Accuracy and reliability of physiotherapists in the interpretation of tape-recorded lung sounds

The effect of clinical experience on the accurate and reliable interpretation of auscultated lung sounds was examined by comparing 16 new graduates (Group A) with 16 experienced cardiopulmonary physiotherapists (Group B). Subjects listened to a tape comprising six different lung sounds, with each sound repeated three times in random order. Group B tended to be more accurate than Group A for five of the six sounds but the difference was significant only for the normal breath sound ($\chi^2 = 6.72, p = 0.01$). Intra-rater reliability was poor; for any individual sound, a maximum of nine subjects recorded the same response on all three occasions. There were no significant inter-group differences in reliability. In conclusion, clinical experience had no significant effect on accuracy and reliability.

Key words: Auscultation; Physical Therapy; Respiratory Sounds

Auscultation is frequently used by physiotherapists as an assessment tool in patient examination. The value of the technique is enhanced when the information gained is accurately communicated to other clinicians (Wilkins et al 1989). A confounding factor when reporting lung sounds is the lack of consistency in the terms used (Pasterkamp et al 1987, Ward 1989, Wilkins et al 1984, Wilkins et al 1989), despite attempts at standardisation (American College of Chest Physicians-American Thoracic Society Joint Committee on Pulmonary Nomenclature 1975, Mikami et al 1987). In addition, the subjective nature of lung sound interpretation may result in variation among clinicians in their ability to accurately and reliably interpret lung sounds.

To date, only two published studies have examined the accuracy and reliability of physiotherapists in the interpretation of tape-recorded lung sounds (Aweida and Kelsey 1990, Brooks et al 1993) and the number of studies examining the same in other health professionals is limited (Godfrey et al 1969, Schilling et al 1955, Smyllie et al 1965, Spiteri et al 1988). The two investigations which studied physiotherapists included 27 subjects who had not specialised in cardiopulmonary care (Aweida and Kelsey 1990) and 26 subjects who had at least one year of current, exclusive practice in cardiopulmonary care (Brooks et al 1993). In both studies, subjects were required to listen to a tape of five different lung sounds, each repeated three times in random order, and identify the sounds by choosing from a list of 10 standardised sounds. The bell of the stethoscope was used in these studies, which is in contrast to the recommendations in several texts where the use of the diaphragm is advocated when listening to normal and abnormal lung sounds (Humberstone 1990, Parker and Middleton 1993, Shapiro et al 1991, Wilkins 1990). The sounds examined were: wheezes (high pitched and low pitched); bronchial breath sounds; stridor; and pleural rub. Thus no complex sounds such as inspiratory crackles and expiratory wheezes were included. The absence of crackles and normal breath sounds in the two studies is important, as these are among the most commonly heard sounds. Furthermore, restricting the choice of responses is not representative of clinical practice. The authors concluded that both groups of physiotherapists were neither accurate nor reliable in the interpretation of tape-recorded lung sounds, and a comparison of the findings of the two studies showed no influence of experience.

The aim of the present study was to compare the accuracy and reliability of
inexperienced and experienced cardiopulmonary physiotherapists in the interpretation of tape-recorded lung sounds. We hypothesised that physiotherapists who were experienced in cardiopulmonary care would be more accurate and more reliable in the interpretation of lung sounds than new graduates. The present study tested these hypotheses in two groups of physiotherapists: newly qualified physiotherapists and experienced cardiopulmonary physiotherapists.

Method

Subjects
The study population comprised 16 physiotherapists (12 females) who had graduated from Curtin University School of Physiotherapy in 1992 (Group A) and 16 experienced cardiopulmonary physiotherapists (15 females) (Group B). Subjects in Group A were required, since graduating, to have had at least one three-month rotation on an adult surgical (thoracic, cardiac or general) ward, intensive care unit, medical respiratory ward or respiratory outpatient department. For inclusion in Group B, subjects were required to be currently working in cardiopulmonary care and, in addition, satisfy at least one of the following criteria: be currently or previously employed as a senior cardiopulmonary physiotherapist; hold a postgraduate diploma in cardiopulmonary rehabilitation; or have at least five years post-registration clinical experience with at least one year of recent experience in cardiopulmonary care. All physiotherapists in Group B had utilised their knowledge and skills in cardiopulmonary physiotherapy during the 12 months prior to the study. Subjects were excluded from Group B if they were only involved in the management of children under nine years of age, since the frequencies of normal breath sounds in this age group vary from that of adults (Hidalgo et al 1991). Any subject who had a history of an auditory deficit or a history of Meniere’s disease, since this affects the audition of both low and high frequency sounds, was excluded from participation (Cull and Will 1991).

Instrumentation
The six lung sounds selected for the study were: 1) normal breath sounds; 2) bronchial breath sounds; 3) fine inspiratory crackles; 4) coarse inspiratory and expiratory crackles; 5) monophonic expiratory wheezes and 6) coarse inspiratory crackles with polyphonic expiratory wheezes. In order to minimise extraneous noises, the tape used for the study was professionally recorded from a newly-purchased tape of lung sounds produced by Wilkins et al (1988). Recording took place in the studios of the Australian Broadcasting Corporation (Perth, Western Australia) and was carried out by a recording engineer. A high quality metal cassette tape was used. Each lung sound was repeated for four respiratory cycles and separated by a 15 second interval to allow subjects time to respond. In order to assess reliability, the six sounds were recorded three times in random order, resulting in a total of 18 sounds to be identified. The number of each sound (ie 1-18) was indicated on the tape by an announcer.

Procedures
The superintendent physiotherapists of the major hospitals in the Perth metropolitan area were contacted to assist with subject recruitment. Potential subjects were given an information consent document outlining the purpose of the study and the testing procedures. The study was approved by the Human Ethics Committee of Curtin University of Technology and consent was obtained in writing from all subjects prior to participation.

Testing took place over a five week period commencing in February 1994 and was performed in a quiet room at
A new Littman Classic II stethoscope was used to listen to the tape of lung sounds which was played through a basic mono-speaker cassette player. The same tape and cassette player were used throughout the study. To ensure that the quality of sound transmission from the speaker was uniform for each subject, no more than four subjects were tested simultaneously at each session. For the same reason, subjects were required to hold the diaphragm of the stethoscope five centimetres above the surface of the cassette speaker. This position was standardised by a protruding marker. After reading the introductory sheet, an example of a lung sound was played to the subjects, followed by an interval of 15 seconds before testing commenced. The tape was then played through once and subjects were asked to identify each of the lung sounds heard and record their responses on the form shown in Figure 1. A separate form was provided for each of the 18 sounds. Subjects were not informed that any of the sounds would be repeated.

In an attempt to avoid subjects conferring, subjects were instructed not to communicate during or following the testing procedure. One of the researchers was present throughout all testing sessions, to ensure these instructions were complied with. Once all testing had been completed, participants were sent a questionnaire which included five questions asking for comments relating to different aspects of the study methodology, for example, whether subjects considered that the instructions provided on the day of testing were clear. Subjects were also asked to comment on whether they felt the results of the study would accurately reflect their clinical expertise in auscultation. Finally, a space was left for any additional comments and suggestions on how the study might be improved. Subjects were informed that responding to the questionnaire was optional.

### Data analysis

#### Accuracy

This was determined by adding the number of correct responses for each sound and converting to a percentage. The possible total number of correct responses for each individual sound was 48, because all of the 16 subjects in each group heard each of the six sounds three times. A response was considered accurate if the nature of the sound, where appropriate, and the primary sound were both identified, for example, coarse crackles. A response was regarded as inaccurate if additional incorrect sounds were included. The exception to this was if the subject included that the breath sound was diminished. The timing of the sound within part of the respiratory cycle (i.e., early, middle, or late) was disregarded in the analyses, due to the arbitrary division of these phases within a cycle and the absence of a standardised correct answer for timing in the original tape. An inter-group comparison of the number of correct answers for each lung sound was made using the chi-square ($\chi^2$) test with

<table>
<thead>
<tr>
<th>Lung Sound</th>
<th>Group A $n = 16$</th>
<th>Group B $n = 16$</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal breath sounds</td>
<td>6 (12.5)</td>
<td>18 (37.5)</td>
<td>0.01</td>
</tr>
<tr>
<td>Bronchial breath sounds</td>
<td>10 (20.8)</td>
<td>15 (31.3)</td>
<td>0.35</td>
</tr>
<tr>
<td>Fine inspiratory crackles</td>
<td>8 (16.7)</td>
<td>9 (18.8)</td>
<td>1.00</td>
</tr>
<tr>
<td>Coarse inspiratory and expiratory crackles</td>
<td>25 (52.1)</td>
<td>30 (62.5)</td>
<td>0.40</td>
</tr>
<tr>
<td>Monophasic expiratory wheezes</td>
<td>25 (52.1)</td>
<td>22 (45.8)</td>
<td>0.68</td>
</tr>
<tr>
<td>Coarse inspiratory crackles with polyphonic expiratory wheezes</td>
<td>16 (33.3)</td>
<td>19 (39.6)</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Group A - new graduates; Group B - experienced cardiopulmonary physiotherapists; $n$ - number of subjects; numbers in table are the number of correct responses out of a maximum of 48; figures in parentheses represent the percentage of correct responses; $p$ - probability value for intergroup comparisons from chi-square test with Yates’s correction for continuity.
From Page 181

Yates's correction for continuity (Altman, 1991). A probability (p) value of less than 0.05 was regarded as significant.

Reliability

Reliability was considered irrespective of accuracy. To determine intra-rater reliability, the number of subjects who gave identical responses on two of the three occasions, and all three occasions that each of the six sounds was played, was counted. Data for intra-rater reliability were analysed descriptively. Inter-group reliability was compared using Fisher's exact test (Altman 1991) to examine differences in the numbers of subjects in each group who reliably interpreted each lung sound on two of the three occasions and all three occasions. A p value of less than 0.05 was regarded as significant.

Results

Accuracy

The number of correct responses and the percentage of correct responses by group for each of the six lung sounds are presented in Table 1.

There was a trend towards subjects in Group B being more accurate than subjects in Group A, but this reached significance for the normal breath sound only. Subjects in Group A tended to include more additional, incorrect adventitious sounds (almost invariably crackles) in their responses than subjects in Group B. When the correct sound was a monophonic expiratory wheeze, subjects in Group A incorrectly included crackles in 43.8 per cent of responses, compared with only 12.5 per cent of the responses from subjects in Group B. This difference was significant ($\chi^2 = 10.09, p = 0.002$). A normal breath sound was incorrectly interpreted as a bronchial breath sound in 39.6 per cent of Group A's responses and 20.8 per cent of the responses from Group B ($\chi^2 = 4.08, p = 0.04$). Analysis of the responses for the complex sound (coarse inspiratory crackles with polyphonic expiratory wheezes) showed that both of the primary sounds, ie crackles and wheezes, were correctly identified in 66.7 per cent and 70.8 per cent of responses from Groups A and B respectively ($\chi^2 = 0.05, p = 0.83$). The nature of the wheeze component (ie polyphonic) of the sound was correctly identified in 85.4 per cent of the responses from Group A and in 83.3 per cent of responses from Group B, whereas the nature of the crackles, (ie coarse) was correctly identified in only 37.5 per cent and 41.7 per cent of responses from Groups A and B respectively. These differences were not significant. Combining the data from both groups showed that the ability to accurately identify both the primary sound and the nature of the polyphonic expiratory wheeze component of the sound (79.2 per cent of responses) was significantly better than identification of the coarse, inspiratory crackle (39.6 per cent of responses) ($\chi^2 = 8.08, p = 0.005$).

Combining the number of correct responses for the six sounds gave pooled accuracy rates of 31.3 per cent and 39.2 per cent for Groups A and B respectively ($\chi^2 = 3.68, p = 0.06$).

<table>
<thead>
<tr>
<th>Lung Sound</th>
<th>Group A</th>
<th>Group B</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 16</td>
<td>n = 16</td>
<td></td>
</tr>
<tr>
<td>Normal breath sounds</td>
<td>2/3</td>
<td>3/3</td>
<td>0.60</td>
</tr>
<tr>
<td>(68.8) (6.3)</td>
<td>(50) (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bronchial breath sounds</td>
<td>5/2</td>
<td>10/2</td>
<td>0.48</td>
</tr>
<tr>
<td>(31.3) (12.5)</td>
<td>(62.5) (12.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fine inspiratory crackles</td>
<td>9/3</td>
<td>9/5</td>
<td>0.44</td>
</tr>
<tr>
<td>(56.3) (18.8)</td>
<td>(56.3) (31.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coarse inspiratory and</td>
<td>7/4</td>
<td>9/0</td>
<td>0.47</td>
</tr>
<tr>
<td>expiratory crackles</td>
<td>(43.8) (6.3)</td>
<td>(56.3) (0)</td>
<td></td>
</tr>
<tr>
<td>Monophonic expiratory wheeze</td>
<td>8/2</td>
<td>8/4</td>
<td>0.42</td>
</tr>
<tr>
<td>(50) (12.5)</td>
<td>(50) (25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coarse inspiratory crackles with</td>
<td>8/3</td>
<td>7/6</td>
<td>0.30</td>
</tr>
<tr>
<td>polyphonic expiratory wheezes</td>
<td>(50) (18.8)</td>
<td>(43.8) (37.5)</td>
<td></td>
</tr>
</tbody>
</table>

Group A - new graduates; Group B - experienced cardiopulmonary physiotherapists; n - number of subjects; figures in table represent the numbers of subjects who gave the same response on 2/3 (two of the three) and 3/3 (all three occasions); figures in parentheses represent percentages; p - probability value for intergroup differences from Fisher's exact test.
Reliability

Table 2 presents the number of subjects in each group who recorded the same response on two out of the three and all three occasions of listening to each of the six lung sounds. Intra-rater reliability was poor, especially for the identification of normal and bronchial breath sounds. For any individual sound, a maximum of nine (28 per cent) subjects (three in Group A and six in Group B) recorded the same response on all three occasions on which the sound was played. There were no significant differences between the groups in the ability to reliably identify any of the six lung sounds.

Questionnaire

Fifteen questionnaires were returned: four from subjects in Group A and 11 from subjects in Group B. Most respondents commented that the testing procedure was clear and easy to follow. Suggestions for improvements to the study were concerned with changes to the response sheet. The majority of subjects commented that they felt the results of lung sound interpretation alone did not fully reflect clinical performance, since chest auscultation invariably is performed with knowledge of a patient's history and in conjunction with other assessment procedures.

Discussion

The findings of this study show that experienced cardiopulmonary physiotherapists are only slightly more accurate than new graduates in the identification of tape-recorded lung sounds. There was no evidence to support the hypothesis that clinical experience improves the ability to reliably identify lung sounds heard on a tape.

It was decided to consider a response as accurate even if, in addition to the correct response, the subject also recorded that the breath sound was diminished. This was considered acceptable given that in the clinical situation, the interpretation of diminished breath sounds is generally made by comparison with the intensity of breath sounds heard over an area of normal lung. Clearly, this comparison was not possible in the present study.

The increased frequency with which new graduates included extra, incorrect adventitious sounds in their responses adversely affected the accuracy for Group A. Experienced physiotherapists were more specific in their responses, possibly indicating a greater degree of certainty when identifying sounds. Interestingly, it was found that the majority of extra, incorrect adventitious sounds recorded by new graduates were crackles. Although every attempt was made to maintain the high quality of the original lung sound tape during the recording, it is possible that extraneous sounds might have been interpreted as crackles. Although this was a limitation for both groups, it is possible that new graduates tended to mistake such sounds as adventitious crackles and that this accounted for some of the differences between the groups. The new graduates were not exposed to tape recordings of lung sounds during their training and thus would have heard normal breath sounds only in the clinical situation. The finding that the primary sound was more accurately identified than the nature of the sound was not unexpected, as it was anticipated that subjects would focus their attention on the most important component of the breath sound.

Both groups were more accurate in identifying wheezes than crackles. This may be explained by the fact that a continuous sound such as a wheeze is of a longer duration, usually more than 0.25s, has a distinct musical quality and may be more obvious than the shorter discontinuous sound of a crackle (Murphy and Holdford 1980).

The overall findings of the present study are similar to those of Aweida and Kelsey (1990) and Brooks et al (1993). The higher accuracy rates in these studies (pooled accuracy rates of 47 per cent for physiotherapists who had not specialised in cardiopulmonary care and 50 per cent for those with at least one year of current, exclusive practice in cardiopulmonary care) may be explained by the choice of individual lung sounds used in these studies. Bronchial breath sounds were accurately identified in 59 per cent of responses from non-experienced physiotherapists (Aweida and Kelsey 1990) and in 77 per cent of responses from experienced physiotherapists (Brooks et al 1993). However, subjects in the present study had difficulty in distinguishing bronchial breath sounds from normal breath sounds. This was especially true for the new graduates. Therefore it might be hypothesised that the accuracy rates reported by Aweida and Kelsey (1990) and Brooks et al (1993) would have been lower had they included normal breath sounds in their studies. The inclusion in the present study of some of the more commonly heard sounds, such as fine crackles, in place of the very distinctive sounds of pleural rub and stridor, also may have adversely affected the accuracy rates.

We chose to consider reliability independently of accuracy, on the basis that if a systematic error was found to occur, for example repeatedly misinterpreting coarse crackles as fine crackles, this might more easily be corrected with training than if a random error was found. However, both intra-rater and inter-rater reliability were poor. Reliability among raters was also poor in both of the previous studies of physiotherapists and there was no evidence that it improved with experience (Aweida and Kelsey 1990, Brooks et al 1993). Direct comparison of the present study with studies investigating the inter-rater reliability of physicians is difficult, as none of the reported studies examined auscultation in isolation from other clinical signs (Godfrey et al 1969, Schilling et al 1955, Smylie et al 1965, Spiteri et al 1988). Nevertheless, these studies all reported that inter-rater reliability was poor and was independent of the degree of clinical experience. Furthermore, in experienced physicians who underwent a training program, there was only marginal improvement in reliability (Godfrey et al 1969).
Several limitations of the present study must be addressed. Tape-recorded lung sounds are not totally representative of lung sounds heard through a stethoscope placed on the chest wall. Another problem encountered when listening to a tape of lung sounds is the lack of visual feedback gained by observing thoraco-abdominal movements. This may cause problems in differentiating between sounds occurring during inspiration and expiration. As auscultation is only one of several assessment tools used by physiotherapists in cardiopulmonary care, examination of the influence of clinical experience on auscultation in isolation is limited. This point was expressed by participants both at the time of the study and in response to the questionnaire. Although the study groups were distinguished by the degree of clinical experience in cardiopulmonary care, examination of the influence of clinical experience on auscultation in isolation is limited. This point was expressed by participants both at the time of the study and in response to the questionnaire. Although the study groups were distinguished by the degree of clinical experience in cardiopulmonary care, it was not possible to obtain an objective measure of the amount of auscultation recently performed by subjects. It is possible that some subjects in Group A might have spent more time auscultating patients' lungs in the months preceding the study than subjects in Group B, especially since Group B subjects may have increased administrative or teaching responsibilities. The relatively small number of subjects studied may also have affected the validity of the findings. However, the strict inclusion criteria required to obtain a homogeneous group of subjects, and the need for easy accessibility to the researchers in the Perth metropolitan area, limited the number of potential subjects.

The findings of poor accuracy and reliability and the lack of influence of clinical experience on the technique of auscultation should not be extrapolated to other skills used in chest assessment. It is recommended that future studies examine the accuracy and reliability of auscultation in vivo both separately and in conjunction with other chest assessment skills.

Conclusions

The findings of this study suggest that, when listening to tape-recorded lung sounds, the accuracy and reliability of both non-experienced and experienced cardiopulmonary physiotherapists in the interpretation of these sounds is poor. The only significant difference between the groups was that the experienced cardiopulmonary physiotherapists were more accurate in the interpretation of the normal breath sound.

Acknowledgements

This study formed an undergraduate research project and the authors gratefully acknowledge the contributions of A Cooper, J Depiazz, T Fischer, S Hendren, H McGlew, R Prosser, T Robinson and K Whelan. We also wish to thank the superintendent physiotherapists for their assistance in recruiting subjects; the physiotherapists who participated in the study; K Akers, recording engineer, Australian Broadcasting Corporation, Perth, WA; and Dr J Sommers, for his statistical advice.

References


