

MORTALITY IN COALWORKERS' PNEUMOCONIOSIS RELATED TO LUNG FUNCTION: A PROSPECTIVE STUDY*

BY

P. D. OLDHAM and C. E. ROSSITER

From the Pneumoconiosis Research Unit, Medical Research Council, Llandough Hospital, Penarth, Glam.

(RECEIVED FOR PUBLICATION OCTOBER 24, 1964)

In 1949, 126 coalminers, chosen to cover the whole radiological range of coalworkers' pneumoconiosis, were subjected to a battery of physiological tests of lung function, clinical investigations, and anthropometry. During the ensuing 12 years 40 of these men died. An analysis of the original measurements was carried out in an attempt to identify the particular features of the disturbance of lung function associated with pneumoconiosis which were most related to the rate of dying.

The men were grouped into seven classes—survivors, and those dying within the six two-year intervals of the 12-year follow-up period. The variation of the means of these classes, for each of the measurements made, then has six degrees of freedom, which implies that six uncorrelated measurements would suffice to describe it completely. Eighteen measurements were in fact available for each man, all of course highly intercorrelated and so in part providing redundant information. The method of analysis adopted was to find a linear combination of the 18 measurements which discriminated best between the seven classes, a second combination, uncorrelated with the first, which was the second best discriminator, and so on. Six such combinations, *canonical variates*, would then contain all the information provided by the variation of the seven classes, and the relative magnitude of the contribution made by each original test to each canonical variate would indicate the relevance of this aspect of the men's health to their survival time.

One canonical variate absorbed 75% of the variation between the seven classes. This variate was, moreover, virtually the same as the variate which best discriminated between those who died and those who survived. Its main ingredients were the amount of progressive massive fibrosis and the size of the residual volume. A second variate, absorbing a further 12% of the variation, was of uncertain significance but could be interpreted as being related to the presence or absence of pulmonary emphysema.

Discrimination between the times of survival of those who died was most markedly related to two clinical indices, the E.S.R. and the systolic blood pressure, none of the indices of respiratory performance contributing much. Those who died within two years of the initial study were well separated by the canonical variate from those who died between two and four years, and from those who survived over four years, but there was no separation of those who died during the final eight years of the study.

A complete re-analysis of these data using logarithmic transformations of each characteristic produced very few changes, which suggests that the results are not very dependent on the distributional assumptions involved.

In 1949 a study of the disturbance of lung function associated with various degrees of coalworkers' pneumoconiosis was carried out at the Medical Research Council's Pneumoconiosis Research Unit.

For this purpose a group of subjects was assembled, including men without dust exposure as well as men diagnosed radiologically as suffering all the degrees of pneumoconiosis, and subjected to a battery of tests of respiratory function. Body measurements and clinical assessments were also made. Factor analysis of the correlation matrix derived from these

*Presented at the 5th International Biometric Conference, Cambridge, 1963, under the title *Multivariate Analysis of Disturbed Lung Function in Pneumoconiosis*.

data produced, as was hoped, factors identifiable in terms of the known functions of the lung: its mechanical properties as a pump; its ability to distribute inspired gas evenly throughout its volume; and its ability to bring gas and blood into close juxtaposition so that diffusion from the one to the other is possible. In addition, of course, the size of the lung appeared as a major source of variation from one man to another. It was concluded that the major disturbance of lung function associated with radiological evidence of pneumoconiosis was in its mechanical properties (Gilson, Hugh-Jones, Oldham, and Meade, 1955).

A further battery of physiological tests was applied to the survivors of the original sample 11 years later, and the interrelations of the two sets of data will, it is hoped, enlarge our understanding of pneumoconiosis and the nature of its progress, with and without continued exposure to dust. Forty of those from the original sample who had had exposure to industrial dust (all in the coalmining industry) died during the 12 years from 1949 to 1961, at the

TABLE 1
DISTRIBUTION OF TIME OF DEATH OF SUBJECTS INITIALLY MEASURED

Date	No.
June 1949-	8
June 1951-	7
June 1953-	8
June 1955-	5
June 1957-	8
June 1959-June 1961	4
Total Survived	40
	86
Total	126

intervals given in Table 1. Eighty-six in this category survived. This paper reports the results of an attempt to discriminate between the groups of men dying earlier or later in the interval, or surviving it, in terms of their characteristics at the beginning of the study.

Eighteen characteristics were included, which fell naturally into four classes, anthropometry (in which for convenience is classed *age*), clinical assessment,

TABLE 2
THE INDICES USED IN THE ANALYSIS

Test No.	Index	Interpretation	Units
1	<i>Radiological</i> Simple pneumoconiosis	Profusion of characteristic opacities in the chest radiograph	Arbitrary
2	Progressive massive fibrosis	Extent of confluent abnormality in the chest radiograph	g. (see text)
7	Total diaphragm movement	Displacement of dome of diaphragm between full inspiration and full expiration	cm.
3	<i>Anthropometrical</i> Age		years
4	Weight		kg.
5	Height		cm.
6	Stem height		cm.
8	<i>Physiological</i> Maximum voluntary ventilation	Volume of air expired in 15 sec. of maximal voluntary hyperventilation	l./min.
9	Change of M.V.V.	Increase in 8 produced by inhaling adrenaline	l./min.
10	Standardized ventilation	Volume of air inspired during 2 minutes of standard exercise	l./min.
11	Total lung volume		l.
12	Functional residual capacity	Volume of air in lungs at resting respiratory level	l.
13	Residual volume	Volume of air in lungs after maximal expiration	l.
14	Mixing index, I_{18}	Number of breaths to attain 90% concentration of inspired helium in expired gas, as a fraction of ideal exponential	%
15	Uptake of CO	Measure of diffusion across lung membrane, as a fraction of inspired concentration	%
16	<i>Clinical</i> Erythrocyte sedimentation rate		mm./hour
17	Systolic blood pressure		mm. Hg
18	Clinical grade	Degree of shortness of breath on exertion	Arbitrary

radiological assessment, and measurements of lung function. They are listed in Table 2, together with brief explanatory matter. Full details of the tests and the manner of their use are given by Gilson *et al.* (1955). The choice of these indices was based on their known reliability, their expected relevance, and, as far as possible, their statistical independence. For example, a familiar index of lung size is the vital capacity. This is not included as such, but it is equal to the difference between the total lung volume and the residual volume. The latter is usually reported as a fraction of the total lung volume, in which form clinicians have assumed it to be more relevant to lung function; it is used here in absolute form to avoid introducing spurious correlations.

Three of the indices required coding, since they are recorded as graded characteristics on more or less arbitrary scales. The clinical grade of dyspnoea was treated as if it were a measured characteristic, using the arbitrary linear scale 1, 2, 3, 4, 5. The radiological category of simple pneumoconiosis is based not on the original readings given by Gilson *et al.* (1955) but on a re-reading of the films by 10 radiologists, using the four-point scale of the I.L.O. Classification (1953) together with category $\frac{1}{2}$ (Fletcher, 1955). Their independent readings were converted to a single measurement by a weighted averaging, after transformation by a method, due to M. E. Wise, based on the postulate of a smooth distribution of abnormality on the continuum underlying the crude classification into stages. The radiological category of progressive massive fibrosis was not used; instead the areas of the radiograph showing massive shadows were traced on to tissue paper, cut out, and weighed, the weight being used in the analysis.

The mode of analysis was a quite conventional application of discriminant functions, the canonical variates of Rao (1952), in a manner closely similar to that used by Ashton, Healy, and Lipton (1957). The basis of this method is as follows. Combinations of the 18 measurements on each man are found which have certain optimal properties in distinguishing between men dying at various intervals after the measurements were taken, or surviving the whole period of the study. These combinations are multiples of the test results added together, and the relative sizes of the multipliers for the various tests give an indication of the relative importance of each test in forecasting the outlook in terms of survival for a man. The survival times were grouped, as in Table 1, in two-year intervals, so that six dying groups ($D_1, D_2 \dots D_6$) and one surviving (A) were distinguished. It follows that a complete specification of the differences between these seven groups can be given by six such combinations of tests, no

matter how many tests there are; the seven groups have six degrees of freedom.

A geometrical analogy may help to make this clear; suppose we have two men on each of whom three measurements have been made. Each man can then be represented as a point in three-dimensional space whose position is determined by making its distances from three axes at right angles correspond to the sizes of the three measurements. A complete description of the differences between the two men would be given by their distance apart on the line joining them. If 18 measurements were made on each man, a space of 18 dimensions would be needed to represent them, but a single line could still be drawn to connect the resulting points. There is but one degree of freedom.

The six combinations can be chosen in a number of ways; the method adopted was to find the first by requiring it to maximize the variance between the seven group means as a fraction of the total variance of the 126 men. The second was then chosen to be uncorrelated with the first and maximizing the variance remaining after that accounted for by the first was removed, and so on.

Geometrically, the seven group means are represented as points in 18-dimensional space. The line is found which traverses these points in the direction in which they are most extended. A second line is found in a plane at right angles to the first, in the direction of greatest extension of the cluster in this plane, and so on.

Mathematically, the procedure is to partition the variance-covariance matrix of the 18 tests into Between and Within components corresponding to the variation of the means of the seven groups and to the variation about the means within each group. This last should not vary materially between the groups; it is difficult to investigate this fully, but past experience has suggested that a reasonable similarity of the scatter of such tests and their inter-correlations might reasonably be expected over a fair range of mean levels. The extent to which the results would be affected if this assumption were untenable was explored empirically by repeating the whole analysis after transforming the data logarithmically.

If then \mathbf{B} is the Between-group variance-covariance matrix, and \mathbf{W} the Within, the six solutions \mathbf{x} of the matrix equation

$$(\mathbf{B} - \lambda\mathbf{W})\mathbf{x} = 0$$

are linear functions of the test measurements having the properties given above. Each is uncorrelated with the others, among subjects in the same group, and has λ times as much variance between groups as within them. The solution with the largest λ is evidently of particular interest, since this is the best single linear function of the measurements for discriminating between the groups.

Certain specific comparisons between the seven group means may have greater practical interest than the overall comparison itself. In particular, one is interested in that linear function which best distinguishes between the subjects who died and those who survived, and in that which best distinguishes those who died early from those who died late. Separate solutions were therefore found for the equations corresponding to these comparisons, the first giving the linear function of the measurements which best discriminated between the living and the dead, the second and third giving those linear functions of the measurements of those dying which had the best relationship to time of death. Time of death was taken first on a linear scale and then on a logarithmic scale (in which the time of survival was contracted to allow for the likelihood that discrimination would be less for those dying later).

As is convenient, the equation

$$(B - \lambda W)x = 0$$

was in each case transformed to

$$(L^{-1}BL^{-1'} - \lambda I)y = 0$$

where L is a lower triangular matrix such that $LL' = W$, and $y = L'x$. The solutions of each equation, and all preliminary and subsequent calculations, were obtained from the Stantec Zebra computer at the Welsh College of Advanced Tech-

nology. The necessary programmes were written by the authors.

Results

Table 3 gives the linear functions obtained from the complete between-groups analysis. The first canonical variate, for example, is $13.3 \times$ amount of simple pneumoconiosis $+ 59.1 \times$ area of P.M.F. $+ \dots$. Tests with positive coefficients increase the value of the variate as they increase. Tests with negative coefficients decrease it. The coefficients of each discriminant have been standardized by expressing the corresponding variables in units of their within-groups standard deviations, so the relative magnitudes of the coefficients are directly comparable. The absolute magnitudes are arbitrary. The fraction of the total variation absorbed by each function is also given, and also the results of applying the approximate χ^2 -test suggested by Bartlett (1947). It will be seen that the first canonical variate is overwhelmingly the largest, absorbing 75% of the total variation. The second might be adjudged significant also, although the total variation of all the variates after the first could well be only random. Figure 1 shows the means of the seven groups plotted on a diagram whose axes are the values of the first and second canonical variates. Their arrangement is

TABLE 3
CANONICAL VARIATES DERIVED FROM THE SEVEN SUB-GROUPS

Test No.*	Canonical Variate No.					
	(1)	(2)	(3)	(4)	(5)	(6)
1	+13.3	+8.6	-3.2	-1.6	-5.7	-8.5
2	+59.1	+66.2	-81.2	-29.8	+37.3	+27.9
3	+27.3	+6.6	-26.8	+15.0	-35.0	+25.6
4	-13.4	+13.0	-41.4	+3.8	+115.8	+25.4
5	-2.2	-32.3	+55.4	+97.4	-51.5	-64.5
6	+7.3	+26.3	-22.9	-14.1	-14.4	+38.7
7	+16.4	+29.9	+4.0	+0.6	-0.7	-2.3
8	+5.8	+32.2	-23.0	+49.7	+32.5	+10.8
9	-0.9	+40.9	-10.5	+16.5	+11.6	-11.5
10	-5.4	-55.0	-36.8	+27.5	+15.7	-5.4
11	-13.8	-18.9	+50.3	-48.5	-132.8	+57.4
12	-38.0	-30.9	-123.5	+15.5	+190.3	-54.5
13	+65.6	+74.2	+75.2	+4.4	-85.4	+82.9
14	+24.8	-26.8	-15.0	-15.3	-12.0	+9.5
15	-17.6	+15.8	+23.8	-22.9	+24.7	+27.9
16	+38.2	-51.5	+52.2	+10.1	+26.9	+25.6
17	-28.8	+29.6	+37.8	-6.5	+25.8	-34.4
18	+10.5	+31.7	+3.5	+18.3	-0.4	-28.5
% of total variation absorbed	75.1	12.2	5.6	3.3	2.3	1.4
χ^2	133.6	35.3				
d.f.	23	21		73.8		
P	<0.001	<0.05		85		
				>0.5		

*For the identity of the tests see Table 2.

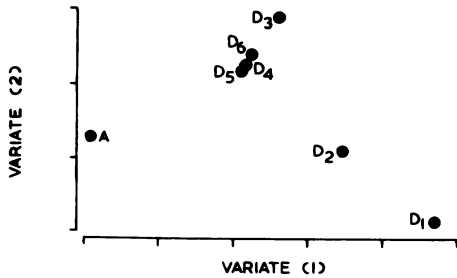


FIG. 1.—The first two canonical variates between groups (unit: one s.d.).

encouragingly logical, with the living subjects (A) at the left centre of the diagram, those who died earliest (D_1) at the right-hand bottom, and a regular progression of those dying later (D_2 — D_6) running to the left and upwards, although discrimination between those dying within eight to 12 years is evidently lacking. It should be commented at once that discrimination between *individuals* on the basis of these variates might not be very effective; the standard deviation of each is equal to one scale division on the diagram so that, for example, the 90% contour for group A just touches that for group D_2 . On the other hand, the main purpose of this study is the elucidation of the relative importance of different aspects of lung function, and the possible application of the findings in practice is of secondary interest.

The tests with the largest coefficients in both the first and second canonical variates are numbers 13 and 2, the size of the residual volume and the amount of massive fibrosis in the chest radiograph. The latter is well established as being of great significance in terms of prognosis (Cochrane, Moore, and Thomas, 1961), but the former has not previously been recognized as of outstanding importance. The third and fourth tests in order of magnitude in the first canonical variate are numbers 16 and 12, the sedimentation rate, of which high values are characteristic of the presence of active destruction of tissue by a disease process, and the functional residual capacity, which appears here with negative sign, so that small values are associated with an unfavourable prognosis. It should be remembered that in practice changes of one index may always be accompanied by changes in another; for example, the residual volume may tend to increase with the F.R.C. These four tests are probably the only important ingredients of the first canonical variate, since between them they account for 80% of the variation absorbed by it.

It appears from Fig. 1 that the separation of groups achieved by the second canonical variate is not of an easily interpretable form. While group D_1 , group D_2 , and groups D_3 , D_4 , D_5 , and D_6 are

separated, the living subjects appear midway between those dying early and those dying late. The main difference between this variate and the first is that the sedimentation rate appears with opposite sign, and the ventilation rate required for standard exercise, test 10, replaces the functional residual capacity in the first four tests in order of importance. It also has a sign which is opposite in significance to those of tests 2 and 13, since the larger the exercise ventilation the less efficient the lungs must be. Inspection of the means of the seven groups reveals that whereas for most tests there is a smooth trend from the group dying earliest to the group surviving, in those tests which reflect size of lung, or are influenced by it, there is an interruption of the trend between groups D_3 and D_4 . Figure 2 shows this for the total lung

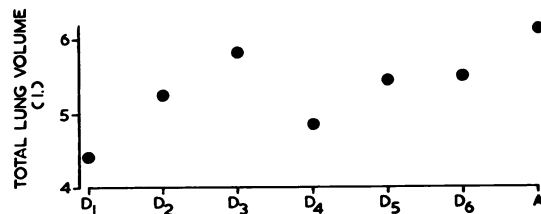


FIG. 2.—Average value of the total lung volume in the seven groups, showing an apparent discontinuity between D_3 and D_4 .

D_1 = died in 1949-51; D_2 = died in 1951-53; D_3 = died in 1953-55; D_4 = died in 1955-57; D_5 = died in 1957-59; D_6 = died in 1959-61; A = alive.

volume. This clearly may be an accident of sampling, but it would also be explicable if a large lung volume were a sign either of normality or of the presence of emphysema. Variate 2 may indicate that the groups are to a certain extent distinguishable, that is that risk of dying or survival times are distinguishable, on a basis of size of lung.

We now consider canonical variates derived by isolating single degrees of freedom from the complete between-groups comparisons; first, the contrast between the surviving subjects and those who died, irrespective of the date of death. From Fig. 1 we might expect the canonical variate best separating these groups to run in a slightly upward direction from the left of this Figure to the right, and so perhaps to be a compound of variates (1) and (2) of the between-groups comparison. Table 4 shows the function actually found; it is clear that it is in fact very similar to variate (1), but it agrees most closely with an axis making 8.7 degrees with variate (1), 81.3 degrees with variate (2).

Complementary to this contrast is the discrimination, among those who died, between different times of dying. The middle two columns of Table 4 give the first two canonical variates of the between-dead comparison, and the means of the six groups of

TABLE 4
CANONICAL VARIATES DERIVED FROM OTHER COMPARISONS

Test No.*	Alive versus Dead	Between Times of Dying		Regression on Time of Death	
		(1)	(2)	Linear	Logarithmic
1	+14.2	+4.1	-2.3	-1.1	+1.9
2	+68.0	-3.0	-6.1	+13.0	+8.9
3	+27.3	+16.3	-27.5	+14.0	+16.1
4	-10.7	-19.9	-42.4	-10.4	-14.8
5	-5.5	+17.7	+52.4	+2.9	+5.4
6	+10.2	-11.4	-21.1	-6.8	-6.8
7	+20.8	-9.6	+57.1	-5.8	-6.6
8	+11.1	-18.0	-28.8	-17.4	-18.7
9	+5.8	-30.1	-10.9	-29.1	-29.7
10	-13.1	+33.3	-37.1	+23.3	+26.1
11	-19.2	+8.3	+53.8	+10.5	+12.5
12	-39.2	-11.6	-125.9	-21.1	-20.6
13	+73.5	+0.4	+75.5	+20.9	+16.1
14	+20.1	+37.0	-14.7	+35.1	+36.7
15	-15.6	-22.1	+23.6	-11.9	-15.7
16	+29.6	+64.2	+50.5	+72.3	+68.4
17	-23.3	-42.5	+39.6	-38.7	-41.5
18	+15.9	-15.5	+2.6	-17.1	-17.1
% of total variation absorbed		60.2	16.5		
χ^2	126.0	54.2	17.6	41.9	47.3
d.f.	18	22	20	18	18
P	<0.001	<0.001	<0.7	<0.01	<0.001

*For the identity of the tests see Table 2.

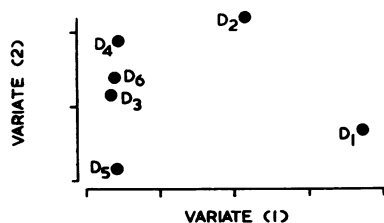


FIG. 3.—The first two canonical variates between the six groups of dead subjects (unit: one s.d.).

subjects who died are plotted in Fig. 3, according to the values of these variates. The second canonical variate is of doubtful significance, but the first separates group D_1 from group D_2 and from groups D_3 , D_4 , D_5 , and D_6 in much the same way as the first canonical variate in Figure 1. The tests with the largest coefficients are however different. The amount of massive fibrosis and the size of the residual volume appear to be of little importance in distinguishing the time of death; a high blood sedimentation rate and a low systolic blood pressure are now, from the first canonical variate, most prognostic of earlier death, while of the physiological indices the index of gas mixing I_R is first. The sign of its coefficient is, however, the reverse of what might be expected, and from inspection of the means of the six groups it appears that its contribution to the value of the first variate is negligible.

The last two columns of Table 4 give the variates obtained by expressing the differences between the

six time-of-death groups as resulting from, first, regression on time of death on a linear scale and, second, regression on time of death measured on a logarithmic scale. In the latter the intervals between the time of death of the groups dying later are counted as smaller than those between the earlier times of death. It is evident that both these variates closely resemble the first variate of the between-dead comparison, which therefore itself has isolated the time of death as a major distinguishing feature of the six groups. The logarithmic time-scale gives slightly better results than the linear.

Logarithmic Analysis.—The analysis was repeated using the logarithms of the values of the indices in Table 2 to assess the stability of the canonical variates. Unity was added to the values for indices 2 and 18 before the logarithms were taken, and instead of change of M.V.V., change of the logarithm of M.V.V. was taken for index 9. For all the other indices, logarithms of the values observed were analysed.

Five of the canonical variates obtained in this parallel analysis are shown in Table 5. There are marked similarities between these and the canonical variates in Tables 2 and 3, and the approximate χ^2 values (Bartlett, 1947) are practically the same.

The tests with the largest coefficients in the first canonical variate between the seven groups are tests 2, 13, 17, and 16, compared with 13, 2, 16, and 12 in the first analysis, the functional residual capacity being replaced in the canonical variate by the systolic

TABLE 5
LOGARITHMIC ANALYSIS

Test No.*	Canonical Variate				
	Between Groups (1)	Between Groups (2)	Alive versus Dead	Between Times of Dying	
				(1)	(2)
1	+7.6	-20.1	+3.3	+21.2	+6.6
2	+52.1	+78.5	+66.4	-34.6	-5.8
3	+19.8	+10.5	+20.9	+3.3	-38.0
4	-18.5	+34.4	-5.2	-38.1	-52.5
5	+5.1	-24.9	-2.1	+25.9	-8.4
6	+9.3	-0.9	+9.6	+5.5	-3.0
7	+6.6	+10.6	+8.7	-5.0	+13.3
8	-26.0	+111.4	-3.8	-106.4	-64.2
9	-7.2	+51.8	+3.0	-46.6	-10.5
10	+5.8	-48.4	-3.5	+42.9	-50.7
11	-12.3	-39.1	-21.7	+27.1	+75.2
12	-20.1	+15.0	-13.5	-27.4	-123.6
13	+40.8	+54.3	+49.0	-19.3	+61.8
14	+22.8	-43.0	+14.1	+48.3	+3.4
15	-25.5	+36.4	-18.1	-44.5	+25.6
16	+27.0	-4.5	+25.6	+19.3	+29.0
17	-31.0	+35.4	-23.4	-47.0	+27.0
18	-15.0	+54.3	-4.1	-53.2	+4.6
% of total variation absorbed	69.4	19.6		69.0	15.6
χ^2	123.6	49.6	119.3	62.4	17.4
d.f.	23	21	18	22	20
P	<0.001	<0.001	<0.001	<0.001	<0.7

*For the identity of the tests see Table 2.

blood pressure, low values of which are here associated with earlier death. The main difference between these five canonical variates and those from the original analysis is that the functional residual capacity does not appear to be of importance in the logarithmic analysis, whereas the maximum voluntary ventilation becomes more important.

The means for the seven groups of subjects, for the first two between-groups canonical variates in the logarithmic analysis, are plotted in Fig. 4, corresponding to Fig. 1 for the first analysis. The only change appears to be a shift in the position of D₂ towards the later dying groups. Similarly, the means for the six groups of subjects who died, for the first two between-deads canonical variates, are shown in

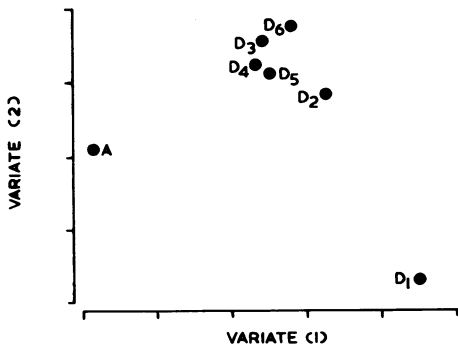


FIG. 4.—Logarithmic analysis: first two canonical variates between groups.

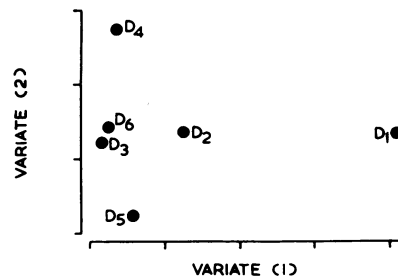


FIG. 5.—Logarithmic analysis: first two canonical variates between dead groups.

Fig. 5, which corresponds to Figure 3. Here, also, the only change is a shift in the position of group D₂.

The logarithmic analysis has shown that the canonical variates are reasonably stable and that the discrimination between groups is virtually unchanged.

Discussion

The main objective of this study has, on the whole, been satisfactorily achieved. Canonical variates have been found which, on the one hand, quite sharply differentiate between the average characteristics of those subjects who died and of those who survived, and, on the other hand, distinguish the average characteristics of those dying soon after the initial study from those surviving for six or more years.

The ingredients of the variates are not entirely what would have been expected from clinical judgment, so it is possible that new insight into the characteristics of coalworkers' pneumoconiosis may have been obtained. Whether or not this is so must await the critical appraisal of the results by physiologists and physicians.

Cotes (1960) has considered differences in the mean values of various measures of lung function between subjects dying within 12 months of assessment and those surviving this period. His indices are for the most part not included in the set we have studied, but this single factor analysis agrees with our multivariate analysis in finding that the extent of pneumoconiosis and the residual volume (in this case as a proportion of total lung capacity) are not related to early death. Subsequently Cotes (1962) reported briefly single factor analyses related to 235 men, many of whom were included in the present series, according to whether they died within four years of assessment, after four years, or survived. He found that the mean age, weight, clinical grade, maximum breathing capacity, vital capacity, and residual volume (%) all varied with survival or time of death, but did not allow for the natural correlations of these indices. Jacob Thomas (personal communication) constructed a discriminant function from these data in which the most important ingredients were: (1) area of P.M.F.; (2) clinical grade of breathlessness; and (3) residual volume % total lung capacity. Save for the appearance of clinical grade, which has little importance in the present analysis, these findings are in broad agreement with our own.

It is noteworthy that age contributes little to the canonical variates we have found. One explanation of this is that chronological age is simply a crude index of physiological changes that naturally occur with the passage of time, and that if in an individual they have not occurred, he is not more likely to die simply because he has lived for a long time. In other words, the relevant changes in his condition are to

be found in the indices of his physiological state. The original selection of subjects for this study was on a basis of age and x-ray category, so that at each age all x-ray categories were equally represented; each x-ray category contained men of all ages. Re-classification of the radiographs has to a slight extent impaired this balance; the correlation between age and the new x-ray score is -0.12 , whereas it originally was $+0.07$. The correlation is still sufficiently small for it to be easily possible for any effect of age additional to its effects on function to reveal itself by age appearing as an important ingredient of the canonical variates. That this has not occurred suggests that no such effect exists.

In one respect the present results are disappointing, in that the more specific physiological measurements appear to play less important roles in discriminating between the groups of subjects than do some less specific measures of sickness and disability. An obvious next step will be to repeat this form of analysis using only the more specific among the tests.

The material analysed in this study was obtained from co-operative work by many present and past members of the Pneumoconiosis Research Unit. We are indebted to all of them for making it available to us, and particularly to Dr. J. E. Cotes, who assembled it into an orderly and accessible form. Our colleagues are in no way to be held accountable, however, for the opinions we have expressed.

REFERENCES

- Ashton, E. H., Healy, M. J. R., and Lipton, S. (1957). *Proc. roy. Soc. B*, **146**, 552.
- Bartlett, M. S. (1947). *J. R. statist. Soc. Suppl.*, **9**, 176.
- Cochrane, A. L., Moore, F., and Thomas, J. (1961). *Tubercle*, **42**, 64.
- Cotes, J. E. (1960). *Thorax*, **15**, 244.
- (1962). *Proc. roy. Soc. Med.*, **55**, 454.
- Fletcher, C. M. (1955). *Arch. industr. Hlth*, **11**, 17.
- Gilson, J. C., Hugh-Jones, P., Oldham, P. D., and Meade, F. (1955). *Spec. Rep. med. Res. Council*, No. 290. H.M.S.O., London.
- International Labour Organisation (1953). *Proceedings of the Third International Conference of Experts on Pneumoconiosis, Sydney 1950: Report of Proceedings*, Vol. 1, p. 130. I.L.O., Geneva.
- Rao, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. Wiley, New York.