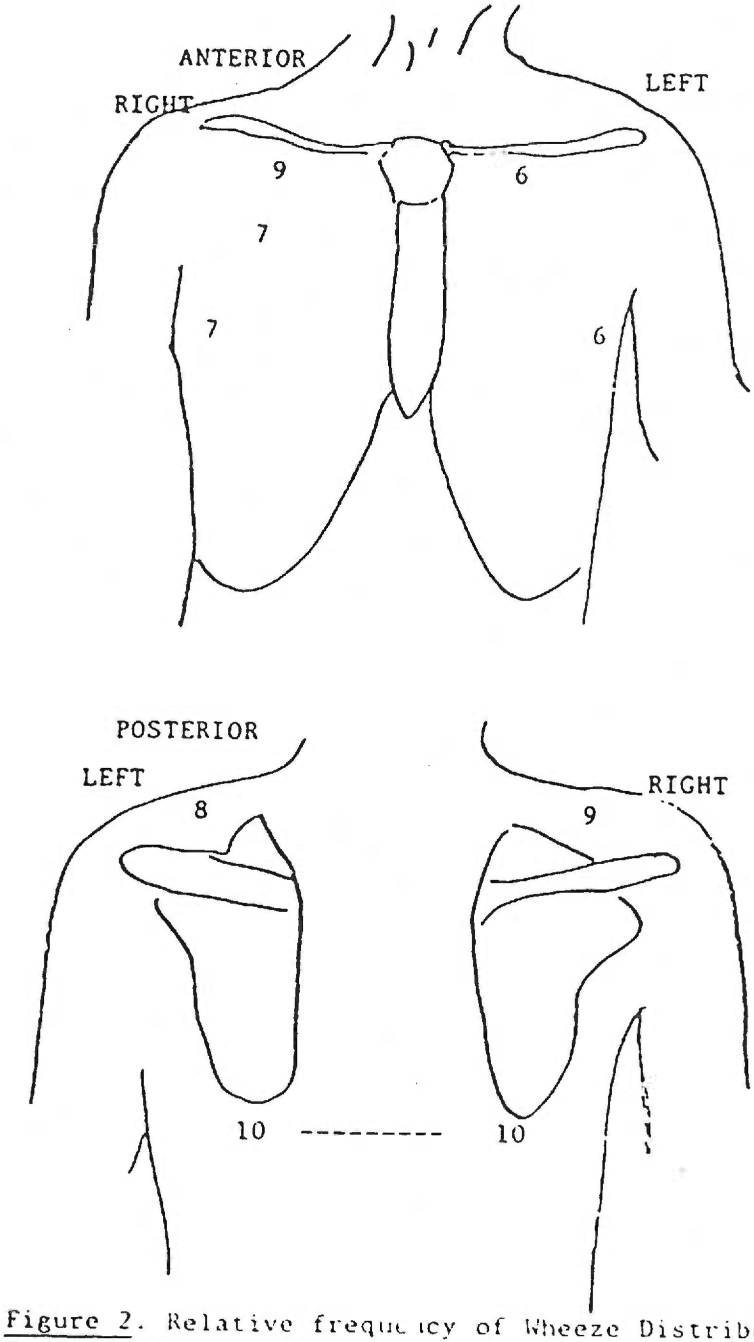


Figure 2. Relative frequency of Wheeze Distribution. Total frequency of wheezes is 72.

Table 4 shows the 34 patients separated into categories by their medical diagnoses and the distribution of reported adventitious sounds among these patients.

Table 4

Adventitious Sounds by Diagnostic Groups

Diagnostic Group	n	Crackle	Wheeze	No Adventitious
Cardiovascular	11	3	1	7
Respiratory	11	10	4	1
Other Primary Diagnosis	12	3	4	6

Note: Some subjects have both crackles and wheezes.

Descriptive Results of Sample

For the 34 subjects at least one of the three raters reported adventitious sounds (either crackles or wheezes) in 20 subjects and no adventitious sounds were reported in the remaining 14 subjects. Because of the relative lack of adventitious sounds reported in the lung areas (see Table 1) the results were aggregated so that any crackle grade rating of 1 through 6 or wheeze grade rating of 1 through 3 were tabulated from the rater responses on the tool as an X. The investigator made the notation O for Grade 0 rater responses. Results reported for patients were divided into either wheeze or crackle categories. Crackles were further differentiated based on whether they were heard on the right side or the left side of the patient's thoracic cavity. The division of the crackles into separate hemispheres is based upon the assumption that crackles occurring on one side would not be heard on the other side of the body. This division between left and right sides was not made for reported wheezes because of the assumption that wheeze sounds may be transmitted throughout the entire chest (Forgacs, 1978). Therefore, for each of the 34 patients agreement or lack of agreement was determined three times for each patient, once for wheezes and twice for crackles. A total of 102 determinations was made.

Missing items of data were reported on one subject, by one rater. The rater could hear the presence or absence of adventitious sounds but could not grade rate them. The data from this subject was retained in the sample. Table 5 summarizes the frequency of rater response patterns for crackles and wheezes.

As shown in Table 5 when considering all three raters' findings, there are eight possible pattern configurations. Two of these patterns indicate complete rater agreement (O,O,O) and (X,X,X). The remaining six pattern configurations are variations of the "O,O,X" and "X,X,O" pattern showing partial agreement and partial disagreement within the raters' findings.

Table 5

Distribution of Auscultation Findings

Possible Pattern Responses			Adventitious Sounds			Total Pattern Frequency
			Crackle		Wheeze	
Rater A	Rater B	Rater C	right	left		
O[a]	O	O	21	23	25	69
X[b]	X	X	3	1	4	8
O	O	X	1	2	0	3
O	X	O	2	2	0	4
X	O	O	5	3	2	10
X	X	O	2	3	1	6
X	O	X	0	0	1	1
O	X	X	0	0	1	1

[a] O indicates raters reported absence of adventitious sound.

[b] X indicates raters reported presence of adventitious sound.

Interrater Agreement

The investigator was interested in estimating the reliability of the instrument. As different observers are using the instrument to measure the same phenomena at the same time, an equivalence reliability may be computed by percent agreement (Polit & Hungler 1983, chap. 18) The investigator used the three pairs of raters (AB), (AC) and (BC) to show percent agreement for each pair. The chance expectation of probability for a pair of raters agreeing was 50% as shown by the following patterns:

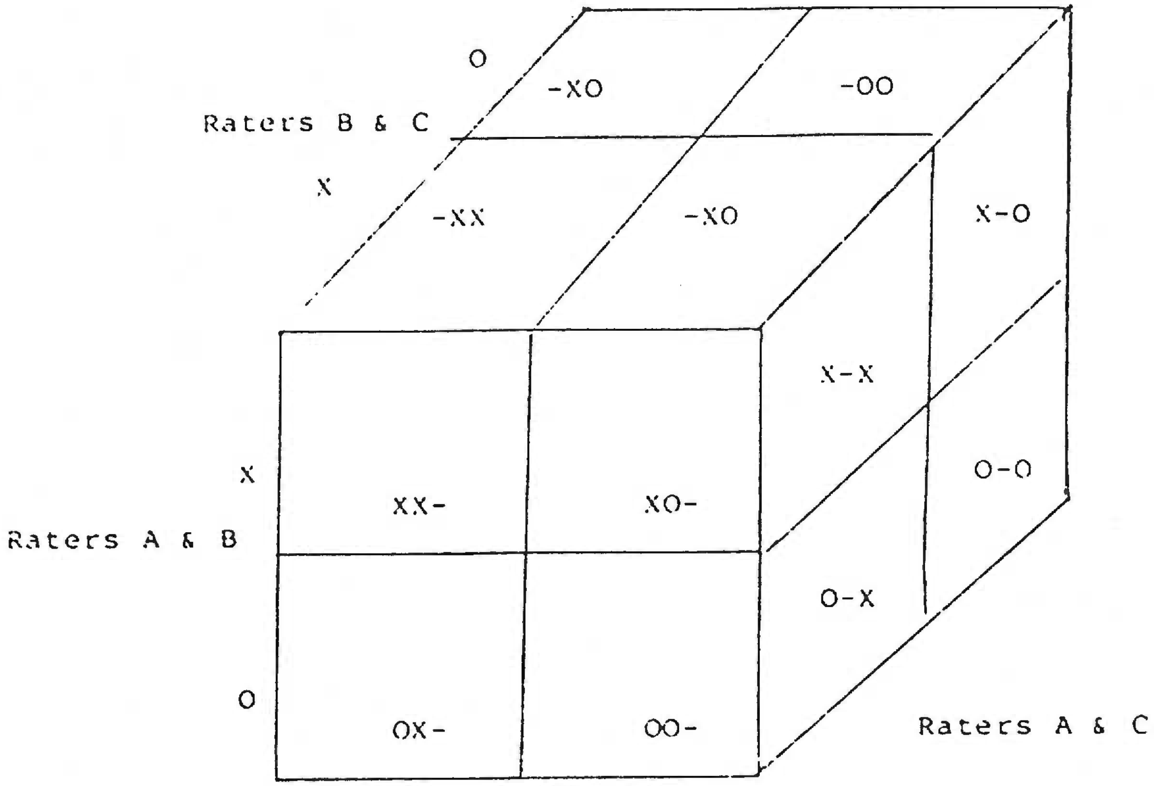
XX
OX
XO
OO

Agreement patterns were XX and OO. Where there existed complete agreement regarding the presence or absence of adventitious sounds, the percent agreement was found to be:

(A&B) = 88%
(A&C) = 80%
(B&C) = 86%

By continuing to analyze the raters by three separate pairs their findings could be categorized into four pattern configurations (XX, OX, XO, OO) with respect to each pair as shown in Figure 3.

The frequency in which each configuration occurred with respect to each pair are shown in Figure 4.



X = presence of adventitious sound

O = absence of adventitious sound

<u>Figure 3</u>. Notation of Agreement Among Three Raters Reporting the Absence and/or the Presence of Adventitious Sound.

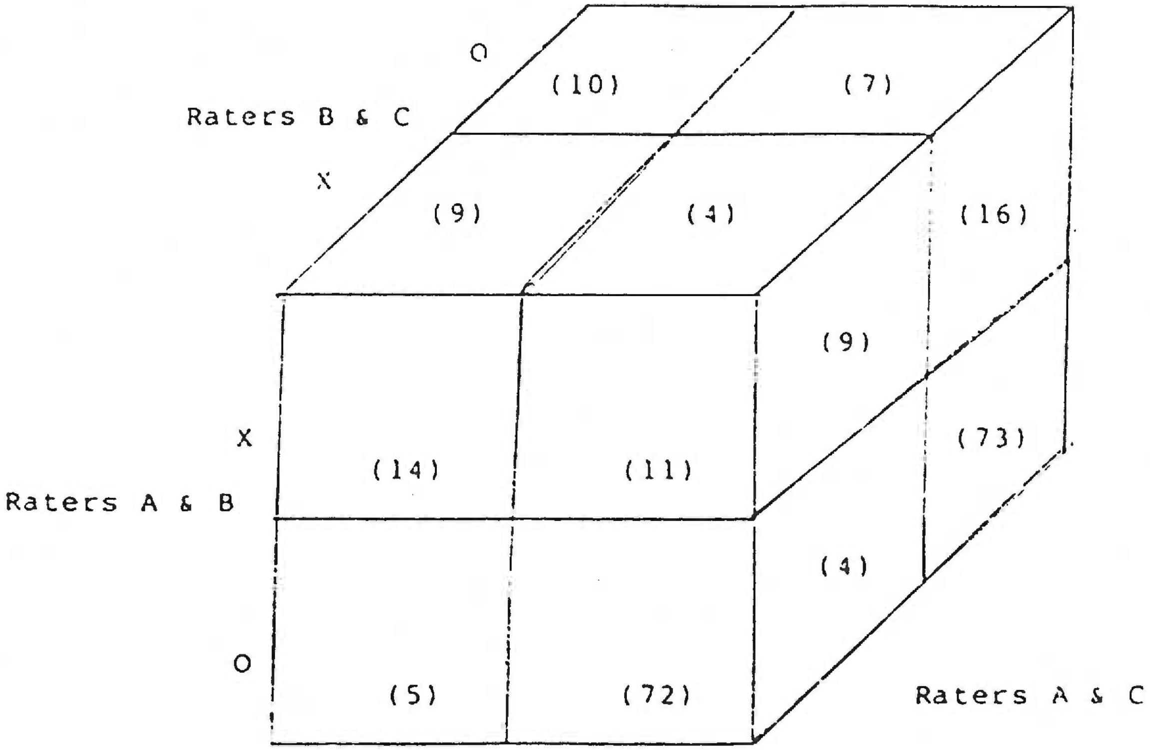


Figure 4. Set of Margin Totals for 102 Determinations.

Comparisons

The results reported in this chapter are analyzed using the coefficient of interrater agreement for nominal scales, known as Cohen's Kappa as extended by Light (1971). The specific comparisons made are as follows:

1. A comparison was made, in which pairs of rater responses were tested for agreement on the presence (X,X) of crackles or the presence (X,X) of wheezes. Of the 102 reported determinations, 69 were excluded as none of the raters reported any adventitious sounds. The remaining 33 determinations were tested for agreement as to the presence of adventitious sounds.
2. A comparison was made in which pairs of rater responses were tested for agreement on the presence (X,X) or the absence (O,O) of adventitious sound. All 102 reported determinations (3 for each subject) were used in testing for agreement.

Cohen's Kappa as Extended by Light

The most applicable statistic available for measuring interrater agreement for nominal data is Cohen's coefficient of agreement referred to as Kappa (K) (Cohen, 1960). The following conditions must exist for the use of Kappa as the coefficient of agreement:

1. The units must be independent.
2. The categories of the nominal scale must

be independent, mutually exclusive and exhaustive.

3. The judges must operate independently and be deemed equally competent to make judgments.

Cohen's K provides the proportion of agreement between the raters after agreement which can be attributed to chance has been removed from the calculation. This adjustment for the amount of agreement which is to be expected by chance makes the statistic Kappa superior to a mere percent agreement. The coefficient of K is simply the proportion of agreement after chance agreement is removed from consideration:

$$(1) \quad K = \frac{P_o - P_c}{1 - P_c} .$$

Where P_o is equal to the proportion of observed units in which the judges agreed and P_c is equal to the proportion of units for which agreement is expected by chance. The test of agreement comes with regard to the $1 - P_c$ of which the remainder is the proportion of agreement. $P_o - P_c$ represents the proportion of the cases that occur in excess of chance agreement.

If there is perfect agreement between the judges, then $K = 1.00$ which becomes the upper limit of K. When the obtained agreement equals chance agreement then $K = 0$. Greater than chance agreement leads to a positive value of K, whereas, less than chance agreement leads to negative values. K cannot be less than -1.00.

Once K has been determined it is then necessary to test the obtained K for significance. In this manner the null hypothesis can be either rejected or not rejected. This is done by determining the difference between the obtained K resulting from the data and K = 0 indicating agreement is due to chance.

Cohen (1960) and Fleiss, Everitt and Cohen (1969) have provided formulas for determining the standard error of K assuming the raters observations had been based on chance and assuming K = 0. By substituting P_c for P_o in the above formula, an estimate of the variance of K is provided, as follows:

$$(2)\ K_{var} = \frac{P_c - P_c}{1 - P_c}$$

The square root of the variance of K equals the estimated standard error (S_o) of K. By dividing K by the standard error ($K/S_o(k_m)$) the resulting ratio can be referred to the normal curve to determine the likelihood of any agreement arising by chance. If K was shown to be at least 1.96 standard deviations away from K = 0 then the agreement would be significant at the .05 level of significance.

Cohen's Kappa was extended by Light (1971) to situations where more than two observers rated the same event. This extended K requires the three raters to be divided into three pairs of raters and the statistic determines the observed proportion of disagreements for each pair and adjusts for the expected proportion of disagreements. The result is a multiple observer agreement statistic which Light refers to as K_m. Just as in Kappa, above, an approximate large sample estimate of the variance of K_m can be found which yields an estimated standard error and a test for level of significance.

In the statistical analysis outlined below, the statistic K_m was used in each case as the data are based upon the observations of three raters. By using K_m and the analysis of the observation of pairs of raters the statistic is sensitive to situations where two out of three raters agreed. For example, if raters A & B found the presence of an adventitious sound which was not reported by rater C there would be agreement by pair A & B but no agreement by pairs A & C and B & C.

Statistical Analysis of Comparisons.

The following results of the two comparisons of pattern distributions use the Cohen's Kappa as extended by Light (1971).

1. Comparison was made of the agreement of the reported presence (X,X) of adventitious sounds by three pairs of raters (AB, AC & BC):

$$K_m = .272.$$

H_o: Using the large sample normal approximation, $K_m / S_o(k_m) = 1.265$, $K_m = .272$ does not differ from zero (chance expectation) at the .05 level of significance.

Decision rule: If $K_m/S_o(k_m) \geq 1.96$ then reject the null hypothesis. If $K_m/S_o(km) \leq 1.96$ then fail to reject the null hypothesis.

Therefore, since 1.265 does not differ from chance expectation with statistical significance, the null hypothesis cannot be rejected.

2. Comparison was made of the reported agreement of presence and absence (XX, OO) of adventitious sounds by three pairs of raters (AB, AC & BC):

$$Km = .463.$$

H_o: Using the large sample normal approximation, $K_m/S_o(k_m) = 10.728$, Km = .463 does not differ from zero (chance expectation) at the .05 level of significance.

Decision rule: If $K_m/S_o(k_m) \geq 1.96$ then reject the null hypothesis. If $K_m/S_o(k_m) \leq 1.96$ then fail to reject the null hypothesis.

Therefore, since 10.728 does differ from chance expectation well beyond the .001 level of significance, the hypothesis must be rejected.

Interrater Response Patterns

Complete agreement patterns (O,O,O) and (X,X,X) were formed a total of 77 times (see Table 5). The remaining 25 response determinations were variation of the "X,X,O" and "O,O,X" patterns.

Pattern "O,O,X". The pattern of subjects with adventitious sounds where only one rater heard a sound (O,O,X) occurred 17 times. Of these patterns responses, Rater A heard an adventitious sound eight times when the two other raters agreed that there were no adventitious sounds, Rater B heard an adventitious sound four times when the two other raters agreed there were no adventitious sounds and Rater C heard adventitious sounds three times when the other two raters reported no adventitious sounds. Conversely, validation of the second rater's agreeing that there was no added sounds occurred as follows:

Rater A agreed seven times;
Rater B agreed 13 times; and
Rater C agreed 14 times.

Pattern "X,X,O". Of the subjects with adventitious sounds (n=20) when only two of the three raters heard sounds (X,X,O), agreement of either crackles or wheeze sounds occurred eight times. Rater A, omitted sound four times, Rater B omitted sound four times and Rater C omitted sound three times. Conversely, Rater A was validated by another rater seven times, Rater B was validated by another rater seven times and Rater C was validated by another rater twice.

CHAPTER IV

DISCUSSION

The purpose of this clinical study was to compare the agreement in auscultatory findings of three nurse raters. The three nurse raters received instruction in current adventitious lung sound nomenclature and received instruction in a uniform reporting method. After achieving 100% on a mastery level test the nurse raters examined and rated the same patients independently but within the same time period. Data were examined for statistical significance of agreement using Cohen's Kappa as extended by Light (1971). The discussion which follows presents the study's findings, suggests limitations occurring in the methods used and relates the results to previous studies.

Statistical Comparisons

A summary of the two statistical comparisons of the data gathered in this study as compared to chance expectation follows:

1. The three pairs of raters (AB, AC & BC) reported complete agreement of the presence (X,X) of adventitious sounds 14, 9 and 9 times respectively, out of the 33 times that sounds were heard. The statistical analysis showed that this occurrence did not differ significantly from chance expectation.

2. The three pairs of raters (AB, AC, BC) reported complete agreement of the presence and

absence (X,X, 0,0) of adventitious sounds 86, 88 and 82 times respectively, out of the 102 determinations. The statistical analysis shows that this occurrence did differ from chance expectancy at the $\underline{p}<.001$ level of significance.

The study found that nurse raters instructed in adventitious breath sounds and the use of a reporting tool agree on presence or absence of adventitious sounds with statistical significance when all 102 determinations were considered. However, it was found that the raters did not agree with statistical significance in the 33 determinations when adventitious sounds were heard by at least one rater. The strength of these findings are dependent upon the design of this clinical study. The raters were randomly chosen. The subject population had diagnoses suggesting a relative frequency of adventitious sound. It is also important to remember that the raters did not know ahead of time whether the patient had adventitious sounds or not, and that all findings were reported independently. Because of the above design it is possible to make inferences about the results of the study.

In the first comparison, rater agreement did not differ with statistical significance from chance expectancy, therefore, the null hypothesis was was not rejected. The failure to reject the null hypothesis may mean that the null hypothesis is actually the case. Nurses, even if they are instructed will not be able to agree on adventitious sounds.

More likely however, the null hypothesis is false. Nurses can learn to agree at a statistically significant level and factors other than ability prevented the null hypothesis from being rejected.

Goodwin (1984) describes a number of such factors that can contribute to this Type II error and suggests the following strategies for increasing the power (the probability of rejecting a false null hypothesis) of the statistics used in the research design to minimize their significance:

> 1. by using a larger population, by relaxing the alpha level and/or by testing a directional hypothesis as compared to the null hypothesis; and
> 2. by increasing the sensitivity of the assessment tool which may also include increasing the rater's instructions.

To increase the probability of rejecting a false null hypothesis in future studies the following method changes should be considered:

> The sample size should be increased and the alpha level relaxed to a lower level. An increase in the sample size results statistically in a decreased error mean square. Relaxing of the alpha level changes the decision rule about rejecting the null hypothesis. It is important to remember that by relaxing the alpha there is a greater probability of making a Type I error. In

this study the consequence of a Type I error is a greater possibility that the rejection of the null hypothesis is due to chance.

Another method approach worth considering to increase power is to test the null hypothesis against a directional rather than a nondirectional alternate hypothesis that way, the critical value for rejecting the null hypothesis is lowered.

The assessment tool should be made less cumbersome, thereby increasing the sensitivity of the instrument as it is used. The length and lack of clarity in the instrument may have been a source of frustration for the raters. The crackle grade scale is lengthy and overlaps in the inspiratory cycles. This may have discouraged the raters from distinguishing between the phases of the cycle. For example, if Rater A reported a Grade 1 crackle (early inspiratory) and rater B reported a Grade 4 crackle (more than one phase of inspiration) it can not be determined how much agreement there was or where the agreement occurred. The tool should be constructed to force the rater into committing to the most specific cycle phase heard. This could be done by reducing the categories to the following: early inspiration; late inspiration; expiration. More than one response could be made if appropriate. The smaller number of categories may also increase the instrument clarity (Polit & Hungler, 1983). Another possible tool simplification would be to present the number of areas auscultated on the patient in a more concrete figure of the chest rather than the abstract boxes. This would give the rater better reference points for remembering where the lung sound was heard. The tool could be placed on computer data cards for easy computer transfer.

The data collection should be modified to increase the reliability of the instrument. A stethoscope with three aurals would allow simultaneous auscultation of the raters. In this study the adventitious sounds could have changed at the area of auscultation for a number of reasons. The auscultation took approximately three to four minutes per subject. As rater A listened to the same spot rater B previously listened to, the time elapsed may have been as little as 10 seconds to as much as three to four minutes. During this time, transpulmonary pressures may have changed due to increased depth of respiration or forced expiration. The changing pressure could have moved fluid in airways or opened deflated airways or closed airways without the raters knowing (one patient did cough during auscultation). Also, slight changes in patients' positions may have affected the lung sounds. The method of auscultating each patient may have caused variations in the sounds heard. Even though raters followed a set protocol, there were times when modifications were made. For example, all patients were assisted to a sitting position during auscultation. However, two very ill patients were placed in semi-fowlers position and assisted to a forward leaning position for access to the posterior lung areas. Finally, faulty stethoscope placement may have prevented the rater from hearing the same sounds.

The clinical setting has many situational contaminants that may have influenced the raters ability to hear, i.e. air conditioners, fans, hallway traffic and conversations. Such contaminants should be minimized, but the importance of using a clinical setting for this study should not be undermined. However, in the future researchers may use high powered stethoscopes to validate the raters' ability to report sounds (Worklum et al., 1982).

Rater instruction should be enhanced to increase the internal validity of the instrument. It is impossible to know whether the raters remembered the sounds as they were instructed throughout the approximate 6 weeks of data collection. Also the mastery learning skill provided to the raters may not have been sufficient to assist the raters in differentiating between the different phases of the inspiratory cycle.

Another threat to internal validity is fatigue. As described by Schneider and Anderson (1965) fatigue and inexperience may have affected the performance of the raters. Other transitory personal factors such as anxiety, hunger or hearing ability may have also influenced the results. On a typical data collection day one rater had already worked several hours while the other two raters had just arrived on the unit.

Finally, individual raters may have response biases. For example, one rater may have consistently ignored faint sounds and another rater may have consistently given more significance to sound than was warranted. Although the sample size is too small to measure such bias, it is noted that Rater C tended not to report sounds when raters A and B reported them and Rater A tended to report sounds when the other two raters did not.

In the second statistical comparison the three pairs of raters (AB, AC & BC) were able to agree on either the absence (O,O) or the presence (X,X) of any adventitious sounds at the .001 level of significance. Since the likelihood of a Type I error (inferring an effect where none exists) in this instance is less than 1 in 1000, it is possible to draw an inference from the results. The raters were able to recognize the absence of adventitious sounds during auscultation of the lungs. However, it is important to note that the statistical measurement used in this instance did not measure whether it was the instruction or any other factor which caused the nurses to agree. A research study should be conducted comparing raters who were given instruction to those who were not.

The use of the Cohen's Kappa, as extended by Light (1971), only shows that it was statistically probable that some factor other than mere chance was responsible for the agreement among raters. It is important to remember that the frequency of the determinations showing the absence of adventitious sounds considerably outnumbered those showing the presence of sounds. This fact may have skewed the results, as the raters may have expected no sounds during auscultation, thereby, preventing any inference that the raters are able to agree on both the presence and absence of adventitious sounds.

The following factors may have influenced the results in favor of showing agreement:

> While the raters did not verbally share information or share their results, unintentional non-verbal communication while auscultating a patient could have influenced other raters.
>
> The Hawthorne effect (Polit & Hungler, 1983) caused by the raters' interest in showing significant results may have been a contributing factor, though after examining the results this possibility seems doubtful.

The results of the study are also dependent upon the assumptions of the study and other factors such as the selection of the raters and selection of the population. The investigator has no reason to believe that any of the assumptions described in Chapter I are not valid. The randomization process used to select the raters and the population should have minimized the errors and increased the ability to generalize the findings.

Previous Studies

Hudson et al. (1978) compared the timing of the respiratory cycle and traditional descriptive terminology for rales using rater agreement. Agreement between observers on qualitative adjectives occurred in 47% of the cases while agreement on timing of the rales occurred in 85% (N=100). The findings did not adjust for the expectation of the chance agreement. The findings of high rater agreement using the respiratory cycle as a basis for comparison were not supported by the current study. A much larger sample of adventitious sounds is needed.

Another area of study focusing on reliability or observer agreement rather than accuracy of findings has shown high disagreement between raters (Schilling et al., 1955). However, when a study has employed criteria and certain methods of measuring observations, observer agreement has increased (Schneider & Anderson, 1965). The present study supports the finding of Schneider and Anderson (1965). Since the specific criteria for observing the criterion variables were measured, agreement was demonstrated.

Validation

The extent to which the findings of this study are consistent with the findings of other researchers tends to validate both the methods used by the raters and their results. For example, Forgacs (1967) found that wheezes are more commonly reported during expiration and are more likely to travel throughout the chest than crackles. The findings of this study support Forgacs' theory (see Figure 1).

Forgacs also found more crackles in the gravity dependent bibasilar lung areas. The findings of this study show 62% of crackles (n=59) were in basilar areas (see Figure 2).

CHAPTER V

SUMMARY, NURSING IMPLICATIONS, AND RECOMMENDATIONS

The purpose of the study was to determine whether the nurse raters would agree in their reporting of adventitious lung sounds. If the finding of the study showed statistically significant agreement then it would suggest the usefulness of the nomenclature, the instruction and the reporting instrument. By demonstrating the validity of the method, nurses might make more frequent use of current auscultatory techniques in providing respiratory care for their patients.

Three randomly selected nurse raters each auscultated nine different anatomical areas of the chest of 34 hospitalized patients. The lung sounds were reported by each rater on a uniform reporting tool designed for use with this study. Each rater was asked to report the presence or absence of both crackles and wheezes and to identify the phase of the respiratory cycle in which the adventitious sounds were heard.

The raters' reported results were compared to determine the amount of agreement. It was found that the number of determinations in which each of the pairs of raters agreed on the presence of a crackle or the presence of a wheeze was not greater than what would have been expected in a chance occurrence. However, the number of determinations in which all three nurse raters completely agreed with each other on either the presence or the absence of crackles or wheezes differed with statistical significance ($p<.001$) from chance expectation.

The results of the study suggest that nurses who have been given instruction on adventitious breath sounds can agree on the absence of adventitious breath sounds in hospitalized patients. While the results do not permit the same conclusion to be drawn for agreement on presence of sound whether crackles or wheezes. Replication of this study with a number of modifications is suggested. The modifications include:

1. enhanced instruction for the raters;
2. simplification of the reporting tool;
3. simultaneous auscultating by the nurse raters; and
4. simultaneous validating devices.

Nursing Implications.

The findings of this study provide a basis for nurses to reliably report the absence of breath sounds. Using mastery learning instruction in adventitious sound nomenclature and a uniform reporting tool resulted in reliable reporting of the absence of adventitious lung sounds in hospital patients. While the absence of abnormal sounds may not mean the absence of underlying disease processes in patients' lungs, the lack of these signs in a patient does provide relevant data to be used in the decision making process. Conversely, upon hearing a suspected adventitious sound, the nurse should inquire further to the exact nature and type of sound.

It is possible that nurses may be required to report their auscultatory findings using cycle sequence, location and other characteristics using the new nomenclature. In the future crackles may be studied as predictor variables. Currently, it is thought that crackles may be predictive of the need to remove secretions in the airway or to indicate an area of lung deflation. The accurate reporting of crackles and wheezes may become more important by suggesting nursing interventions. For example, specific auscultatory findings may indicate that the patient needs repositioning, deep breathing, coughing, and/or suctioning. In addition, the ability to record valid and reliable indices of lung sounds may enable nurses to map patient progress on an hourly, 8 hourly and daily basis resulting in better continuity of patient care.

Recommendation

This descriptive study is a first step in determining if nurse raters using the new nomenclature can reliably agree on auscultatory findings. This step suggests that nurses are able to identify those patients who do not have adventitious sounds. This study should be replicated to validate these findings. Further studies should determine the amount and kind of instruction that are most effective in teaching nurses to assess adventitious sounds in the same way. It is suggested that this be done by using a larger sample, by changing the instrument as recommended above and/or by enhancing the instruction by using a measurable validating device; for example, waveform analysis or a stethoscope that has three aurals with a common diaphragm.

The next step is to isolate variables to determine those that most affect the ability to discriminate the criterion variables. Once a degree of reliability of hearing adventitious sounds is established, crackles should be studied as a predictor variable for possible nursing interventions.

References

Alexy, B. J. (1978). Monitoring cardiovascular status with non-invasive techniques. *Nursing Clinics of North America*, *13*, 423-435.

Amborn, S. A. (1976). Clinical signs associated with the amount of tracheobronchial secretions. *Nursing Research*, *25* (2), 121-126.

American College of Chest Physicians-Joint Ad Hoc Subcommittee of Pulmonary Nomenclature. (1975). Pulmonary terms and symbols. *Chest*, *67* (5), 583-593.

Banaszak, E. F., Kory, R. C. & Snider, G. L. (1973). Phonopneumography. *American Review of Respiratory Disease*, *107*, 449-455.

Bates, B. (1983). *A guide to physical examination* (3rd ed.). Philadelphia: J. B. Lippincott.

Bunin, N. J. & Loudon, R. G. (1979). Lung sound terminology in case reports. *Chest*, *76* (6), 690-692.

Bohadana, A. B., Preslin, R. & Uffholtz, H. (1978). Breath sounds in the clinical assessment of airflow obstruction. *Thorax*, *33*, 345-351.

Bunin, N. J. & Loudon, R. G. (1979). Lung sound terminology in case reports. *Chest*, *76* (6), 690-692.

Carrieri, V., Stotts, N., Murdaugh, C., Levinson, J. & Holzemer, W. (1982). The use of cardiopulmonary assessment skills in the clinical setting. *Western Journal of Nursing Research*, *4* (1), 5-19.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, XX (1), 37-47.

Cohen, J. & Fleiss, J. L. (1969). Large sample standard errors of Kappa and weighted Kappa. Psychological Bulletin, 72 (5), 323-327.

Cugell, D. W. (1971). Use of tape recordings of respiratory sound and breathing pattern for instruction in pulmonary auscultation. American Review of Respiratory Disease, 104, 948-950.

Cugell, D. W. (1978). Sounds of the lungs. Chest, 73 (3), 311-312.

Downie, N. M. & Heathe (1974). Basic statistical methods, (4th ed.). New York: Harper & Row.

Fletcher, C. M. (1952). The clinical diagnosis of pulmonary emphysema - an experimental study. Proceedings of the Royal Society of Medicine, 45, 577-584.

Forgacs, P. (1969). Lung sounds. British Journal of Diseases of the Chest, 63 (1) 1-12.

Forgacs, P. (1978). The functional basis of pulmonary sounds. Chest, 73 (3), 399-405.

Forgacs, P. & Lond, M. D. (1967). Crackles and wheezes. Lancet, 22, 203-205.

Forgacs, P., Nathoo, A. R. & Richardson, H. D. (1971). Breath sounds. Thorax, 26, 288-295.

Goodwin, L. D., P. (1984). The use of power estimation in nursing research. Nursing Research, 33 (2), 118-120.

Grossbach-Landis, J. & Mclane, A. M. (1979). Tracheal suctioning: a tool for evaluation and learning needs assessment. Nursing Review, 28 (4), 237-242.

Hudson, L. D., Conn, R. D., Matsubara, R. S. & Pribble, A. A. (1978). Rales - Diagnostic Uselessness of Qualitative Adjectives. American Review of Respiratory Disease, 113, 187.

Koran, L. M. (1975). The reliability of clinical methods, data and judgments. The New England Journal of Medicine, 13, 642-646.

Krumme, U. S. (1975). The case for criterion-referenced measurement. Nursing Outlook, 23 (12), 764-770.

Kulik, J. A., Kulik, C. & Hertzler, E. C. (1977). Modular college teaching with and without required Mediation. Journal of Personalized Instruction, 2, 70-75.

Light, R. J. (1971). Measures of response agreement for qualitative data: some generalizations and alternatives. Psychological Bulletin, 76 (5), 365-377.

Loudon, R. G. (1982 Fall). Auscultation of the lung. Clinical Notes on Respiratory Diseases, 21, 3-7.

Murphy, R. L. (1981). Auscultation of the lung: past lessons, future possibilities. Thorax, 36, 99-107.

Marini, J. J., Pierson, D. J., Hudson, L. D., & Lakshminarayan, S. (1979). The significance of wheezing in chronic airflow obstruction. American Review of Respiratory Disease, 120, 1069-1072.

Murphy, R. L., Gaensler, E. A., Holford, S. K., DelBono, E. A. & Epler, G. (1984). Crackles in the early detection of asbestosis. American Review of Respiratory Disease, 129, 375-379.

Nairn, J. R. & Turner-Warwick, M. (1969). Breath sounds in emphysema. British Journal of Diseases of the Chest, 63, 29-37.

Nath, A. R. & Capel, L. H. (1974). Inspiratory crackles and mechanical events of breathing. Thorax, 29, 695-698.

Nath, A. R. & Capel, L. H. (1974). Inspiratory crackles -- early and late. Thorax, 29, 223-227.

Nath, A. R. & Capel, L. H. (1980). Lung crackles in bronchiectasis. Thorax, 35, 694-699.

Nathoo, A. R. & Capel, L. H. (script producers) Med-One (proctor) (1981). Lung Sounds (video cassette). Nutley, N.J.: LaRouche-Hoffman.

Pardee, N. E., Martin, C. J., & Morgan, E. H. (1976). A test of the practical value of estimating breath sound intensity. Chest, 70 (3), 341-344.

Ploy-Song-Sang, Y., Martin R. R., Ross, W. R. D., Loudon, R. G., & Macklem, P. T. (1977). Breath sounds and regional ventilation. American Review of Respiratory Disease, 116, 187-199.

Polit, D., & Hungler, B. (1983). Nursing research, (2nd ed.). Philadelphia: J. B. Lippincott.

Price, S. A. & Wilson, L. M. (1978). Pathophysiology, New York: McGraw Hill.

Robertson, A. J. (1957). Before our time: Rales, Rhonchi and Laennec as recorded by Robertson, M.D. of Liverpool.

Lancet, 2, 417-423.

Schare, B. L. & Stehlin, C. S. (1981). What those breath sounds are telling you to do. R N, 44, 48-49.

Schilling, R. S. F., Hughes, J. P. W.,& Dingwall-Fordyce, I. (1955). Disagreement between observers in an epidemiological study of respiratory disease. British Medical Journal, 1, 65-68.

Schneider, I. C. & Anderson, A. E., Jr. (1965). Correlation of clinical signs with ventilatory function in obstructive lung disease. Annals of Internal Medicine, 62 (3), 477-485.

Thorne, G. L. (1983). Measuring student learning. Nurse Educator, 8, 7-10.

Weiss, E. B.,& Carlson, C. J. (1972). Recording of breath sounds. American Review of Respiratory Disease, 105, 835-839.

Wilkins, R. L., Dexter, J. R. & Smith, J. R. (1984). Survey of adventitious lung sounds terminology in case reports. Chest, 85 (4), 523-525.

Worklum, P., Holford, S. K., Delbono, E. A., & Murphy, R. L. (1982). The prevalence and character of crackles (rales) in young women without significant disease. American Review of Respiratory Disease, 126, 921-923.

APPENDIX A
PRELIMINARY STUDY ASSESSMENT TOOL

APPENDIX A

Subject #____________

Rater #____________

Date ____________

Time____________

AUSCULTATION & PALPATION

CRACKLES

0 = No crackles

1 = Late inspiratory; Early inspiratory and/or Late expiratory;

2 = Crackles throughout inspiration or expiration;

3 = Crackles throughout inspiration and expiration.

WHEEZES

0 = Absent

1 = Present

PALPATION

0 = Fremitus absent

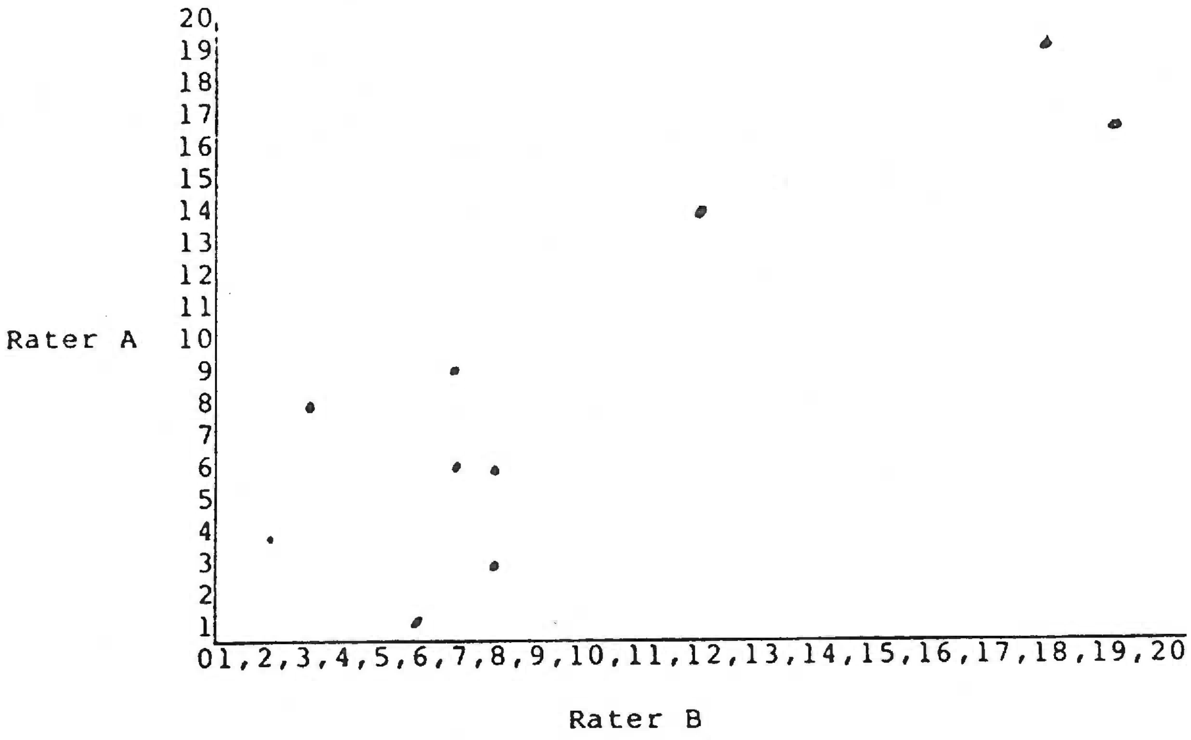
1 = Fremitus present

AUSCULTATORY AREA	CRACKLES	WHEEZES	PALPATION
Angle of Louis			
Left 3rd ICS, MCL			
Right 4th ICS, SB			
Left 5th ICS, AAL			
Right 5th ICS, AAL			
Left 7th ICS, PAL			
Right 7th ICS, PAL			
TOTAL			

APPENDIX B
PRELIMINARY STUDY TABLE AND SCATTERPLOT

APPENDIX B

SCATTERPLOT



Note. Scatterplot shows a positive relationship between Raters A and Rater B.

APPENDIX B

Table B-1
Crackle Grades Frequency Distribution

Subject	Rater A	Rater B
1	3	8
2	17	19
3	6	7
4	8	3
5	14	12
6	0	0
7	9	7
8	4	2
9	0	0
10	1	6
11	6	8
12	0	0
13	19	18

APPENDIX C
HUMAN RESEARCH EXEMPTION

THE OREGON HEALTH SCIENCES UNIVERSITY

MEMO

Date: June 26, 1984

To: Mary M Farland R.N., M.S.N. MacH 4162

From: The Committee on Human Research L 106

Subject: ORS# EXEMPT
TITLE: Interrater Reliability: A Study of Adventitious Pulmonary Sounds Using a Grade-rated Scale.

Your above entitled study falls under category # 4 and is considered to be exempt from the requirement for Committee on Human Research review. We, therefore, have put your study into our exempt files and you will receive no further communication from our Committee concerning this study.

If the involvement of human subjects changes in this study you should contact the Committee on Human Research to find out whether or not these changes need to be reviewed.

If you have any questions regarding the status of this study, please contact Donna Buker at X 7887.

Sincerely,

[redacted]

Michael A. Wall, M.D., Chairman
Committee on Human Research

APPENDIX D
CRITERION REFERENCED OBJECTIVES

APPENDIX D

Criterion Referenced Objectives

At the end of the instruction session the nurse raters will be able to:

1. Identify ten auditory lung sound simulations as either crackles or wheezes.

2. Place the grade value next to the respiratory timing segments for crackles.

3. Place the grade value next to the respiratory timing segment for wheezes.

4. Locate on the chest diagrams the nine areas to be auscultated (to be marked X).

5. Order in sequence the procedure for patient preparation on a written test.

6. Demonstrate the understanding of the importance of preventing bias caused by the sharing of auscultatory information from one rater to another.

APPENDIX E
CRITERION TEST FOR GRADE-RATERS

APPENDIX E

Code no. ____________
Date ____________

CRITERION TEST FOR GRADE-RATERS

1. You will hear several auditory lung sound simulations, on a tape. Please identify them as either: A. Crackles/rales, or B. Wheezes/rhonchi. Place the letter beside the number.

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

2. Place the numerical value of the crackle/rale during segments of the respiratory cycle.

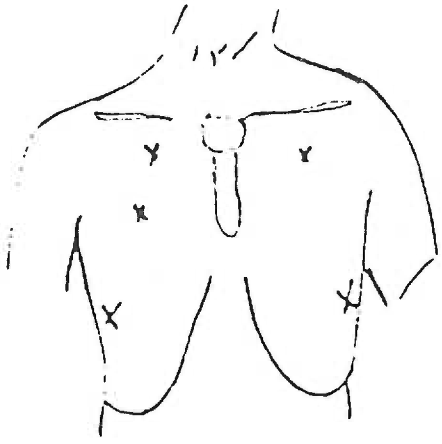
A. Occurring only during early inspiration, (1)
B. Occurring during more than one phase of inspiration but not during expiration, (4)
C. Occurring during both inspiration and expiration, (6)
D. Occurring only during expiration, (5)
E. Occurring only during mid-inspiration, (2)
F. No crackles/rales heard, (0)
G. Occurring only during late inspiration, (3)

3. Place the numerical value of wheezes/rhonchi during segments of the respiratory cycle.

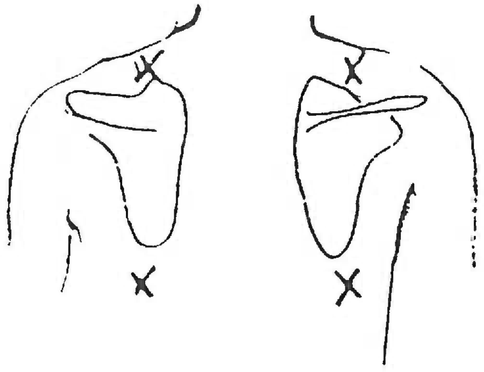
A. Occurring only during expiration, (2)
B. Occurring during both inspiration and expiration, (3)
C. No wheezes or rhonchi heard, (0)
D. Occurring during inspiration, (1)

4. Place X's over the 9 lung areas to be auscultated for this study.

ANTERIOR and LATERAL



POSTERIOR



5. Place the items below which describe patient preparation for lung auscultation for this study in proper sequence.

A. patient undress to waist,
B. provide privacy, comfortable room, good lighting, warm stethoscope, relaxed atmosphere,
C. coach patient: cough, open mouth breathe deeper than normal,
D. patient sitting upright,

1st (B)
2nd (A)
3rd (D)
4th (C)

6. Grade-raters who discuss findings are not in danger of biasing the study. Circle one:

A. True
[✓ B. False.]

APPENDIX F
AUSCULTATORY RATING TOOL

APPENDIX F

Auscultatory Rating Scale

Date____________

Subject_________
Rater__________
Time___________

Crackles/rales

0 = no crackles
1 = only early inspiration
2 = only mid-inspiration
3 = only late inspiration
4 = occuring during more than phase of inspiration but not during expiration
5 = only expiration
6 = both inspiration and expiration.

Wheezes/rhonchi

0 = no wheezes
1 = only during inspiration
2 = only during expiration
3 = both inspiration and expiration

AUSCULTATORY AREAS		CRACKLES		WHEEZES	
		right	left	right	left
POSTERIOR:	Above scapula, at mid-scapular line				
	At level T 10, mid-scapular line				
ANTERIOR:	Right 2nd ICS, MCL		▨		▨
	4th ICS, MCL		▨		▨
	Left 3rd ICS, MCL	▨		▨	
LATERAL:	7TH ICS, anterior-axillary line.				
	Subtotal________				
TOTAL					

Primary Diagnosis______________________(to be completed by investigator).

AN ABSTRACT OF THE THESIS OF

Julie R. Wellman

For the MASTERS OF NURSING

Date of Receiving this Degree: June 14, 1985

Title: A STUDY OF ADVENTITIOUS BREATH SOUNDS: INTERRATER AGREEMENT. [redacted]

Approved:

Mary McFarland, R.N., M.S.N., Thesis Advisor

Widespread application of adventitious breath sounds as predictors for nursing action is dependent upon both the validity of the breath sound terminology used and the reliability of auscultated results. The purpose of this descriptive study was to determine if nurse-raters could agree on the presence or the absence of adventitious breath sounds using the current American Thoracic Society (ATS) nomenclature.

Nurse-raters were instructed in the use of the ATS terms, crackle and wheeze. Each nurse-rater was required to obtain one-hundred percent on a criterion referenced test. Next, they practiced the auscultation procedure and the reporting method on patients.

The study was conducted on acute care hospital units where patients could be expected to have adventitious breath sounds. Patients were positioned and coached to control for

the rate and depth of respirations. Using the diaphragm of binaural, single tubed stethoscopes, three nurse-raters examined each patient at the same time and reported their findings independently on the assessment tool. They auscultated nine specified apical and basilar lung areas on the anterior and posterior chest. Prior to auscultation it was unknown to the raters if the patient had adventitious sounds. Raters did not share information verbally or non-verbally during the study. Thirty-four patients were auscultated.

Findings showed that crackles were heard most often in the basilar lung areas while wheezes were more evenly distributed throughout the chest. In this study the nurse-raters agreed upon the absence of adventitious breath sounds. However, when an adventitious sound was heard by one nurse-rater, agreement by the other nurse-raters did not exceed that which would have been expected by chance alone.

Further study is needed to support or refute these findings. To potentially increase the interrater reliability and the validity of the instrument the following modifications are suggested: Increase the sample size; Increase the number of raters; Simplify the instrument; Use a measurable validating device.