Enhanced Accuracy and Reproducibility in Reporting of Lung Scintigrams by a Segmental Reference Chart

John S. Magnussen, Peter Chicco, Amanda W. Palmer, Douglas W. Mackey, Michael Magee, I. Provan C. Murray, George Bautovich, Kevin Allman, Geoffrey Storey and Hans Van der Wall

Department of Nuclear Medicine, Concord Hospital; Graduate School of Biomedical Engineering, University of New South Wales; Department of Radiology, St. George Hospital; Nuclear Diagnostics; and Department of Nuclear Medicine, Royal Prince Alfred Hospital, Sydney, New South Wales, Australia

The diagnostic probability of pulmonary embolic disease is based on the recognition of unmatched segmental perfusion defects. Although interobserver and intraobserver reproducibility have been studied, accuracy has been an elusive goal due to the lack of a gold standard. We investigated the accuracy and reproducibility of reporting in a virtual scintigraphic model of the lungs, with and without the use of a lung segmental reference chart. **Methods:** A Monte Carlo package was used to model lung scintigraphy from a digital phantom of the human lungs. An ideal lung segmental reference chart was created from the phantom. Five experienced nuclear medicine physicians reported a set of all possible defects involving 100% of a segment, without and with the chart. A further set of defects involving 45%–55% of a segment in the lower lobes was investigated using the chart. **Results:** There was a significant improvement in accuracy (from 48% to 72%) and intraobserver agreement (from 61% to 77%) with the chart. The accuracy of reporting defects in the upper and middle lobes was consistently better than that in the lower lobes. There was no significant difference between the accuracy of reporting large defects and that of reporting moderate defects in the lower lobes. **Conclusion:** The lung segmental reference chart significantly improves both the accuracy and reproducibility of reporting lung scintigrams; however, although reporting in the lung bases is improved, absolute accuracy is substantially less than that in the upper and middle lobes. This emphasizes the need for caution because the lung bases are the most common site of embolic disease.

Key Words: Monte Carlo simulation; lung scintigraphy; lung segmental reference chart

J Nucl Med 1998; 39:1095–1099

Pulmonary embolism is a disease that is most often investigated by ventilation and perfusion scintigraphy. Determination of the probability of thromboembolic disease in virtually all classification systems of pulmonary scintigraphy is based on the recognition of unmatched segmental and subsegmental perfusion defects (*1,2*). The prospective validation of the importance of the recognition of segmental and subsegmental perfusion defects was achieved with the Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) trial (*2*). The modified PIOPED criteria (*3,4*) rely on a robust interpretation of scan findings, which presupposes both accuracy and reproducibility of the results.

Although there is a body of literature that assesses interobserver and intraobserver reproducibility (*5,6*), there is little that investigates the accuracy of scan interpretation. Accuracy of lung scan reporting is a difficult goal to achieve in vivo because there is no real gold standard available other than angiography, which has its limitations (*7*). Consensual agreement between observers has been used as "truth" in an alternative methodology (*5,6*). This method implies that observers also can be reproducibly inaccurate in a proportion of cases, i.e., they can agree on an incorrect result.

A novel solution to the problem is to use a virtual scintigraphic model of the segmental anatomy of the lungs that allows precise determination of defect size and location. Because these two variables are controlled, an absolute calculation of accuracy can be achieved. Such a model provides an ideal basis for investigating accuracy and reproducibility, as well as the impact of a lung segmental reference chart on these variables.

MATERIALS AND METHODS

The phantom used in the construction of the model for the investigation of the perception of defect size was based on CT data (*8*) from a supine man who was 178 cm tall and weighed 70 kg; he was chosen for his similarity to the dosimetry standard mathematical phantom (*9*).

Lung tissue in the phantom was segmented in several phases. An experienced radiologist, using patient CT data showing the interlobar fissures, marked the lobes as a guide to the segments. The segmental bronchi of human cadaveric lungs were injected with color-coded dyes. These were finely sectioned and digitized (*10*). The lobar and segmental boundaries were transferred to the phantom dataset with reference to anatomical texts (*11–13*). Subsegmental boundaries were added using the descriptions from anatomical studies (*14*).

The Monte Carlo simulation package used for this work was the Photon History Generator (*15,16*), which models the emission, scatter and attenuation of photons in a heterogeneous phantom, followed by the photons' subsequent collimation and detection.

Simulations were performed for a 23.6-mm-thick parallel-hole collimator, using a 32.5-cm radius of rotation. The isotope modeled was ^{99m}Tc, collected with a symmetric 20% energy window centered around 140 keV into a 256 × 256 matrix, resulting in counts of 600,000–700,000 per view when no defects were present. The views collected were anterior, posterior, both laterals, anterior obliques and posterior obliques.

A lung segmental reference chart was created (Fig. 1), consisting of a diagram outlining the segmental boundaries for the eight standard views of the anatomical phantom used in this study. A clear overlay, which could be removed, was used to identify the segments and views during reporting.

The study consisted of two phases, the first of which examined defects involving 100% of each segment, whereas the second

Received Mar. 12, 1997; revision accepted Aug. 22, 1997.

For correspondence or reprints contact: John S. Magnussen, MB, BS, FRSM, Department of Nuclear Medicine, Concord Hospital, Hospital Road, Concord, New South Wales, 2139, Australia.

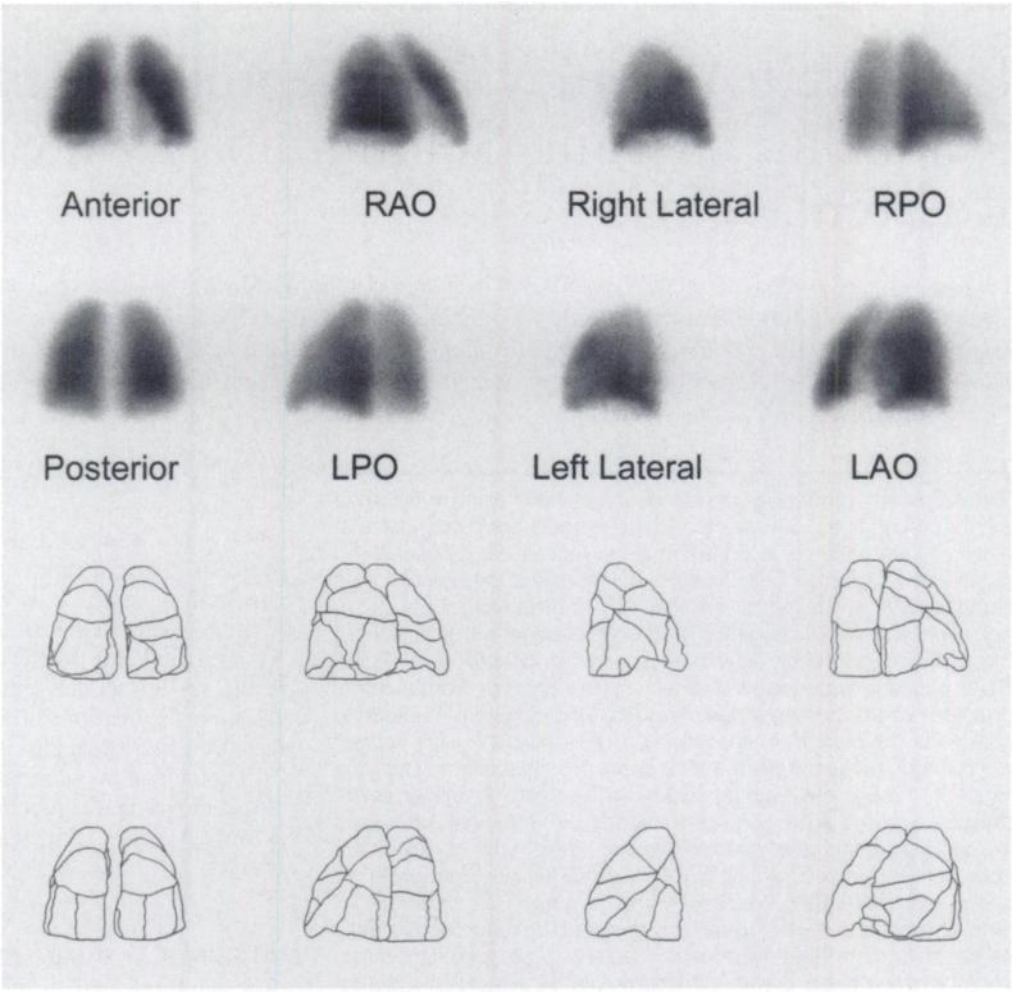

FIGURE 1. Eight views of the simulated study (upper two rows), with the corresponding outlines of the segmented phantom (lower two rows), as used in this study. The segmental labels, which were present on the clear overlay for the study, are removed for clarity. RAO = right anterior oblique; RPO = right posterior oblique; LAO = left anterior oblique; LPO = left posterior oblique.

examined defects involving 45%–55% of a segment in the lower lobes.

In the first phase, a series of 18 studies was performed, each of which contained a single defect involving 100% of a segment. Eighteen film images showing all eight standard views were presented to five experienced nuclear medicine physicians, in two stages: without and, 2 mo later, with the use of a lung segmental reference chart.

In the second phase, all anatomically possible combinations of subsegmental defects involving 45%–55% of a segment in the lower lobes, which formed a single, contiguous defect, were simulated, for a total of 18 studies. Eighteen film images showing all eight standard views were presented to five experienced nuclear medicine physicians for reporting with a lung segmental reference chart.

The observers were informed that there was only one defect present per study, which varied in size between studies. They were then asked to categorize each defect into one of the following three categories: <25%, 25%–75% or >75% of a segment. To assess intraobserver reproducibility, the same studies were represented to the clinicians 1 mo after the initial reporting session for each phase, in a different order and under the same reporting conditions.

The images were considered suitable for diagnostic purposes by all the clinical observers. Clinicians were not given any information regarding the true size of defects until the completion of both phases of the study.

Intraobserver agreement was assessed using two methods. The first was descriptive, showing the percentage of complete agreement between the first and second readings of each defect for all categories. The second used a weighted kappa statistic (*17,18*), which is a statistical measure of agreement, adjusted for the agreement that would have occurred by chance. The weighted kappa statistic is used when analyzing a categorical diagnosis, containing more than two categories, which are ordered, as is the case with the criteria for defect size used here. The kappa statistic takes a value between 1 and −1, where a value of 1 represents complete agreement, 0 represents only chance agreement and −1 shows complete disagreement.

RESULTS

The accuracy and intraobserver agreement of the reporting for defects involving 100% of a segment are shown in Table 1. The accuracy of each observer, without and with a lung segmental reference chart, is shown as the number of defects reported as 75%–100% of a segment. For four of the five observers, the accuracy improved significantly with the chart ($p < 0.05$).

TABLE 1
Accuracy and Intraobserver Agreement With 100% Defects

	Accuracy (%)			Intraobserver agreement (%)		Weighted kappa	
Observer	Without chart	With chart	p value	Without chart	With chart	Without chart	With chart
1	67	72	0.31	67	89	0.49	0.78
2	42	67	<0.01	67	78	0.60	0.64
3	44	67	<0.05	50	56	0.34	0.46
4	53	78	<0.01	67	83	0.49	0.59
5	36	78	<0.01	56	81	0.31	0.63

TABLE 2
Comparison of Accuracy Without and With a Chart for the Upper Middle and Lower Lobe Segments With 100% Defects

	Upper and middle lobes (%)			Lower lobes (%)		
Observer	Without chart	With chart	p value	Without chart	With chart	p value
1	89	100	0.157	44	44	1
2	61	89	0.025	22	44	0.045
3	56	78	0.248	33	56	0.045
4	78	100	0.014	28	56	0.025
5	44	100	0.003	28	56	0.025

TABLE 3
Comparison of Accuracy Between 100% and 50% Defects in the Lower Lobes With a Chart

	Accuracy (%)		
Observer	100% defect	50% defect	p value
1	44	69	0.06
2	44	27	0.19
3	56	61	0.71
4	56	66	0.45
5	56	69	0.32

Intraobserver agreement is shown using the percentage of complete agreement and the weighted kappa measurement.

Table 2 shows a comparison of the accuracy of reporting for 100% defects when the upper and middle lobes are considered separately from the lower lobes. For the upper and middle lobes, there was a significant difference for three of the five observers when reporting with a chart ($p < 0.05$), whereas in the lower lobes, there was a significant difference for four of the five observers ($p < 0.05$).

A histogram of the defects involving 45%–55% of a segment in the lower lobes is shown in Figure 2, which combines the results of the first and second readings. This represents two reports from each of five observers, giving a total of 180 defects, all of which were 45%–55% of a segment in size. Of these 180 defects, 26% were reported as <25% of a segment, 59% as 25%–75% and 15% as >75% of a segment.

A comparison of the accuracy of reporting defects involving 100% and 45%–55% of a segment in the lower lobes is shown in Table 3. The difference in accuracy of reporting was not significant for any of the observers ($p > 0.05$).

Examples of segmental defects, for which the use of the chart did not significantly alter the observers' accuracy and reproducibility, are shown in Figures 3 and 5. Figures 4 and 6 show examples in which the chart improved the observers' accuracy and reproducibility.

DISCUSSION

This virtual scintigraphic model of the lung allows the production of defects that can be precisely controlled in terms of size and placement. This allowed the production of 36 scans, each of which contained a unique defect, shown in eight standard views. Half the defects were 100% of a segment, and the remainder were between 45% and 55% of a segment. Most importantly, the lung segmental reference chart (Fig. 1) used in reporting these defects was ideal because it was derived from the original segmental model in which counts were simulated to produce the studies. Accordingly, these qualities created nearly ideal conditions for testing the observers, perceptions of the size of the defects.

The results of the perception of defects sizes for 100% of a segment and 45%–55% of a segment (Fig. 2) confirm that there is an underestimation of size in over a quarter of defects, as noted by Morrell et al. (*19*). It most likely represents scatter into the defects from surrounding unaffected segments.

Accuracy in determination of defect size was variable between the five observers without the lung segmental reference chart, ranging from 36% to 67%. A significant improvement in accuracy occurred in four of the five observers, and there was an improvement in intraobserver reproducibility for all observers with the chart.

Assimilation and recall of the information about the segmental anatomy of the lungs is obviously a complex process for the observer. It requires integration of planar information to produce a three-dimensional construct that is either memorized or recreated with each reading of a lung scintigram. It is conceivable that, if this process is not enacted frequently, there will be some loss of information and, therefore, accuracy, which can be compensated for by the use of the lung segmental reference chart. It also is apparent that a minority of observers could perform well without the use of a chart and apparently pay no

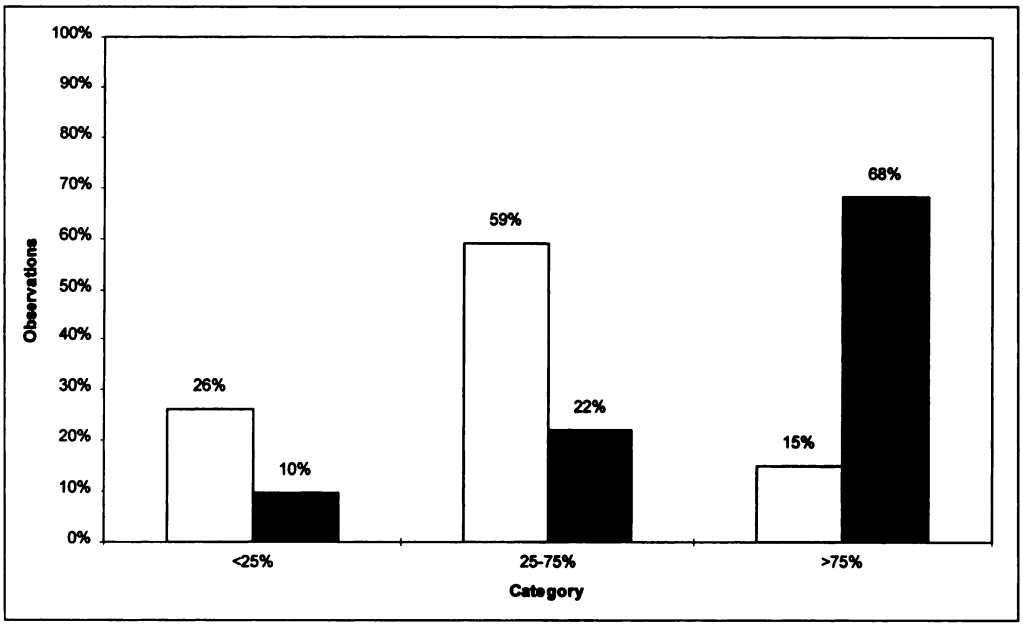

FIGURE 2. Histogram showing the frequency (%) with which the observers classified the size of the defects. ■ = segmental defects, from all lobes, that should have been classified into the >75% category. □ = subsegmental defects, from the lower lobes only, that should have been classified into the 25%–75% category.

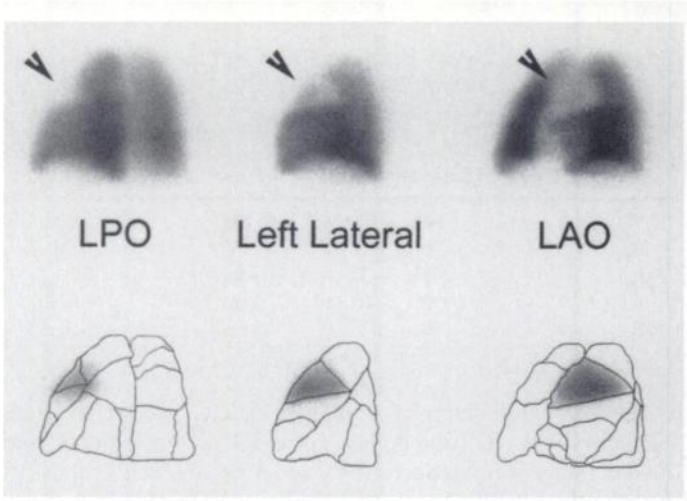

FIGURE 3. (Top) A defect in the anterior segment of the left upper lobe, for which the use of the chart did not significantly alter the observers' performance. The left posterior oblique (LPO), left lateral and left anterior oblique (LAO) views are shown. (Bottom) Segmental outlines are superimposed on the positive images.

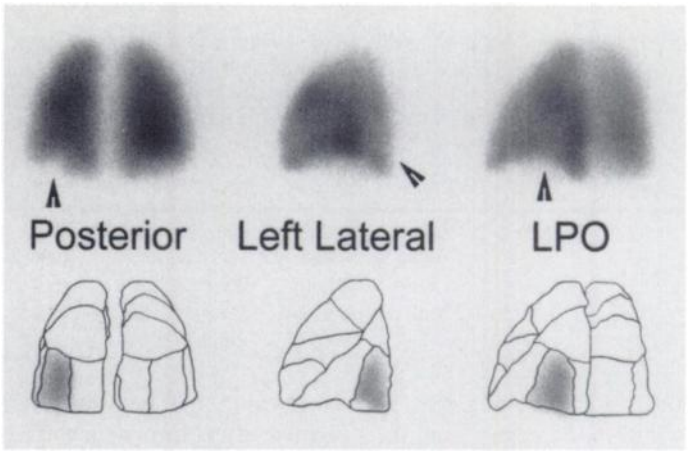

FIGURE 5. (Top) A defect in the lateral basal segment in the left lung, for which the use of the chart did not significantly alter the observers' performance. The posterior, left lateral and left posterior oblique (LPO) views are shown. (Bottom) Segmental outlines are superimposed on the positive images.

heed to the chart when it is available, as was the case with Observer 1, shown in Table 1. This would account for the failure to improve accuracy with the chart. However, other observers, whose accuracies improved with the chart, have the capability of producing a higher level of accuracy than this minority of observers.

Intraobserver agreement about defect size is uniformly improved with the chart, eliminating another source of variability in reporting. As Table 1 illustrates, the weighted kappa values improved for all five observers when the chart was used. Intraobserver agreement is reported as poor if the kappa value is less than 0.40, fair if the value is between 0.40 and 0.75 and excellent if it is above 0.75 (*20*). Observers' kappa values increased from poor to fair in two cases and from fair to excellent in one case. A chart has the advantage of standardizing reporting in terms of where segmental boundaries are perceived, which is apparently quite variable if memory is relied on. Other workers also have demonstrated this effect, which, however, presupposes the accuracy of the chart. Lensing et al. (*5*) showed that a chart could eliminate both interobserver and intraobserver variability but did not address the accuracy of the chart itself.

A different pattern of accuracy emerges when the 100% segmental defects are classified by site into lower lobes or upper and middle lobes combined. Table 2 shows that even without a chart, observers achieve a high, albeit variable, accuracy in the upper and middle lobes that can be significantly improved with the chart, reaching 100% accuracy for three of five observers. Accuracy in the bases is, however, low without the chart, with the average accuracy of the five observers being 31%. Even with the chart, accuracy improves to an average of 51%, there being a statistically significant improvement in four of five observers. This is a crucial observation, given that emboli are most common in the bases of the lungs (*21*). Arrangement of the segments is more complex in the bases of the lungs, particularly on the right side. An extra segment, the medial basal segment, is found on the right side in a corresponding position to where the heart indents the anteromedial border of the left lung. It produces a double row of tightly packed segments, with the peripheral segments grouped in an arc around the medial basal segment. Thus, any defects in the peripheral arc will be obscured to some degree by shine-through from the centrally placed medial basal segment, in addition to scatter from the surrounding unaffected segments. This is not the case in the base of the left lung, where a single row of peripheral segments is grouped around the heart. Scatter from

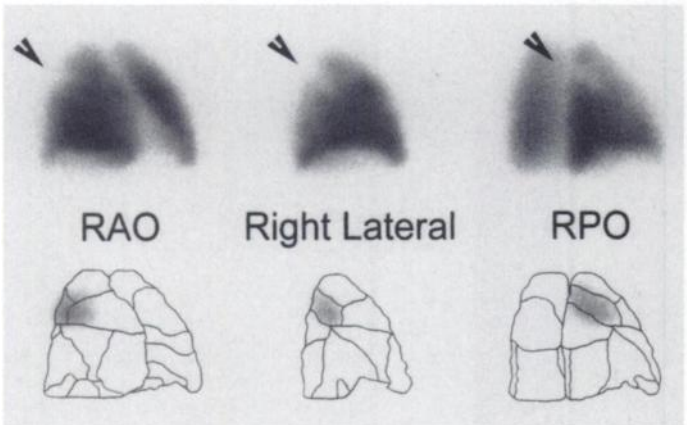

FIGURE 4. (Top) A defect in the posterior segment of the right upper lobe, for which the use of the chart significantly improved the observers' accuracy and reproducibility. The right anterior oblique (RAO), right lateral and right posterior oblique (RPO) views are shown. (Bottom) Segmental outlines are superimposed on the positive images.

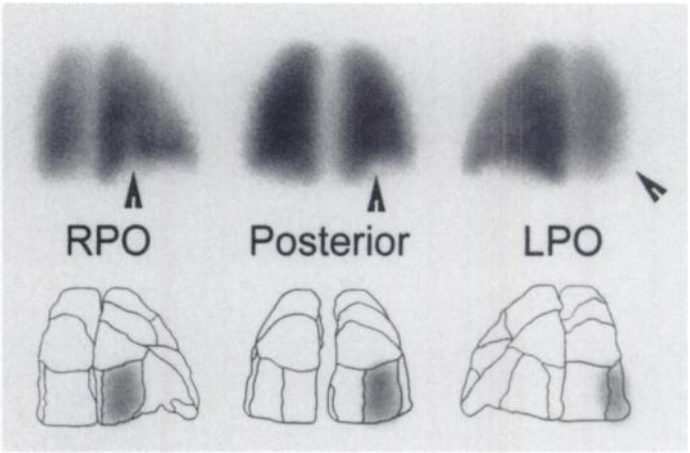

FIGURE 6. (Top) A defect in the lateral basal segment in the right lung, for which the use of the chart significantly improved the observers' accuracy and reproducibility. The right posterior oblique (LPO), posterior and left posterior oblique (LPO) views are shown. (Bottom) Segmental outlines are superimposed on the positive images.

the unaffected segments will be the main contributor in obscuring the defect.

The accuracy of detection of basal defects analyzed by size is shown in Table 3. Although the trend was for higher accuracy in the 50% defects, this did not reach statistical significance. It highlights the problem associated with the accurate perception and classification of all defects in the bases. This finding is a cause for some concern, given that pulmonary emboli are most common in the lung bases.

CONCLUSION

A virtual scintigraphic model of the lungs that produces high-quality images and is based on anatomical data has been generated. The model also has led to the production of an anatomically accurate lung segmental reference chart as an aid for reporting scintigrams. This chart has the capability of significantly improving the accuracy and intraobserver reproducibility in the reporting of lung scintigrams. Although it does improve overall accuracy in reporting defects in the lungs, the absolute level of accuracy in the lung bases is still only moderate. It is of particular concern in the setting of multiple, contiguous segmental involvement. Notwithstanding these deficiencies, it has the potential to standardize the reporting of lung scintigrams at a high level of accuracy and reproducibility.

ACKNOWLEDGMENTS

This work was supported by research grants from the Nuclear Medicine Research Foundation, Sydney, Australia.

REFERENCES

1. Biello DR, Mattar AG, McKnight RC, Siegel BA. Ventilation perfusion studies in suspected pulmonary embolism. *AJR Am J Roentgenol* 1979;133:1033–1037.
2. PIOPED Investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism: results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). *JAMA* 1990;263:2753–2759.
3. Gottschalk A, Juni JE, Sostman HD, et al. Ventilation-perfusion scintigraphy in the PIOPED study. Part I. Data collection and tabulation. *J Nucl Med* 1993;34:1109–1118.
4. Gottschalk A, Sostman HD, Coleman RE, et al. Ventilation-perfusion scintigraphy in the PIOPED study. Part II. Evaluation of the scintigraphic criteria and interpretations. *J Nucl Med* 1993;34:1119–1126.
5. Lensing AWA, van Beek EJ, Demers C, et al. Ventilation-perfusion lung scanning and the diagnosis of pulmonary embolism: improvement of observer agreement by the use of a lung segment reference chart. *Thromb Haemostasis* 1992;68:245–249.
6. Vanbeek EJR, Tiel van Buul MMC, Hoefnagel CA, Jagt HHT, Vanroyen EA. Reporting of perfusion/ventilation lung scintigraphy using an anatomical lung segment chart: a prospective study. *Nucl Med Commun* 1994;15:746–751.
7. Greenspan RH. Pulmonary angiography and the diagnosis of pulmonary embolism. *Prog Cardiovasc Dis* 1994;37:93–105.
8. Zubal IG, Harrell CR, Smith EO, Rattner Z, Gindi G, Hoffer PB. Computerized three-dimensional segmented human anatomy. *Med Phys* 1994;21:299–302.
9. NM/MIRD. *Estimates of specific absorbed fractions for photon sources uniformly distributed in various organs of a heterogeneous phantom.* New York: Society of Nuclear Medicine; 1978; 5.
10. Magnussen J, Chicco P, Vu D, Pandey P, D'Mello A. Study of sectional segmental anatomy of the human lung using gelatin-embedded sections and CT imaging techniques [Abstract]. In: *Proceedings of the 1996 Conference of the Anatomical Society of Australia and New Zealand.* 1996; 22.
11. Netter FH. *Atlas of human anatomy.* Basel, Switzerland: Ciba-Geigy, 1989.
12. Gray H. *Gray's anatomy*, 37th ed. Edinburgh: Churchill Livingstone, 1989.
13. Agur AMR, Lee MJ. *Grant's atlas of anatomy*, 9th ed. Baltimore: Williams and Wilkins, 1991.
14. Boyden EA. *Segmental anatomy of the lungs: a study of the patterns of the segmental bronchi and related pulmonary vessels.* New York: McGraw-Hill, 1955.
15. Lewellen TK, Anson CP, Haynor DR, et al. Design of a simulation system for emission tomographs [Abstract]. *J Nucl Med* 1988;29:(suppl) 871.
16. Haynor DR, Harrison RL, Lewellen TK. The use of importance sampling techniques to improve the efficiency of photon tracking in emission tomography simulations. *Med Phys* 1991;18:990–1001.
17. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;70:213–220.
18. Carbonell AM, Landis GA, Miale A, Moser KM. Construction and testing of a thorax-lung phantom to aid in scintiphotograph interpretation. *Invest Radiol* 1969;4: 275–285.
19. Morrell NW, Nijran KS, Jones BE, Biggs T, Seed WA. The underestimation of segmental defect size in radionuclide lung scanning. *J Nucl Med* 1993;34:370–374.
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
21. Moser KM, Harsanyi P, Rius-Garriga C, Guisan M, Landis GA, Miale A. Assessment of pulmonary photoscanning and angiography in experimental pulmonary embolism. *Circulation* 1969;39:663–674.