# The Underestimation of Segmental Defect Size in Radionuclide Lung Scanning

N.W. Morrell, K.S. Nijran, B.E. Jones, T. Biggs and W.A. Seed

*Departments of Medicine and Nuclear Medicine, Charing Cross and Westminster Medical School, London, United Kingdom*

Criteria used to place ventilation-perfusion lung scans into categories with different probabilities for pulmonary embolism depend largely on the size and anatomical distribution of defects recognized. These criteria assume that actual segmental defects *appear* segmental on the lung scan. This study examined the accuracy with which four experienced observers were able to estimate the size of defects of known anatomical location and size, using images of segmental defects in ventilation produced with a bronchoscopic technique and $^{81m}$Kr. Of the 24 segmental defects produced in this study, 17% were interpreted as being <25% of a segment; 23% were interpreted as being 25%–50% of a segment; 17% were interpreted as 50%–75% of a segment; 40% were interpreted as being 75%–100% of a segment and 4% were interpreted as being >100% of a segment. Intra- and interobserver agreement as assessed by the Kappa statistic varied with the number of size categories used but was generally poor. Underestimation of defect size observed in this study may explain why many patients with pulmonary embolism do not have high probability scans. We conclude that the subjective impression of the size of a defect on a lung scan is an unreliable indication of a defect's true segmental or subsegmental nature and that scoring systems based on these criteria should be viewed with caution.

**J Nucl Med 1993; 34:370–374**

In the interpretation of ventilation-perfusion lung scans for the diagnosis of pulmonary embolism, a number of scoring systems have been devised for categorizing the scans according to probability of pulmonary embolism (*1–3*). These classifications depend largely on the presence of ventilation-perfusion mismatch and the size and anatomical distribution of defects recognized. In the recent Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) study (*3*), defects were classified into small (<25% of a segment), moderate (25%–75% of a segment) or large (>75% of a segment). In that study for example, two large mismatched segmental defects on a perfusion scan carried a high probability for pulmonary embolism whereas the presence of two moderate sized defects carried only an intermediate probability for embolism. The high probability lung scan has a specificity approaching 90% for the detection of pulmonary emboli when judged by pulmonary angiography (*3,4*). However, the majority of patients with pulmonary emboli do not have high probability lung scans (*3*). One reason for this may be an underestimation of defect size by those interpreting scans because the size of a defect on the lung scan may not be a reliable indication of a defect's true segmental or sub-segmental nature. The aim of this study was to determine the accuracy with which experienced observers are able to estimate the size of defects of known anatomical location and size on the lung scan.

Received Jun. 10, 1992; revision accepted Oct. 19, 1992.

For reprints or correspondence contact: Dr. N.W. Morrell, Department of Medicine, Charing Cross Hospital, Fulham Palace Road, London, United Kingdom, W6 8RF.

## METHODS

### The Production of Defects of Known Anatomical Location and Size

Healthy volunteers with normal lung function underwent fibreoptic bronchoscopy (Olympus BF-1T20D) under local anesthesia positioned supine over a gamma camera. The gamma camera was fitted with a low-energy, long-bore, parallel-hole collimator to reduce septal penetration and improve resolution using the relatively high energy $^{81m}$Kr. The energy window was set at 20%. Prior to bronchoscopy, control ventilation scans were obtained in the posterior, posterior-oblique and lateral positions with $^{81m}$Kr. Between 250K and 300K counts were collected in each position. A fine balloon catheter was then inserted through the instrument channel of the bronchoscope and the balloon inflated under direct vision in a selected segmental bronchus. With the occlusion in place, images of the resulting defect in ventilation were obtained in the same views as for the controls collecting 250K to 300K counts in each view (Fig. 1). We shall refer to this image as the "negative" image of the segment involved. After allowing time for the washout and decay of $^{81m}$Kr, the site and surface projection of the occluded segment were then confirmed by ventilating the isolated segment independently with $^{81m}$Kr and air introduced with a hypodermic syringe via the lumen of the balloon catheter (Fig. 2). This we shall refer to as the "positive" image of that segment. This also allowed detection of any leak of isotope around an imperfect occlusion. These images (25K–30K counts) were ob-

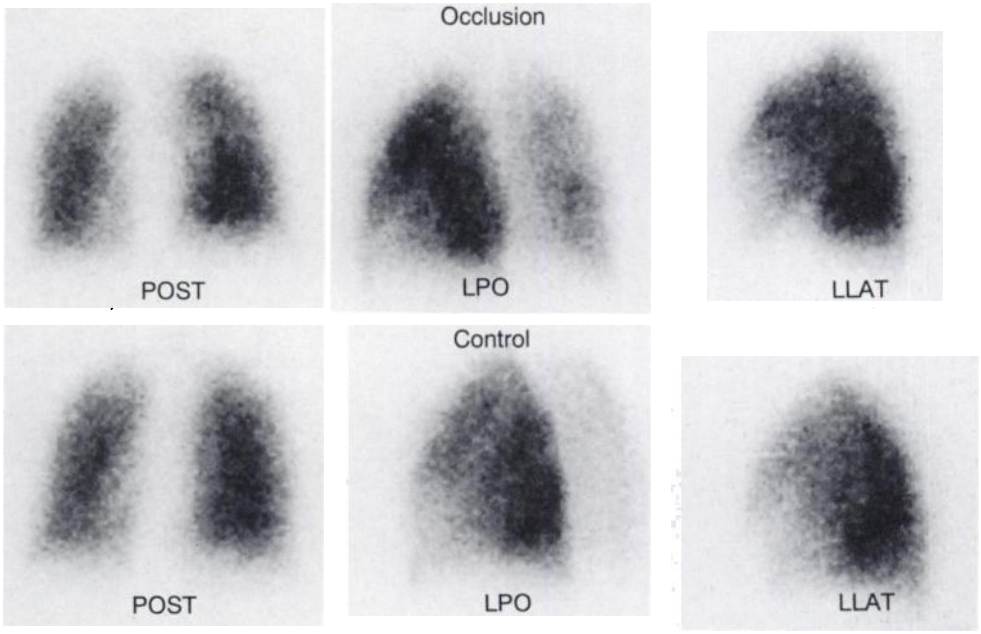


**FIGURE 1.** Krypton-81m ventilation images produced by occlusion of the bronchus to the anterior segment of the left lower lobe. Beneath the experimental images are the corresponding control images. This defect is only clearly visible in the left lateral view. (POST = posterior, LPO = left posterior oblique, LLAT = left lateral).

same projections. Further details of this technique have been previously published (*5*).

### Analysis of Images

Images were obtained of all 18 lung segments, the majority on at least two occasions. Twenty-four "negative" images were selected as being technically acceptable and free of ventilation defects caused by the instillation of local anesthetic. These 24 images were used in the subsequent interpretations. Seventeen of the 18 lung segments were represented in this selection. The single segment omitted was the medial segment of the right lower lobe which was not detectable in any view, including the anterior, despite its presence being detected on a "positive" image.

Black and white polaroid images of the 24 defects were presented to four experienced nuclear medicine physicians. The view selected in each case was that which demonstrated the segmental defect at its largest. The observers were informed of the location of a defect and were told that a defect could be smaller, equal to or larger than a full segment. The observers were instructed to place each defect into one of five categories based on their subjective estimation of its size: <25% of a segment, 25%–50% of a segment, 50%–75% of a segment, 75%–100% of a segment or >100% of a segment. The scan readings were performed twice, at least 1 mo apart.

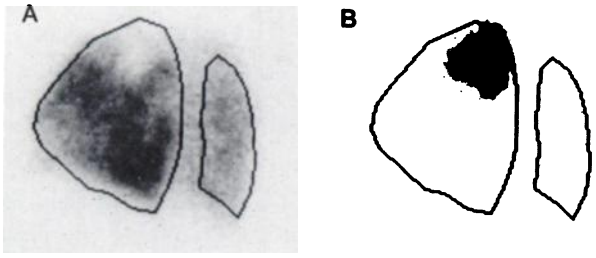


**FIGURE 2.** "Negative" (A) and "positive" (B) images of a defect involving the apico-posterior segment of the left upper lobe. The lung outline has been added from a control image to aid localization.

Two methods were chosen to assess intra- and interobserver agreement. The first was descriptive in which the estimate of each defect's size by an observer in a reading was compared with the estimate from each other observer in turn. The percentage mean complete agreement between any two readings could then be calculated and the frequency with which observers disagreed by one or more categories. The second method was the Kappa test, which is a statistical index of agreement beyond that expected by chance (*6*). Kappa has a value between −1 and +1. A Kappa of 0 implies no agreement. A positive Kappa implies agreement beyond that expected by chance and a negative Kappa implies disagreement beyond chance. Kappa may be subjected to statistical evaluation to show whether it is significantly different from zero and the difference between two independent Kappas may be tested for significance (*7*). Intra- and interobserver agreement were also estimated for three size categories as used in the PIOPED (*3*) study: <25% of a segment, 25%–75% of a segment and >75% of a segment to test the effect of such categorization on agreement.

### Spatial Resolution of $^{81m}$Kr Compared to $^{99m}$Tc

Our experimental method was based on the production of defects in ventilation using $^{81m}$Kr. For our results to have application to $^{99m}$Tc perfusion images, the spatial resolution of our system using the two isotopes would have to be similar. Therefore, as a measure of spatial resolution we measured the full width at half maximum (FWHM) for both isotopes with the gamma camera and collimator used in our experiments. The FWHM is analogous to the minimum distance required between two radioisotope point sources for them to be separately resolved.

A parallel line source (20 mm separation) was used and a lead transmission phantom (50 mm deep) with slits of 0.5 mm thickness. The collimator energy window settings were at 20% as before. The line sources were filled either with a solution of

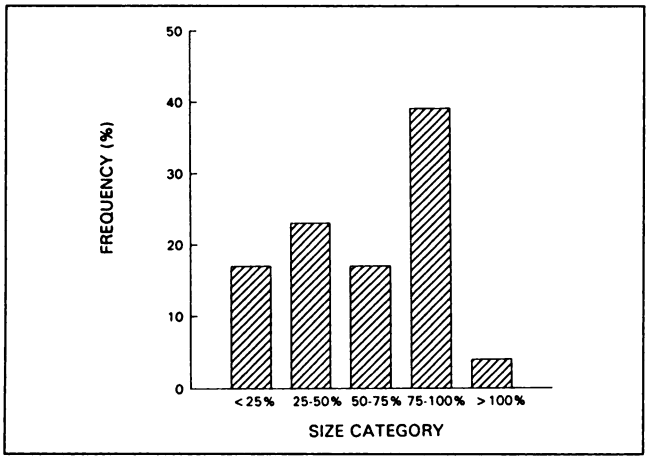


**FIGURE 3.** Bar chart illustrating the frequency (%) with which the four observers placed actual segmental defects into one of five size categories.

$^{99m}$Tc or with $^{81m}$Kr gas at a constant flow rate. Images were obtained both at the camera face (0 cm) and with attenuation from 10 cm water. The data were displayed in graphical form and the FWHM was calculated in millimeters.

## RESULTS

### Estimation of Defect Size

These results are represented graphically in Figure 3. Of these known segmental defects, 17% were interpreted as being <25% of a segment; 23% were interpreted as being 25%–50% of a segment; 17% were interpreted as 50%–75% of a segment; 40% were interpreted as being 75%–100% of a segment and 4% were interpreted as being >100% of a segment.

### Intraobserver Agreement

The mean percentage complete intraobserver agreement using five size categories was 56.8% (s.e.m. 1.33) (Table 1). There was disagreement by one scan category in 34.1% of readings, by two scan categories in 5.7% and by 3 categories in 3.4%.

The mean percentage complete agreement using three size categories increased to 75% (s.e.m. 2.93) ($p < 0.01$). There was disagreement by one size category in 21.6% of readings and by two categories in 3.4%.

**TABLE 1**
Intraobserver Agreement

| | 5 Categories | | 3 Categories | |
|---|---|---|---|---|
| | % | Kappa | % | Kappa |
| Observer 1 | 54.5 | 0.413 | 81.8 | 0.720 |
| Observer 2 | 59.1 | 0.378 | 68.2 | 0.486 |
| Observer 3 | 59.1 | 0.445 | 75 | 0.647 |
| Observer 4 | 54.5 | 0.379 | 72.7 | 0.495 |
| Total | 56.8 | | 75 | |

**TABLE 2**
Interobserver Agreement

| | 5 Categories | | 3 Categories | |
|---|---|---|---|---|
| | % | Kappa | % | Kappa |
| First reading | 35.6 | 0.132 | 54.5 | 0.268 |
| Second reading | 47.7 | 0.297 | 62.9 | 0.418 |
| Total | 41.4 | | 58 | |

Intraobserver agreement as measured with Kappa ranged from 0.378 to 0.445 with five size categories. With three size categories Kappa ranged from 0.486 to 0.720 (Table 1). All values of Kappa were significantly greater than zero. Within observers, there was a tendency for agreement to improve with a smaller number of categories although the number of observations ($n = 24$) is probably too small to show a real difference if one existed (*6*).

### Interobserver Agreement

The mean percentage complete interobserver agreement using five size categories was 41.4% (s.e.m. 2.46) (Table 2). There was disagreement by one size category in 43.9% of readings, by two size categories in 10.9% of readings, by three categories in 3.4% and by four categories in 0.4%.

If three size categories were used, total percentage agreement increased to 58% (s.e.m. 2.52) ($p < 0.001$). There was disagreement by one category in 38.8% of readings and by two categories in 3.6% of readings.

Interobserver agreement measured by Kappa (Table 2) for five size categories in the first reading was 0.132 and in the second reading was 0.297. Using three size categories in the first reading, Kappa was 0.268 and in the second reading was 0.418. All values of Kappa were significantly different from zero. Using three scan categories significantly increased the value of Kappa in the second reading ($p < 0.05$), but the increase in Kappa in the first reading did not reach statistical significance ($0.1 < p > 0.05$).

### Spatial Resolution of $^{81m}$Kr Compared to $^{99m}$Tc

At the camera face (0 cm), the FWHM for $^{99m}$Tc and $^{81m}$Kr was identical at 5.2 mm. With the attenuation from 10 cm water, the FWHM for $^{99m}$Tc increased to 8.1 mm, and for $^{81m}$Kr increased to 8.3 mm.

## DISCUSSION

In this study, segmental defects in ventilation were produced using an isotope ($^{81m}$Kr) that gives a similar resolution to $^{99m}$Tc used in perfusion scanning (*8*). This similar resolution was confirmed in the present study by the very comparable values for the FWHM measured with our gamma camera and collimator. In the lung, the territory perfused by the artery supplying a bronchopul-

monary segment is the same as that ventilated via its segmental bronchus (*9*). Therefore, defects produced in this way are directly comparable with perfusion defects that occur in pulmonary embolism.

These results demonstrate that the size of known segmental defects on a lung scan may be greatly underestimated by experienced observers. Only 44% of segmental defects were interpreted as being greater than 75% of a segment. According to the criteria introduced by Biello (*10*) and modified in the recent PIOPED study, these defects would be classified as large. By the same criteria, 40% of known segmental defects produced in the present study would be classified as moderate sized defects (25%–75% of a segment) and 17% would be classified as small subsegmental defects (<25% of a segment). In the set of criteria proposed by McNeil (*2*), any defect or defects less than a full segment carries a low probability for pulmonary embolism. In the assessment of perfusion defects, it has been emphasized that the location, presence or absence of corresponding roentgenographic abnormality and ventilation-perfusion mismatch must be taken into account (*10*). However, the estimated size of perfusion defects clearly influences the probability category for pulmonary embolism. The present study demonstrates that largely empirical classifications based on segmental size are unreliable because of the false assumption that actual segmental defects *appear* segmental on the lung scan.

Studies have shown that the probability of angiographic evidence of embolism is around 90% for multiple segmental or larger mismatched defects in perfusion (*4,11*). However, these high probability lung scans diagnose only a minority of patients with pulmonary embolism (*3,4*). It has also been shown (*3,4*) that a normal or near normal lung scan effectively excludes the diagnosis. It has been suggested that scan categories between these two extremes be considered as "nondiagnostic" because they do not identify the 30%–40% of patients in this group who have pulmonary embolism (*12*). A proportion of "nondiagnostic" scans will be in patients with co-existent cardio-pulmonary disease which makes interpretation difficult. However, a further explanation may be underestimation of defect size caused by inadequate visualization of some defects in standard planar views (*5*). In addition, it is likely that observers misinterpret segmental size from a lack of appreciation of both the scintigraphic appearances of actual segmental defects and the range of size differences between individual segments in the lung which may be threefold (*13*). Therefore, defects that are actually segmental in nature may be underestimated from their lung scan appearances and be misclassified as subsegmental. This hypothesis is supported by the findings of this study.

The most frequently underestimated defects in this study involved the anterior and lateral basal segments of both lower lobes, the medial segment of the right middle lobe and the posterior and apical segments of the right upper lobe. We have previously reported that a defect involving the medial basal segment of the right lower lobe is undetectable on any view (*5*).

In the present study, the image containing a defect in the medial basal segment of the right lower lobe was not presented to the observers because the defect was not visible in the negative images. If this image had been included, it would presumably have increased the frequency with which defects were classified as <25% of a segment.

Intraobserver agreement was moderate (Table 2). Predictably, decreasing the number of size categories from five to three increased the total percentage agreement from 56.8% to 75%. However, such an analysis does not take into account the increase in agreement expected by chance. Kappa was consistently higher with three size categories than with five for each observer, but in each case the increase did not reach statistical significance. Our study has specifically addressed the ability of experienced observers to reliably interpret the size of defects of known anatomical location and size. Previous studies of lung scan observer agreement (*14,15*) have usually addressed the question of presence or absence of disease which makes comparisons with our findings difficult. However, Hoey et al. (*16*), using lung scans performed for suspected pulmonary embolism, tested variation between two observers in the classification of abnormalities as nonsegmental, subsegmental, segmental or lobar (*16*). For defects judged to be segmental or smaller, the Kappa within observers varied widely from 0.008 to 0.604, which is comparable to values obtained in the present study.

Interobserver agreement in the present study was consistently poor (Table 2). Decreasing the number of size categories from five to three significantly increased the total percentage agreement. Kappa also increased but the increase was statistically significant in the second reading only. In the previously mentioned study (*16*), interobserver agreement for defects judged to be segmental or less gave a value for Kappa ranging between 0.202 and 0.510.

Koran (*17*) has reviewed the reliability of clinical judgments including radiographic interpretations. Two observers interpreting pyelograms for the presence or absence of pyelonephritis managed an overall agreement of 80% but with a Kappa of only 0.425 (*18*). Agreement on presence or absence of osteoarthritis in individual joints on hand x-rays was 67% with a Kappa of 0.348 (*19*). Agreement on presence or absence of pneumoconiosis on a chest radiograph was found to be 78% with a Kappa of 0.471 (*20*). Direct comparisons of Kappa from different studies may be misleading because of the dependence of Kappa on the prevalence of the attribute being measured (*21*). However, the findings of the present study suggest that the subjective estimation of segmental size on a radioisotope lung scan has only poor reproducibility.

In conclusion, it is suggested that empirical scoring criteria based on segmental size be viewed with caution because of the inaccuracy of predicting the segmental or subsegmental nature of a defect on the lung scan and because of low observer agreement. Taking into account anatomical data (*5*), it is suggested that any defect that is pleurally based and triangular or concave in shape and in the anatomical distribution of a lung segment should be considered segmental in nature. It should also be remembered that some segmental defects (mentioned above) are particularly prone to underestimation and that a defect involving the medial basal segment of the right lower lobe may be undetectable in any view.

## ACKNOWLEDGMENTS

The authors would like to thank Peter Ryan for his excellent technical assistance, Dr. R.F. Jewkes of Charing Cross Hospital and Dr. J.B. Bomanji, Dr. K.E. Britton and Dr. Q. Siraj of St. Bartholomew's Hospital, London for their assistance and advice. This work was supported by a North West Thames Regional Health Authority Grant.

## REFERENCES

1. Biello DR, Mattar AG, McKnight RC, Siegel BA. Ventilation-perfusion studies in suspected pulmonary embolism. *AJR* 1979;133:1033–1037.
2. McNeil BJ. Ventilation-perfusion studies and the diagnosis of pulmonary embolism: concise communication. *J Nucl Med* 1980;21:319–323.
3. The PIOPED Investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism. *JAMA* 1990;263:2753–2759.
4. Hull RD, Hirsh J, Carter CJ, et al. Diagnostic value of ventilation-perfusion lung scanning in patients with suspected pulmonary embolism. *Chest* 1985;88:819–828.
5. Morrell NW, Roberts CM, Jones BE, Nijran KS, Biggs T, Seed WA. The anatomy of radioisotope lung scanning. *J Nucl Med* 1992;33:676–683.
6. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Measures* 1960;20:37–46.
7. Siegel S, Castellan NJ. *Nonparametric statistics for the behavioural sciences*, 2nd edition, New York: McGraw-Hill Book Co.; 1988:284–291.
8. Fazio F, Jones T. Assessment of regional ventilation by continuous inhalation of radioactive krypton-81m. *Br Med J* 1975;3:673–676.
9. Appleton AB. Segments and blood-vessels of the lungs. *Lancet* 1944;2: 592–594.
10. Biello DR. Radiological (scintigraphic) evaluation of patients with suspected pulmonary embolism. *JAMA* 1987;257:3257–3259.
11. McNeil BJ. A diagnostic strategy using ventilation-perfusion studies in patients suspect for pulmonary embolism. *J Nucl Med* 1976;17:613–619.
12. Hull RD, Raskob GE. Low-probability lung scan findings: a need for change. *Ann Intern Med* 1991;114:142–143.
13. Horsfield K. M.D. thesis. University of Birmingham, 1967.
14. Bateman NT, Coakley AJ, Croft DN, Lyall JRW. Ventilation-perfusion lung scans for pulmonary emboli: accuracy of reporting. *Eur J Nucl Med* 1977;2:201–203.
15. Sullivan DC, Coleman RE, Mills SR, Ravin CE, Hedlund LW. Lung scan interpretation: effect of different observers and different criteria. *Radiology* 1983;149:803–807.
16. Hoey JR, Farrer PA, Rosenthall LJ, Spengler RF. Interobserver and intraobserver variability in lung scan reading in suspected pulmonary embolism. *Clin Nucl Med* 1980;5:508–513.
17. Koran LM. The reliability of clinical methods, data and judgments. *N Engl J Med* 1975;293:642–646, 695–701.
18. Norden C, Phillips E, Levy P, et al. Variation in interpretation of intravenous pyelograms. *Am J Epidemiol* 1970;91:155–160.
19. Wright EC, Acheson RM. New Haven survey of joint diseases. XI. Observer variability in the assessment of x-rays for osteoarthritis of the hands. *Am J Epidemiol* 1970;91:378–392.
20. Felson B, Morgan WKC, Bristol LJ, et al. Observations on the results of multiple readings of chest films in coal miner's pneumoconiosis. *Radiology* 1973;109:19–23.
21. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *Br Med J* 1992;304:1491–1494.