

Macroscopic assessment of pulmonary emphysema by image analysis

P A Gevenois, J Zanen, V de Maertelaer, P De Vuyst, P Dumortier, J-C Yernault

Abstract

Aim—To propose a computerised image analysis based method for measuring, on paper mounted lung sections, the area macroscopically occupied by emphysema.

Methods—The study was based on the assessment of 69 lung sections prepared following a modified Gough-Wentworth technique. The results obtained from image analysis, point counting, and panel grading methods were compared, as was the repeatability of image analysis and panel grading.

Results—The results from image analysis and from point counting were not significantly different ($p=0.609$) and significant quadratic regressions ($r=0.96$, $p<0.001$) were found between measurements from image analysis and from panel grading, the computerised technique being shown to be the most reproducible.

Conclusions—Image analysis is a valuable and reproducible method to measure the area of lung macroscopically involved by emphysema.

(*J Clin Pathol* 1995;48:318-322)

Keywords: Pulmonary emphysema, image processing, computer assistance.

Two main methods are used to quantify macroscopically the degree of emphysema: panel grading established by Thurlbeck and co-workers,¹ and point counting initially developed by Dunnill.² The panel grading method is based on the comparison of a paper mounted sagittal section against a set of standards.¹ This method is quick but scoring is intuitive and does not provide an estimate of the extent of emphysema. Moreover, this technique does not permit combined grading of several sections from the same lung specimen, although it has been shown that an adequate assessment of emphysema cannot be made from a single lung slice.³ On the other hand, point counting is truly quantitative and can be performed on several sections obtained throughout a lung specimen.²⁻⁴ Depending on the proportion of emphysema involving the sample, Turner and Whimster have shown that a minimum number of points has to be counted in order to reach a predetermined standard error. This number may be so high that counting would have to be made several times on all the slices obtained from a lung specimen.³

We propose an image analysis based technique for measuring the surface area of lung occupied by emphysema on paper mounted lung sections. We have compared this new technique with point counting as well as with the panel grading method. The repeatability of the image analysis technique and of the panel method was also compared.

Methods

LUNG SLICES

The pathological material used in this study was made of horizontal lung slices prepared according to a modified Gough-Wentworth technique.⁶ They originated from 69 surgical specimens (62 resections for peripheral cancer and seven from transplantation) collected between October 1991 and September 1992. All the selected specimens, free of pneumonia or atelectasis, were prepared following the same procedure. The segmental bronchi were cannulated and 10% buffered formalin was poured through a perfusion set from a bottle at a height of 20 to 30 cm above the lung level for 48 hours. After fixation, each specimen was horizontally sectioned from apex to base into 1.5 cm thick slices, and one paper mounted lung section was obtained from each slice. The mean number of sections per specimen was 11, ranging from five to 23 (see table 1 for details). For the purpose of the present study, one slice per specimen was randomly selected; 62 of the selected slices involved a single lobe and seven involved two lobes.

IMAGE ANALYSIS

Each slice was submitted twice to a SAMBA (TITN-Alcatel, Grenoble, France) image analysis system. The lung sections were examined by transparency with a 0.5 inch CCD matrix video camera, the objective of which had a 25 mm focal distance. The distance between

Table 1 Number of patients and number of lung sections obtained in each type of surgical resection

	Number of patients	Number of sections (range)
Right pneumonectomy	7	8-16
Right upper lobectomy	23	8-19
Right middle lobectomy	2	5-8
Right lower lobectomy	5	8-13
Right middle and lower bilobectomy	2	8
Left pneumonectomy	11	10-23
Left upper lobectomy	14	9-17
Left lower lobectomy	5	8-14

Hôpital Erasme,
Route de Lennik 808,
1070 Brussels,
Belgium: Department
of Radiology
P A Gevenois

Department of Chest
Medicine
P De Vuyst
P Dumortier
J-C Yernault

Statistical Unit,
IRIBHN
V de Maertelaer

Service of Histology,
Medical School,
Université de Mons-
Hainaut, Mons,
Belgium
J Zanen

Correspondence to:
Dr P A Gevenois.

Accepted for publication
25 August 1994

the front lens of the objective and the sample was 31 cm. Lung sections were divided into

Table 2 Results from the three methods applied on the 69 lung sections

	Point counting (%)	Image analysis (%)		
		Session No. 1	Session No. 2	Panel grading
Mean	13.5	12.9	12.6	—
Standard deviation	16.3	15.8	15.3	—
Minimum	0	0.3	0.1	0
Maximum	71.4	74.3	74.3	100

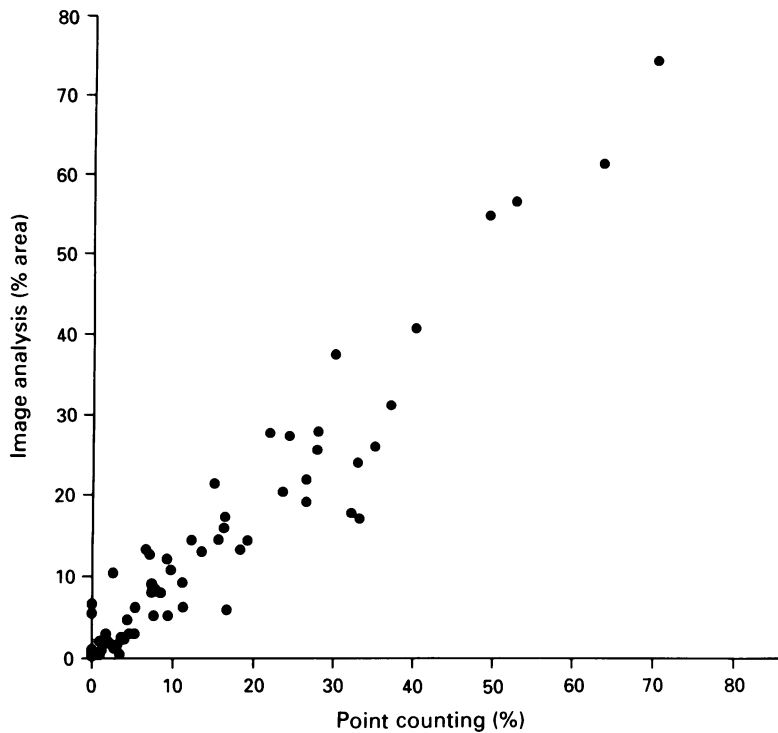


Figure 1 Scatter plot of the results from image analysis versus point counting applied to the 69 paper mounted lung sections.

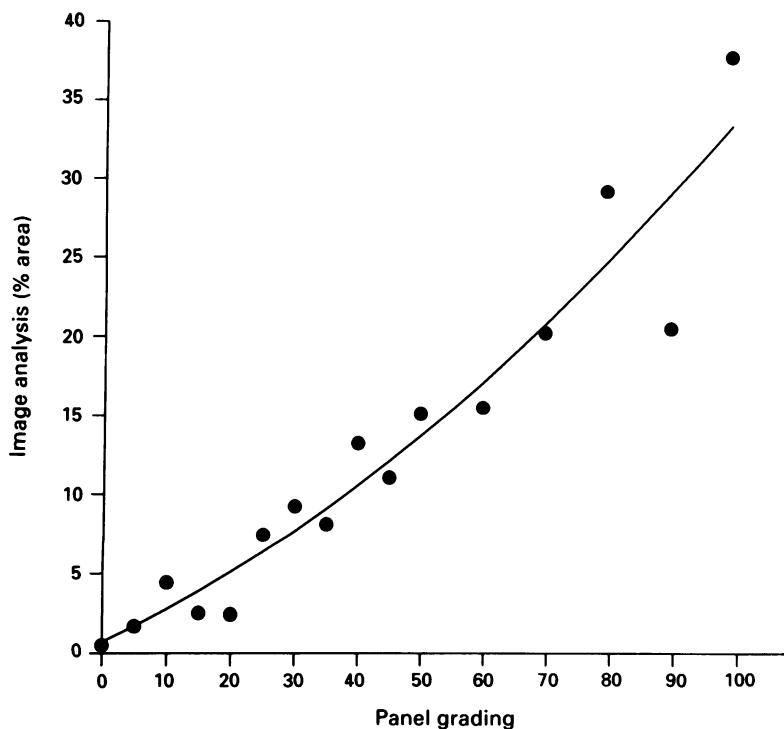


Figure 2 Quadratic regression of the results from image analysis applied to the panel versus the corresponding scores. There was a significant correlation ($r=0.96$).

fields of 7×7 cm. To assess the whole surface area, one to five fields per section had to be examined; to avoid overlapping between adjacent zones, a paper mask was used and the limits of each field were traced on the plastic sheet covering the lung section.⁶ Each image was digitised on a colour monitor (512×512 pixels) and stored in a microcomputer. The image thresholding was selected by the operator so that highest grey levels correspond to emphysematous spaces, contrasting with the lowest grey levels corresponding to the lung structures. Before calculation of the number of pixels per field corresponding to emphysema and to the total lung area, the vessels and bronchi were erased. After addition of the results obtained per field, the final result was expressed as the percentage of emphysematous area related to total lung area. A 7×7 cm square corresponded to 200 000 pixels on the monitor; one pixel was formed to be equal to 0.025 mm^2 . The total procedure was very rapid, since it requires about 15 seconds per field. The Thurlbeck's standards were themselves submitted to our image analyser. Since they could not be studied by transparency, indirect lighting was required. The operator of the image analyser was a radiologist not familiar with image analysis.

POINT COUNTING

A piece of transparency film for plain paper copiers, with the point counting grid drawn on it, was placed over each lung section. The points lay 1 cm apart and were situated at the angles of equilateral triangles with 1 cm sides.² The percentage of the lung involved by emphysema was given by the number of points superposed on emphysema multiplied by 100 and divided by the number of points on emphysema plus the number of points on non-emphysematous parenchyma. Blood vessels and bronchi were excluded from this counting. Point counting was performed by the experimenter who had done the image analysis.

PANEL GRADING

The panel grading method is based on the comparison of the paper lung section under consideration against a panel of standards. This panel consists in photographs of 16 sagittal whole lung sections arranged according to increasing severity of emphysema. A score, over the range of emphysema from none (score 0) to the most severe (score 100), corresponds to each standard.¹ In the present study, the 69 sections were submitted twice to four observers who independently compared them to the panel and gave the score of the most closely similar standard. Reader No. 1 was the laboratory assistant who performed the whole lung sections; reader No. 2 was the person who adapted the program used on the image analyser. Reader No. 3 was a pneumologist who often uses the 1980 ILO classification of radiographs for pneumoconioses.⁷ Reader No. 4 was a scientist working on an electron microscope in our research unit.

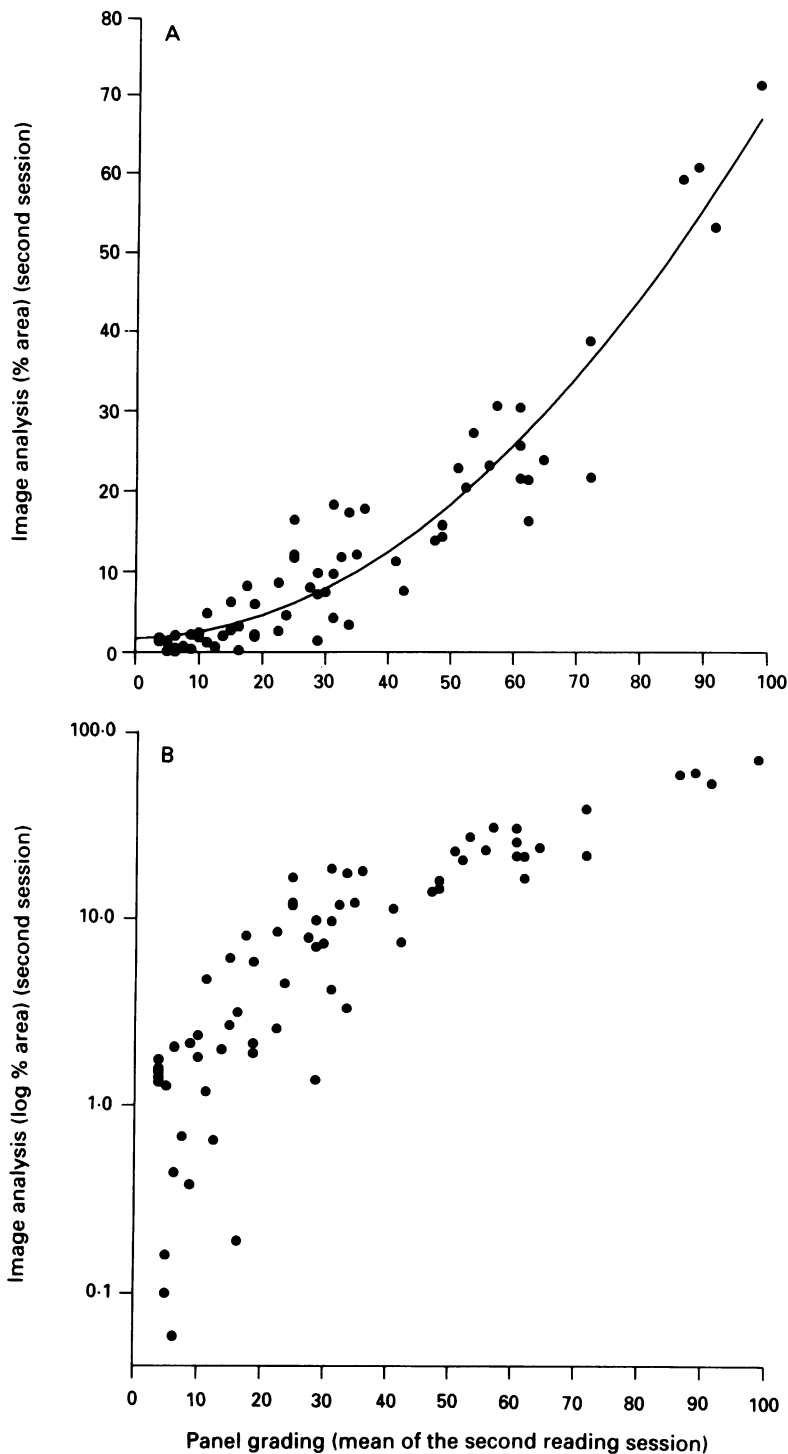


Figure 3 A: Quadratic regression of the results from the image analysis session No. 2, applied to the 69 paper mounted lung sections, versus the mean of the sessions No. 2 from the four readers. There was a significant correlation ($r=0.96$). B: Same relationship as shown in fig 3A expressed on a semilogarithmic scale.

STATISTICAL ANALYSIS

The results from image analysis and from point counting were compared in order to investigate the compatibility of these methods. A Student t test for paired samples was used at $\alpha=0.10$ level of significance. Data were previously transformed by taking their square roots in order to tend towards normality. (A logarithmic transformation was not considered because for some sections point counting yielded a zero value.) As there was no significant difference between the results of these two methods, we also checked that the power of the test was

reasonably large. (The power is the probability of detecting a difference in the mean results if such a difference really exists. A power of at least 0.70 is considered satisfactory.⁸)

We also compared the data obtained from image analysis and from panel grading. In this case, testing the equality of means is inappropriate since these two techniques express the severity of emphysema in different units; thus we investigated the correlation between the results obtained by image analysis and by panel grading (Pearson's correlation coefficient).⁹

The reproducibility of the image analysis method and the intraobserver reproducibility of panel grading were evaluated using the repeatability coefficients.⁹ The 95% confidence intervals on the mean differences between the first and second measurements were also calculated. The interobserver reproducibility of the panel grading method was evaluated by using the intrapatient correlation coefficient, expected to be equal to 1 in case of perfect correlation between the readers.¹⁰

Results

The results obtained by applying the different techniques on the 69 lung sections are summarised in table 2.

IMAGE ANALYSIS AND POINT COUNTING

The means calculated from the image analysis and from the point counting methods were not significantly different ($p=0.609$). The data obtained by both methods are plotted in fig 1. The power of the test is 0.77 for detecting a 5% difference in the results of the two methods, and 0.97 for detecting a 10% difference.

IMAGE ANALYSIS AND PANEL GRADING

When image analysis was applied to the 16 standards of the panel, the percentage area of emphysema ranged from 0.5% to 37.6% and a significant quadratic regression was found between this quantification and the scores from the panel ($r=0.96$, $p<0.001$) (fig 2).

In order to calculate the correlation between the results obtained from image analysis and from the grading panel method applied to the 69 lung sections, we averaged the results of the second session of the four readers. A highly significant quadratic regression ($r=0.96$, $p<0.001$) was found between the grading scores and the data from the image analysis (fig 3A).

REPEATABILITY

The repeatability coefficients and the 95% confidence intervals on the mean differences are given in table 3. The best performance, indicated by the lowest repeatability coefficient, was obtained with the computerised method. For the grading panel method, the best performance was obtained by reader No. 3. Nevertheless, the lower and upper limits of the confidence intervals of the mean differences

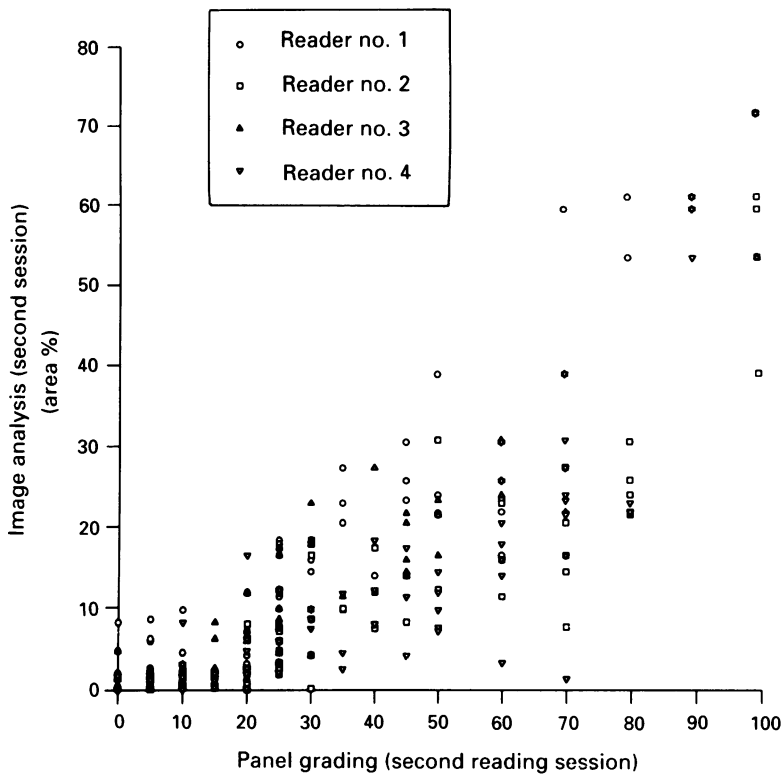


Figure 4 Comparison between the results from the image analysis session No. 2 versus the session No. 1 from each reader.

between the first and second reading sessions are both positive in reader No. 1, and both negative in reader No. 3. This indicates that the data obtained in the second reading session were significantly lower than in the first reading session for reader No. 1, and significantly higher for reader No. 3, and thus reveals a bias in two of the four readers, contrasting with the best repeatability of one of them. In addition, the intrapatient correlation coefficient assessing the interobserver reproducibility of the panel grading method was 0.81, with a 90% confidence interval ranging from 0.75 to 0.86. As this interval did not include 1, a perfect correlation is not plausible, consistent with the data obtained by the four readers and by image analysis plotted in fig 4.

Discussion

Until now, the point counting method described by Dunnill² with its adaptation proposed by Turner and Whimster³ was the only available quantitative method expressing the proportion of lung macroscopically involved by emphysema. Computerised techniques have been developed for microscopic measurements, and we propose the adaptation of such a technique for macroscopic quantification. Our study shows that the results of image analysis are similar to those of point counting. One

particular advantage of image analysis is that it yields data more rapidly and more precisely than point counting. Depending on the proportion of emphysema involving the sample, Turner and Whimster have shown that a minimum number of points has to be counted in order to reach a predetermined standard error.³ For example, the total number of points required to give a standard error of 5% ranges from 45 to 399 600 for proportions of emphysema ranging respectively from 90% to 0.1%. To obtain a high number of points, all the slices from a lung specimen might have to be counted several times. Using image analysis, 200 000 adjacent points are simultaneously counted on each 7 × 7 cm field, contrasting with multiple counting necessary with the Dunnill method.³

The lowest repeatability coefficient obtained with image analysis shows that the computerised technique is more reproducible than the panel grading method. The differences between two successive image analysis sessions remain small; they can be explained by tenuous variations in the thresholding performed in the beginning of each procedure. Using the panel method, the best reader in our experience was the pneumologist who had been familiar with the ILO classification of radiographs of pneumoconioses⁷ for several years. It is possible that the intensive use of a system based on comparisons of images is a valuable training for another system, but the reproducibility is never as high as with the image analyser. In addition, a bias was observed between two reading sessions with this particular reader as well as with reader No. 1. The comparison of the results from the four readers against the image analysis (fig 4) and the intrapatient correlation coefficient show significant interobserver discrepancies. Despite significant correlations between results from the image analysis and from panel grading, these discrepancies reveal major limitations of the grading method and suggest that analysing gathered data obtained by different observers as well as by the same observer on different occasions could be unreliable. In addition, the comparison between image analysis and panel grading scores reveals that the latter, ranging from 0 to 100, are not proportional (figs 2 and 3A). This is not related to a lack of detection of the subtle differences between the lowest grades, as shown by the plot of results on a semilogarithmic scale (fig 3B).

Since new programs, available on computed tomography (CT) scanners, provide data established in vivo on a set of scans,^{11,12} pathological quantitative data, obtained in order to validate these CT procedures, should also be based on a set of lung sections. The area of emphysema, expressed as a percentage, should be calculated on sections obtained through an entire lobe or lung and compared to quant-

Table 3 Repeatability coefficients and 95% confidence intervals on the mean differences established from two sessions

	Image analysis	Reader No. 1	Reader No. 2	Reader No. 3	Reader No. 4
Repeatability coefficients	0.898	2.118	3.006	1.292	2.582
95% confidence intervals	[-0.506; 1.286]	[4.407; 8.636]	[-4.376; 1.622]	[-4.334; -1.753]	[-4.908; 1.054]

itative CT data, also expressed as a percentage area. In addition, it has been shown that an adequate assessment of emphysema cannot be made from a single lung slice.³ Providing valid results, the image analysis method has the advantages of being quick, precise, and highly reproducible, and it permits the combination of data from several representative sections obtained throughout a lung specimen.

Supported by The Foundation Erasme.

- 1 Thurlbeck WM, Dunnill MS, Hartung W, Heard BE, Heppleston AG, Ryder RC. A comparison of three methods of measuring emphysema. *Hum Pathol* 1970;1:215-26.
- 2 Dunnill MS. Quantitative methods in the study of pulmonary pathology. *Thorax* 1962;17:320-8.
- 3 Turner P, Whimster WF. Volume of emphysema. *Thorax* 1981;36:932-7.
- 4 Anderson JA, Dunnill MS. Observations on the estimation of the quantity of emphysema in the lungs by the point-sampling method. *Thorax* 1965;20:462-6.
- 5 Medical Research Council. Quantitative assessment of chronic non-specific lung disease at necropsy. *Thorax* 1975;30:241-51.
- 6 Gevenois PA, Koob MC, Jacobovits D, De Vuyst P, Yernault JS, Struyven J. Whole lung sections for CT-pathologic correlations: modified Gough-Wentworth technique. *Invest Radiol* 1993;28:242-6.
- 7 International Labour Organisation. Guidelines for use of ILO international classification of radiographs of pneumoconioses. Geneva: International Labour Office, 1980 (Occupational Safety and Health Series No 22, revised).
- 8 Anonymous. *Solo power analysis*. Los Angeles: BMDP Statistical Software Inc, 1992.
- 9 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;i:307-10.
- 10 Snedecor GW, Cochran WG. *Statistical methods*, 7th ed. Iowa City: The Iowa State University Press, 1980:243-6.
- 11 Kalender WA, Fichte H, Bautz W, Skalej M. Semiautomatic evaluation procedures for quantitative CT of the lung. *J Comput Assist Tomogr* 1991;15:248-55.
- 12 Archer DC, Coblenz CL, de Kemp RA, Nahmias C, Norman G. Automated in vivo quantification of emphysema. *Radiology* 1993;188:835-8.