FACTOR ANALYSIS OF THE DISABILITIES OF THE ELDERLY WITH THE AID OF AN ELECTRONIC CALCULATOR

BY

JOHN MADDISON, R. C. F. SMITH, AND O. OLBRICH

Twickenham, Middlesex *Boldon, Co. Durham*

Sunderland

There have been a number of surveys of old people by doctors (Walker, 1947; Sheldon, 1948; Adams and Cheeseman, 1951; Simonds and Stewart, 1954), but no survey that we know of has attempted full medical examination of a sample in a localized community.* Early in 1952 we thought it would be a useful piece of work to determine accurately the diseases and disabilities of a sample of the men and women 60 years of age and over in a circumscribed area. We chose the villages of Tunstall and Silks-worth in Co. Durham, which together are typical of a mixed urban community with an average normal healthy population and with housing conditions and wages just a shade higher than average. The estimated population is 7,500; the acreage is 1,720; there are 2,300 houses; the number of persons per house is a fraction over 3 but under the expected average of $3\frac{1}{2}$. These conditions are a little better than usual, but to offset this there are several small groups of miners' houses with twelve houses to the acre, which are small and poorly built. For the whole area the death rate and incidence of infectious disease are about the same as for England and Wales. The workers are employed mostly in the coal mining and shipping industries and in small shops.

Arranging the Medical Examinations.—The names and addresses of all people aged 60 and over were obtained by using every source of information possible, such as the voting lists, organizations and clubs, churches, welfare officers, and medical practitioners. The completed list contained 550 names, or 7.3 per cent. of the population, which is slightly less than that expected in comparable areas of England and Wales. All the local practitioners offered to encourage the old people they met in their practices to attend for examination.

By courtesy of Dr. O. Olbrich, consulting physician in charge of geriatrics at Sunderland General Hospital, the facilities for the medical examinations in his department were placed at Dr. Smith's disposal. He arranged for about sixty people from the list to be invited each week to attend the hospital. About half of them ac-

cepted; but some feeling quite well and not in need of medical attention, some fearing something being found out, some saying they did not wish to tempt providence, some recently attended by their own doctors or already attending hospital, some being ill in bed, and some for other reasons, were reluctant or refused to attend. These Dr. Smith visited and in this way more were persuaded to come. Some private-car owners, including the Red Cross, volunteered to provide transport. The number which finally attended was 238 out of the original 550; the sample is obviously biased.

The medical examinations were very thorough, covering every system of the body and including x rays and laboratory tests. Dr. Smith did much of the clinical work himself; he was assisted by four other doctors in the geriatric department, a sister, two nurses, and a clerk. The findings for each person examined were entered on a separate form specially designed to ensure that nothing was overlooked.

Analysis of the Data.—The technique of factor analysis can be used as a tool for probing medical data to see if any factors or common influences can be detected to account for the individual differences in the people examined, and as a method of classification (Maddison, 1953).

We decided to try a factor analysis of the clinical findings of our sample. This begins with a Table showing the correlation between every pair of attributes. In our investigation some of the observations such as blood pressure took the form of measurements, but most of them were qualitative; that is the observation showed whether or not a person suffered from a disease, there being no question of degree. The first step in the analysis was the transformation of the qualitative observations into correlations. Burt (1950) has shown how this is done. For the benefit of readers who are not well acquainted with the methods of factor analysis, we have given a simplified account of the procedures used in this part of the paper. Full accounts are given by Thurstone (1947), Thomson (1950), and Cattell (1952).

FACTOR ANALYSIS

TABULATION.—We selected certain "determinables" from the clinical records of the 238 old persons examined; these were sex, age, percentage of

* Since this was written there have appeared Hobson and Pemberton (1955), and Exton-Smith (1955).

TABLE I—continued

Diseases Found		Number of Times each Disease was Found (order of importance)									
		First		2nd		3rd		4th		Total	
		Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
Diseases of Digestive System	Gingivitis; septic gums	—	—	—	—	—	1	—	—	—	1
	Pyloric stenosis	—	1	—	—	—	—	—	—	—	1
	Hernia— ^{inguinal}	3	1	—	—	—	—	—	—	3	1
	Hernia— ^{ventral}	—	—	—	1	—	—	—	—	—	1
	Diverticulitis	—	1	—	—	—	—	—	—	—	1
	Constipation	—	—	—	1	—	—	—	—	—	1
	Colon—spastic	—	1	—	—	—	—	—	—	—	1
Diseases of Genito-Urinary System	Fistula—rectum	1	—	1	—	—	—	—	—	2	—
	Cholecystitis	—	3	—	2	—	3	—	—	—	8
	Pyelitis	—	—	1	—	—	—	—	—	1	—
	Calculi of kidney and ureter; renal calculus	1	1	—	—	—	—	—	—	1	1
	Cystitis	1	—	—	—	—	1	—	—	1	1
	Caruncle of urethra	—	1	—	—	—	—	—	—	—	1
	Prostate—enlargement	3	—	5	—	3	—	1	—	12	—
Diseases of Skin	Phimosis	1	—	—	—	—	—	—	—	1	—
	Cystocoele; stress incontinence; uterine prolapse	—	3	—	5	—	2	—	—	—	10
Diseases of Skin	Dermatitis (unqualified)	—	—	—	1	—	—	—	—	—	1
Diseases of Bones and Joints	Rheumatoid arthritis	—	—	—	1	—	1	—	—	—	2
	Osteo-arthritis	3	2	—	6	1	4	1	1	5	13
	Fibrositis	—	2	—	3	—	—	—	—	—	5
	Displacement of intervertebral disk	—	—	—	1	—	—	—	—	—	1
	Bunions	—	—	—	1	—	—	—	—	—	1
	Synovitis { non-occupational	—	—	—	1	—	—	—	—	—	1
	Synovitis { occupational; bursitis	—	1	—	—	—	—	—	—	—	1
Congenital Malformations	Kyphosis; lumbar lordosis	—	—	—	1	—	1	—	—	—	2
	Flat feet	1	—	—	—	—	—	—	—	1	—
	Congenital heart disease; dextra cordia; situs inversus of liver	—	1	—	—	—	—	—	—	—	1
	Ichthyosis	—	—	—	—	—	—	1	—	1	—
Symptoms and Ill-defined Conditions	Congenital deformity—foot	—	—	—	—	1	—	—	—	—	—
	Insomnia	1	—	1	—	—	—	—	—	2	—
	Sympathetic block	—	—	—	1	—	—	—	—	—	1
Symptoms and Ill-defined Conditions	Backache—chronic	1	—	—	2	—	—	—	—	1	—
	Headache	—	1	—	—	—	—	—	1	—	4
Total		79	131	42	72	11	27	4	8	136	238

Each determinable was divided into a number of subclasses or attributes. To reduce the number of attributes to manageable size, some items with small frequencies were pooled; the 42 attributes finally chosen are set out in Table II (overleaf).

Each attribute was given a code number, those for diseases being taken from the W.H.O. International Classification (1948). The information about each person was then recorded in code form on a punch card for use with Hollerith machinery. The set of cards for the 238 persons, each with appropriate details of the 42 attributes, constituted a Table showing the presence or absence of the attributes for the given sample of persons. If the presence of an attribute is denoted by one, and its absence by nought, the Table will appear as in the sample illustration Table III (overleaf).

FREQUENCY TABLE.—The cards were sorted on a machine into groups and subgroups so as to form a Table of frequency of persons, characterized by the attributes specified, taken in pairs. The Table contained 42×42 entries; the top left-hand corner is shown in Table IV (overleaf).

Table IV can be obtained by multiplying each element in each row of Table III by the corresponding element in each column in turn of the same Table when transposed, that is with the rows arranged as columns (Table IIIA). The products thus formed are summed and entered in turn in the cells of each row of Table IV. The product-sum formed by Row j of Table III and Column k of Table IIIA is always entered in the cell at the junction of Row j and Column k of Table IV.

TABLE II
DETERMINABLES, ATTRIBUTES, AND FREQUENCIES IN PERSONS

Determinable	Attribute	Attribute Reference Number	Number of Persons	
			Men	Women
Sex	Man	1	90	—
	Woman	2	—	148
Age Group	60-64	3	19	59
	65-69	4	28	41
	70-74	5	25	28
	75-79	6	14	13
	80-84	7	4	7
Percentage of Haemoglobin	30-69	8	7	17
	70-79	9	13	47
	80-89	10	30	41
	90-99	11	20	26
	100-129	12	20	17
Blood Pressure	Systolic	130-159	13	22
		160-179	14	26
		180-199	15	16
		200-219	16	15
		220-239	17	8
		240-299	18	3
	Diastolic	60-89	19	27
		90-99	20	27
		100-109	21	13
		110-119	22	10
		120-159	23	13
Most Important Disease	First	Nil	24	12
		Ineffective, neoplasm, allergic	25	5
		Anaemia	26	—
		Mental, nervous, senses	27	11
		Heart, arteries	28	10
		Benign hypertension	29	11
		Malignant hypertension	30	9
		Respiratory	31	17
		Digestive, genito-urinary	32	9
		Skin, bone, malformation, etc.	33	6
	Second	Nil	34	46
		Infective, neoplasm, allergic	35	4
		Anaemia, mental, nervous, senses	36	9
		Heart, arteries, hypertension	37	15
	Third	Nil	39	79
		All diseases	40	11
	Fourth	Nil	41	86
		All diseases	42	4

TABLE III
ATTRIBUTES OF PERSONS

Determinable	Attribute	Persons (on punch cards)						
		A	B	C	D	E	F	G
Sex	Man	1	0	1	0	0	0	1
	Woman	0	1	0	1	1	1	0
Age Group	60-64	1	0	1	0	1	0	0
	65-69	0	1	0	1	0	0	1
Most Important Disease	First	None Found	0	0	1	0	1	0
		Respiratory	1	0	0	0	0	1
		Hypertensive	0	1	0	1	0	0
		Allergic	0	0	0	0	1	0
	Second	None Found	1	0	1	0	1	0
		Digestive	0	1	0	1	0	0
			0	0	0	0	0	1
			0	0	0	0	0	0

Thus in diagrammatic form:

Table III	Table IIIA (III transposed)	Table IV
N Persons	n Attributes	n Attributes
<div> <div>Row 1</div> <div>Row j</div> <div>Row k</div> </div>	<div> <div>Column 1</div> <div>Column j</div> <div>Column k</div> </div>	<div> <div>—</div> <div>—</div> <div>—</div> </div>

and in symbols: post-multiply Table III by its transpose Table IIIA to give Table IV.

$$\text{III} \times \text{IIIA} = \text{IV} \quad \dots \quad \dots \quad \dots \quad (i)$$

TABLE IV
FREQUENCY TABLE OF ATTRIBUTES

		Sex		Age Group					Haemoglobin					Systolic Blood Pressure					
		Men	Women	60-64	65-69	70-74	75-79	80-84	30-69	70-79	80-89	90-99	100-129	130-159	160-179	...	...	...	...
Sex	Men Women	90 —	— 148	19 59	28 41	25 28	14 13	4 7	7 17	13 47	30 41	20 26	20 17	22 15	26 36	..	..	..	..
Age	60-64	19	59	78	—	—	—	—	7	23	23	11	14	17	24	..	..	..	..
	65-69	28	41	—	69	—	—	—	5	18	19	17	10	9	19	..	..	..	..
	70-74	25	28	—	—	53	—	—	6	9	19	11	8	8	13	..	..	..	..
	75-79	14	13	—	—	—	27	—	5	6	8	5	3	2	6	..	..	..	..
	80-84	4	7	—	—	—	—	11	1	4	2	2	2	1	—	..	..	..	..
Haemo- globin	30-69	7	17	7	5	6	5	1	24	—	—	—	—	1	9	..	..	..	..
	70-79	13	47	23	18	9	6	4	—	60	—	—	—	11	14	..	..	..	..
	80-89	30	41	23	19	19	8	2	—	—	71	—	—	14	22	..	..	..	..
	90-99	20	26	11	17	11	5	2	—	—	—	46	—	4	11	..	..	..	..
	100-129	20	17	14	10	8	3	2	—	—	—	—	37	7	6	..	..	..	..
Systolic Blood Pressure	130-159	22	15	17	9	8	2	1	1	11	14	4	7	37	—	..	..	..	..
	160-179	26	36	24	19	13	6	—	9	14	22	11	6	—	62	..	..	..	..
	...	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	...	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	...	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..

TABLE V
ATTRIBUTES COMMON TO EACH PERSON

Determinates				29 Persons																	Total
				1	2	3	4	5	6	7	8	9	10	11	.	.	.	29			
Row <i>j</i> from III	Men	..	..	1	0	1	0	0	0	1	0	1	0	0	.	0	.	0	4 = <i>N_j</i> ; i.e. three men with respiratory disease and one without		
Column <i>k</i> from IIIA	Respiratory Disease			1	1	0	1	1	1	1	1	1	0	.	0	.	0	9 = <i>N_k</i> ; i.e. three respiratory diseases in men and six not men			
Cross Products	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	3 = <i>N_{jk}</i> ; i.e. three men with respiratory disease; entry in a cell of Table IV or VI.			

Suppose one Row *j* of persons from Table III and a different Column *k* of persons from Table IIIA were arranged as in Table V, we should have the result shown in Table V (above).

The third row of Table V (Cross Products) gives the attributes possessed in common by each person; these entries are also the cross products between row and column. The totals such as N_{jk} form the entries in the cells of tables like Table IV or Table VI (overleaf).

FORMING THE CORRELATIONS.—A correlation is a ratio. It provides a measure of the amount of association between two sets of observations on a series of persons or things. If the correlation is one there is perfect association; if it is nought there is no association; and if it is something in between, e.g. 0.5, there is partial association. The correlation is obtained by finding the geometric average (that is square root of product) of two slopes called regression lines.

Suppose the observations are plotted on a graph with axes *X* and *Y* (Fig. 1, overleaf). One regression line (*y* on *x*) tells us the average amount of increase up the *Y* axis for each unit length along

the *X* axis; the other regression line (*x* on *y*) tells us the average amount of increase along the *X* axis for each unit up the *Y* axis. The geometric mean of these two is the correlation.*

Each regression slope is not a simple average of the slopes of single points. It is itself a ratio: the numerator is an area—the sum of the rectangles formed by the product from each *x* and *y* value. The denominator is an area—the sum of the squares formed from the *x* values, or the sum of the squares formed from the *y* values.

The correlation may be expressed thus in words:
Sum of Cross Products

$$\frac{\text{Sum of Cross Products}}{\sqrt{(\text{Sum of Squares on } X \text{ Axis})(\text{Sum of Squares on } Y \text{ Axis})}}$$
and in symbols thus:

$$\frac{N_{jk}}{\sqrt{N_j \cdot N_k}} = \sqrt{\frac{N_{jk} \cdot N_{jk}}{N_j \cdot N_k}} \dots \dots (ii)$$

where N_{jk} is an entry in a side cell of Table IV, and N_j , N_k are the entries in the respective row and column of the diagonal.

* We are deliberately avoiding the complication in the usual computation of correlations in which the sums of cross products and squares are measured from the mean.

Suppose the totals of Row j and Column k from Table V were plotted on a graph as in Fig. 1, then:

Product of one occurrence "man" (x) $\times$ one occurrence "respiratory disease" (y)	$= 1 \times 1 = 1$		
Sum of three products (xy) for three occurrences "men with respiratory disease"	$= 3 \times 1 = 3$	$= 3$	$\dots N_{jk}$
Square for one man (x)	$= 1 \times 1 = 1$		
Sum of four man squares (x) for four men	$= 4 \times 1 = 4$	$= 4$	$\dots N_j$
Square of one respiratory disease (y)	$= 1 \times 1 = 1$		
Sum of nine respiratory disease squares (y)	$= 9 \times 1 = 9$	$= 9$	$\dots N_k$
Slope or regression (of y on x) rise of y for 1 x		$= 3/4$	$\dots N_{jk}/N_j$
Slope or regression (of x on y) increase of x for 1 y		$= 3/9$	$\dots N_{jk}/N_k$
Geometric mean = Square root of $3/4 \times 3/9 = \sqrt{9/36} = \sqrt{1/4}$		$= 1/2$	$\dots \sqrt{N_{jk} \cdot N_{jk}/N_j \cdot N_k}$

Suppose for example we make a small table (Table VI), which is like Table IV but complete. The correlations formed by applying Equation (ii) would then be as in Table VII.

In Table VII, first the rows of Table VI and then the resulting columns are divided by the square root of the number in the appropriate corresponding diagonal cell. Thus the entry 3 opposite Respiratory and under Men is divided by $\sqrt{9}$, i.e. 3, and then by $\sqrt{4}$, i.e. 2. The result is 0.500 in Table VII. This procedure converts each diagonal cell of Table VII to unity and makes the entries in

TABLE VI
FREQUENCIES

		Sex		Disease		
		Men	Women	None Found	Respiratory	Hypertensive
Sex	Men	4	0	0	3	1
	Women	0	25	4	6	15
Disease	None Found	0	4	4	0	0
	Respiratory	3	6	0	9	0
	Hypertensive	1	15	0	0	16

TABLE VII
CORRELATIONS

		Sex		Disease		
		Men	Women	None Found	Respiratory	Hypertensive
Sex	Men	1.000	.000	.000	.500	.125
	Women	.000	1.000	.400	.400	.750
Disease	None Found	.000	.400	1.000	.000	.000
	Respiratory	.500	.400	.000	1.000	.000
	Hypertensive	.125	.750	.000	.000	1.000

the side cells of relative comparability. The rows and columns of Table IV were treated in this way; the answers were found by machinery and punched on cards in the binary scale ready for the next operation.

FIG. 1.—Graph of Table V plotted on rectangular axes.

TABLE VIII
NORMALIZED VALUES

Determinates	29 Persons																	Total	Sum of Squares for Each Row
	1	2	3	4	5	6	7	8	9	10	11	.	.	.	.	.	29		
Row <i>j</i> from III Men	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0	.	0	.	0	.	0	2	1
Column <i>k</i> from IIIA Respiratory Disease	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	.	0	.	0	.	0	3	1
Cross Products	$\frac{1}{2}$	0	0	0	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0	.	0	.	0	.	0	$3/6 = \frac{1}{2}$	

THE COSINE CONVENTION.—In Table III we gave a value one to each person who possessed an attribute and the value nought when he did not. The following step is simplified if we regard the correlations as having been obtained in a different way. In Table V, Row *j* (men) contains four values each of one making a total of four. Each value one² is one, so that the sum of the squares is 4 and its square root 2. If we divide each value one by this square root we get four values each $\frac{1}{2}$. The values are now said to be “normalized”. Similarly the second Row *k* (respiratory disease) has nine values of one unit = total 9. The root of the sum of the square values is 3. Each “normalized” value becomes $\frac{1}{3}$. The normalized values for Table V now become as in Table VIII.

The sum of the squared values for each row of Table VIII is one. Thus, when normalized values are used in Equation (ii), the denominator becomes unity; the numerator—the sum of the normalized cross products—becomes equal to the correlation; in our example the correlation is $\frac{1}{2}$.

If now we prepare a graph (Fig. 2) in which the axes, instead of being at right angles, are placed at an angle whose cosine is equal to the correlation coefficient, then, when we plot the normalized entries from Table VIII, the sum of the areas formed by the cross products is zero.

In Fig. 2 the axes *OX* and *OY* are placed at an angle of 60° whose cosine is $\frac{1}{2}$ to correspond with the correlation from Table VIII. The area *OBCD* is that formed by the normalized cross products. The formula for this area, which consists of two right-angled triangles *OBC* and *ODC* put together is:

$$\frac{OB \cdot OD}{\sin DOB} - \frac{(OB^2 + OD^2) \cot DOB}{2} \dots (iii)$$

FIG. 2.—Graph of Table VIII plotted on oblique axes showing the “circular distribution”. The algebraic sum of the cross products is zero.

Substituting the numerical values, we obtain 0.2646. Similarly, area *OEB* which represents one "man-not-respiratory" is -0.0722 ; and area *ODF* which represents six "respiratory-not-men" is -0.1924 . The two minus areas exactly balance the plus area so that the algebraic sum is zero.

The diagram (Fig. 2) illustrates the use of two axes *X* and *Y* placed so that the angle between them has a cosine equal to the correlation between the two attributes—men and respiratory disease. This correlation appears in the appropriate side cell at the junction of a row and a column of Table VII. Every side cell can be similarly represented by a pair of axes at an angle. In three dimensions they are like the edges of a three-cornered pyramid standing on its point. In many dimensions they have to be imagined; but we may in the meantime think of some of them arranged like the ribs of a half-opened umbrella or pins in a pin cushion.

Thus, by using the convention of plotting persons on a graph with axes at an angle whose cosine is equal to the correlation coefficient, we get the association of the persons on two attributes into a circular distribution and in geometrical balance. The two axes with their appropriate angle form a pair of vectors which summarize the distribution of the persons, and which can be used to represent the persons in diagrammatic form on paper and in space.

RESOLVING THE VECTORS.—If we draw a line, *OX* say (Fig. 3), on a sheet of paper, we can give it a certain direction and a certain length; it then becomes a vector. From one end of the line, say *O*,

FIG. 3.—The resolution of the attribute vectors on the principal axes.

we can draw a pair of Axes I and II which can be either at right angles or oblique. We may imagine *OX* split along its length; one part projected or casting its shadow on to I and the other part on to II. Let the shadow on I be of length *OP*, and that on II of length *OQ*. These lengths are called the resolutions, projections, loadings, or saturations of *OX* on I and II respectively.

Now let us imagine another vector *OY* placed in the diagram and let it form an angle with *OX* whose cosine is equal to the correlation between *OX* and *OY*. Now *OY* can also have shadows on I and II. Let these be *OR* and *OS*. We can now think of I as containing the parts of *OX* and *OY* which they share in common, that is we can think of *OP* and *OR* as measures of the amount of the influence which the vectors have in common, or as measures of the strength of classifying them together. Thus suppose *OX* and *OY* form the vector system representing the distributions "men" and "respiratory disease" respectively. The angle between the two vectors is fixed by the cosine convention. We can now imagine that the influence which is common to "men" and "respiratory disease" is given on Axis I and that the contributions towards the influence are represented by the length *OP* for the men, and *OR* for the respiratory disease. Similar considerations apply to the projections on Axis II.

So far I has been placed arbitrarily between *OX* and *OY*; if we imagine I placed in various positions between *OX* and *OY* it is clear that the lengths of *OP* and *OR* will vary as I swings from one side to the other. One good choice of position is that in which the sum of squares of the loadings or shadows of *OX* and *OY* on I are made as large as possible. This is achieved when we minimize the sum of squares of the distances *YR* and *PX* of the points *Y* and *X* from *OR*. In the diagram this position is attained when I lies mid-way between *OX* and *OY*. In the jargon of factor analysis we "maximize" the sum of squares of the loadings or saturations of the vectors on the first or principal axis.

When there are three dimensions there will be three principal axes, I, II, and III; we then arrange for the sums of squares of the loadings on axes I and II to be as large as possible. When there are four dimensions* and four axes we maximize the sums of squares on I, II, and III; and so on for as many axes as necessary.

When there are many vectors the problem has to be solved in multidimensional space. The method of doing this is by finding the "latent roots and

vectors" of the table of correlations, which can be done on an electronic calculator. After adjustment, these calculations provide the loadings of all the attributes on the first, second, third, etc. principal axes.

EXTRACTING THE LOADINGS ON THE PRINCIPAL AXES.—From the table of correlations (like Table VII) we have to obtain a table of loadings of each attribute on each principal axis. As we have seen in the previous paragraph these loadings are nothing more than the shadows or projections of the vectors representing correlations on axes like I and II.

To compute the loadings for Table VII we first note that there are two determinables and five attributes. The number of axes is "attributes minus determinables plus 1", which in this case is $5 - 2 + 1 = 4$ axes. In Table VII there are five rows and five columns, one for each attribute. We write a column with five numbers—any numbers will do; the totals of the rows are a good guide. The first row of the table is post-multiplied by the new column, element by element, and is summed to form the top entry of a second new column; similarly with the second and other rows of the table to complete the second new column of five numbers. If the first and second new columns are not in proportion to each other we use the second column for a new trial and repeat the process. We go on performing the repetition for as many times as are necessary to make the two columns proportional. When we are satisfied, let us call the final columns u and v ; u is a "latent vector"—a column of numbers. We next reduce the items in u so as to make their sum of squares equal to a number called the "latent root" which is the largest item in v . When this is done we have the loadings on the First Principal Axis I.

In theory we next construct a 5×5 Table; this is done by writing, outside the Table along the top and the left-hand edges, the numbers we have just found for the first principal axis; the body of the Table is then filled in by forming all the products in order. This Table is subtracted element by element from Table VII and we get the "first factor residuals". This Table is just like Table VII; it is submitted to the same procedure of the u and v columns which eventually lead to the loadings on the Second Principal Axis II. This procedure goes on for as many axes as are needed—in the present case four.

In practice none of these calculations appeared on paper; the correlations from the frequency Table VI were punched on cards in binary notation and subjected to a "programme" in the electronic calculating machine at the National Physical Laboratory. The loadings on the first four principal axes for

Table VII were done in this way and appear in Table IX.

TABLE IX
LOADING ON FOUR PRINCIPAL AXES

Attributes	Axis I	Axis II	Axis III	Axis IV	Sum of Squares
Men	+ .37	+ .77	+ .51	.0	1
Women ..	+ .93	— .31	— .21	.0	1
No Disease ..	+ .37	— .32	+ .21	+ .84	1
Respiratory ..	+ .56	+ .68	— .45	+ .13	1
Hypertensive ..	+ .74	— .35	+ .23	— .52	1
Sum of Squares	2.00 1st Latent Root and Determinables	1.38 2nd Latent Root	+0.61 3rd Latent Root	1.00 4th Latent Root	5 Attributes

The first latent root is equal to the number of determinables—2.00 in this case shown as the sum of squares under Axis I. The sum of squares for each line across is one as shown in the last column, making a total 5, one for each attribute. The loadings on Axis I are proportional to the roots of the frequencies, *i.e.* to the diagonal entries in Table VI.

It is possible to apply a probability test to see if any of the columns, especially the last ones, could have arisen by chance.

THE SIGN PATTERN.—Table IX also shows a trend (which is more apparent in Table XI) for the plus and minus signs in each column to fall into a hierarchical order; thus the first column has all plus signs, the second column is roughly half plus and half minus, the third again sub-divides the halves in column two into plus and minus divisions; column four has two lines with zero entries and is not characteristic; the Table is too small, and because it is artificial may not show the phenomenon clearly at all.

Burt (1954) has pointed out that, in some investigations, the material itself has the intrinsic property of having signs which fall into a true hierarchical order and which can be used as the basis of a classification. In a perfect hierarchy, with eight attributes, we should get the ideal scheme (Table X, overleaf)

In the present investigation we extracted the loadings on three principal axes. We did not go any further because out of the total 42 for the sum of squares for 42 attributes, Axis I absorbed 9.00—corresponding with the number of determinables—Axis II absorbed 2.88, and Axis III absorbed 1.99, a total of 13.87 for the three axes. This left 28.13 squares to be divided over 39 axes—less than one square per axis. As each axis had to be split over 42 items, it was considered that the enormous labour and time on the electronic calculator of extracting

TABLE X
HIERARCHICAL FACTOR SIGN SCHEME FOR EIGHT ITEMS

Attribute	Axis I	Axis II	Axis III
1	+	-	+
2	+	-	+
3	+	-	+
4	+	-	+
5	+	+	-
6	+	+	-
7	+	+	-
8	+	+	-
	Group 0 8 items ++++ As these are the roots of the fre- quencies they are not of further interest	Group 1 4 items ----- Group 2 4 items ++++	Group 3 2 items ++ Group 4 2 items -- Group 5 2 items -- Group 6 2 items ++

By a suitable conversion the minus entries can be turned into groups with positive signs, as in the lower part of the Table, and if need be these groups can be regarded as factors, or influences, or classifications as we please.

further axes was not worthwhile. We could have applied a probability test but even this on such a large table was a formidable undertaking, and we decided not to do it.

It is interesting that this material did indeed form a hierarchical pattern, which is set out in Table XI.

ROTATING THE AXES.—Thurston (1947) recommends a further step. Instead of leaving the loadings on the principal axes as in Table XI, he rotates the axes round the origin *O* into various positions of obliquity. The new positions must have the property of forming a "simple pattern" in which the sums of squares of some loadings on certain axes are a minimum and on others a maximum. There are two methods of doing this, both dependent on geometrical manipulation. We submitted our data to both of these processes and obtained three factors from each method. As the manipulation often involved a choice of positions for the axes, and is to some extent arbitrary, we are disinclined to place much reliance on these methods, and have not included any results based on them.

THE PRINCIPAL AXIS SOLUTION.—In the principal axis solution, the fixing of the axes is unique in the sense that it is determined mathematically. With a Table of as many as 42 attributes, the extraction of the principal axes from the correlation table by desk methods would have required such an enormous amount of computational labour as to be almost prohibitive. We were fortunate in having the use of the electronic calculator at the National Physical Laboratory. Even with the great resources of this

TABLE XI
LOADINGS ON PRINCIPAL AXES

Attribute Reference Number	Attribute	Loadings	
		on Axis II	on Axis III
		Group 1	Group 3
13	Systolic blood pressure 130-159 ..	- 385	+ 217
39	3rd Disability: None	- 215	+ 202
24	1st Disability: None	- 301	+ 200
34	2nd Disability: None	- 353	+ 184
12	Haemoglobin 100-129	- 009	+ 173
19	Diastolic blood pressure 60-89 ..	- 383	+ 143
1	Men	- 166	+ 142
41	4th Disability: None	- 096	+ 090
3	Age 60-64	- 110	+ 082
27	1st Disability: Mental, nervous, senses ..	- 234	+ 060
		Group 4	Group 5
15	Systolic blood pressure 180-199 ..	- 073	- 255
8	Haemoglobin 30-69	- 043	- 174
14	Systolic blood pressure 160-179 ..	- 317	- 164
31	1st Disability: Respiratory	- 145	- 159
28	1st Disability: Heart, arteries	- 011	- 152
5	Age 70-74	- 085	- 107
20	Diastolic blood pressure 90-99 ..	- 270	- 103
33	1st Disability: Bones, malformations, ill-defined	- 156	- 080
26	1st Disability: Anaemia	- 125	- 078
32	1st Disability: Digestive, genito-urinary ..	- 149	- 043
9	Haemoglobin 70-79	- 066	- 037
		Group 2	Group 6
40	3rd Disability: All	+ 494	+ 464
21	Diastolic blood pressure 100-109 ..	+ 040	+ 411
42	4th Disability: All	+ 415	+ 390
35	2nd Disability: Infective, neoplasms, allergic	+ 402	+ 350
29	1st Disability: Benign hypertension ..	+ 276	+ 200
4	Age 65-69	+ 140	+ 192
25	1st Disability: Infective, neoplasms, allergic	+ 020	+ 132
2	Women	+ 129	+ 111
36	2nd Disability: Anaemia, mental, nervous, senses	+ 105	+ 092
16	Systolic blood pressure 200-219 ..	+ 212	+ 054
22	Diastolic blood pressure 110-119 ..	+ 162	+ 052
38	2nd Disability: Respiratory, digestive, genito-urinary	+ 075	+ 041
10	Haemoglobin 80-89	+ 033	+ 016
		Group 6	
30	1st Disability: Malignant hypertension ..	+ 619	+ 530
23	Diastolic blood pressure 120-159 ..	+ 608	+ 515
7	Age 80-84	+ 085	+ 288
18	Systolic blood pressure 240-299 ..	+ 484	+ 252
17	Systolic blood pressure 220-239 ..	+ 346	+ 183
6	Age 75-79	+ 029	+ 134
37	2nd Disability: Heart, arteries, hypertension	+ 201	+ 070
11	Haemoglobin 90-99	+ 073	+ 033

machine, the extraction of the latent root and vector for the third principal axis required over 4 hours running time. Without the machine we should have had to be content with a less accurate method.

DISCUSSION ON THE FACTOR ANALYSIS

The purpose of a factor analysis is to try to discover a set of factors whose interpretation might help us to understand the underlying order or grouping of the attributes; possibly to suggest the influences affecting the people in a group, and lead us to the causes of their disabilities; and perhaps to help to suggest remedies and means of prevention. We must, however, be careful to remember the

limitations of our data. Our sample contained only about half the population at risk. The set of attributes chosen for analysis is to some extent arbitrary in that attributes can be included or excluded at will; so that some attributes, perhaps those of the greatest importance, may be overlooked altogether. These variations lead to differences in the make-up of the factors and to variations, especially of emphasis, in the interpretations.

With these reservations in mind, the four groups of loadings on Principal Axis III in Table XI lead us to believe that our sample can be classified according to the four groups of attributes set out there. Each group is arranged in order of magnitude so as to put those with the highest loadings at the top. The first few attributes in each group are thus closely associated with the factor and with each other, and can form the basis of a group classification of the population. In each group it is possible that there might be a common influence operating either as cause or effect. For convenience, let the groups be re-named A, B, C, and D; Table XI then gives us the following:

Groups from Factor Analysis

MAIN ATTRIBUTES IN ORDER OF IMPORTANCE

Group A.—Healthiness:

Systolic blood pressure 130–159;
Absence of disease;
Haemoglobin 100–129;
Diastolic blood pressure 60–89;
Men.

Group B.—Rising Systolic Pressure, Anaemia, and Chest Diseases:

Systolic blood pressure 160–199;
Haemoglobin 30–69;
Respiratory disease;
Heart and arterial disease.

Group C.—Multiple Disabilities, Rising Diastolic Pressure, and Obesity:

Three or four diseases;
Diastolic blood pressure 100–109;
Infective, neoplastic, and allergic diseases (especially obesity);
Age 65–69.

Group D.—Malignant Hypertension:

Malignant hypertension;
Diastolic blood pressure, 120–159;
Age 80–84;
Systolic pressure 220–299.

Relationship to Blood Pressure.—The four groups are closely related to blood pressure. Persons with systolic blood pressure 130–159 tend to be healthy and to have high haemoglobin; those with systolic blood pressure 160–199 tend to have low haemoglobin and to suffer from respiratory, heart, and arterial diseases. Those

with diastolic pressure 100–109 tend to have multiple disabilities and obesity; those with diastolic pressure 120–159 tend to have malignant hypertension. Thus, as blood pressure rises, people so afflicted tend to have more and severer disabilities.

Deficiency Diseases.—In Group B (rising systolic pressure, anaemia, and chest diseases) there may be diet deficiencies, since deficiencies of iron and possibly of other trace elements like boron, manganese, molybdenum, and copper, as well as deficiencies of protein, render animals and plants susceptible not only to anaemia and deficiency diseases but also to infections.

It is tempting to wonder if rising blood pressure is a deficiency disease. In Group C, persons with three or four disabilities tend to have rising diastolic pressures and to be fat. Wilgram, Hartroft, and Best (1954) have suggested that deficiency of choline, a derivative of lecithin found in egg-yolk, may be a possible cause of cardiovascular damage. Most of our obese subjects were women (29 women, 5 men). In working-class households it is common for the women to deny themselves the expensive protein and protective foods, and to live mostly on starchy foods instead, especially bread and butter or margarine and preserves, or bread and fat dripping. In the North they also like tea-cakes, scones, and cakes. They tend to put on weight in the 30s, and become fatter as they grow older; some reach 18 stones in weight.

Corrective and Preventive Measures.—To combat the anaemia and deficiencies in Group B we can bring to the attention of doctors the prevalence of these conditions so that appropriate remedies can be prescribed. By health education we can teach the value and importance of a diet rich in proteins and minerals and encourage good habits in the choice of foodstuffs. The overweight subjects of Group C can be warned about the dangers of obesity and shown how to control their weight.

SIGNIFICANCE OF SYMPTOMS AND SIGNS IN HYPERTENSION.—From the clinical records we made a series of frequency tables showing the association of various symptoms and signs with both systolic and diastolic pressures above and below the average. All the following had a probability of association in random sampling greater than 0.05, and so do not pass the 20 to 1 test of significance:

In both men and women and for both systolic and diastolic pressures: mitral presystolic murmur, mitral systolic murmur, mitral diastolic murmur, split first sound, aortic systolic murmur, aortic diastolic murmur, accentuated aortic second sound, pulmonary systolic murmur, accentuated pulmonary second sound, enlarged right ventricle, pendulum rhythm, tic-tac rhythm, gallop rhythm, tremor, headache, tinnitus, vertigo, and ataxia.

AGE AND BLOOD PRESSURE.—With age the systolic blood pressure shows a significant rise in women but not in men; the rise in diastolic blood pressure is not significant in either. There is a large scatter

in pressure on each side of the mean at every age. Fig. 4 shows the regression of blood pressure on age in women. One standard deviation above or below the mean gives the point where 340 persons out of 1,000 would fall by random selection so that 680 would fall between the two standard deviation lines as shown.

ENLARGED LEFT VENTRICLE.—With enlarged left ventricle the association with both systolic and diastolic blood pressure is close. The probabilities were:

men systolic 0·05,
 men diastolic 0·01,
 women systolic 0·00001,
 women diastolic 0·004.

UNFOLDED AORTA.—With unfolded aorta the association was close with systolic blood pressure but not with diastolic. The probabilities were:

men systolic 0·01,
 men diastolic 0·10,
 women systolic 0·01,
 women diastolic 0·37.

Droller, Pemberton, and Roseman (1952) found no significant association between blood pressure and vertigo, tinnitus, angina of effort, clinically detectable arteriosclerosis, radiological size of the heart, and the subject's well-being and activity.

SUMMARY AND CONCLUSIONS

(1) We were able to carry out a thorough physical examination of 238 men and women aged 60 and over, out of 550 living in Tunstall and Silksworth, where the estimated combined population is 7,500.

(2) The observations were coded, punched on cards, and sorted; a frequency table was prepared and from this a factor analysis was made. In addition, correlations were ascertained, and probability tests by analysis of variance and χ^2 were carried out on certain parts of the data.

(3) The factor analysis gave four groups:

Group A.—Healthiness:
 Systolic blood pressure 130–159;
 No disease;
 Haemoglobin 100–129.

Group B.—Rising Systolic Pressure, Anaemia, Respiratory and Circulatory Diseases:

Systolic blood pressure 160–199;
Haemoglobin 30–69;
Respiratory, heart, and arterial diseases.

Group C.—Multiple Disabilities, Rising Diastolic Pressure, and Obesity:

Three or four diseases;
Diastolic blood pressure 100–109;
Obesity.

Group D.—Malignant Hypertension:

Diastolic blood pressure 120–159;
Systolic blood pressure 220–299.

(4) Groups A and B are related to systolic blood pressure; Groups C and D to diastolic blood pressure.

(5) It is possible that diet deficiencies may be the cause of some of the disabilities in Groups B and C.

(6) Correction and prevention can be obtained by prescription of remedies for deficiencies, by education in dietetics, and by instruction in weight control.

(7) Systolic blood pressure rises significantly with age in women. Many symptoms and signs traditionally associated with hypertension proved not to be statistically significant. When diagnosed by screening, the association of enlarged left ventricle and unfolded aorta with systolic and diastolic blood pressures is close.

We acknowledge with grateful thanks the assistance we have received from many people during the course of this research. We would specially like to mention our gratitude for the facilities placed at our disposal in the department of geriatrics at the Sunderland General Hospital, and to Drs. E. W. Williams, H. S. Stott, and W. Lloyd, and the sister, nurses, and laboratory technicians for their assistance with the medical examinations. Also to the Director of the National Physical Laboratory for permission to use the electronic calculator, and Dr. Goodwin, Mr. Vickers, and Miss Rigg of the Mathematics Division for assistance in sorting and in the extraction of the latent roots and vectors.

REFERENCES

- Adams, G. F., and Cheeseman, E. A. (1951). "Old People in Northern Ireland". A Report to the Northern Ireland Hospitals Authority, Belfast, pp. 25–47.
Burt, C. (1950). *Brit. J. Psych., Stat. Sect.*, 3, 166–85.
— (1954). *Brit. J. of stat. Psych.*, 7, 15.
Cattell, R. B. (1952). "Factor Analysis". Harper, New York.
Droller, H., Pemberton, J., and Roseman, C. (1952). *Brit. med. J.*, 2, 968.
Exton-Smith, A. N. (1955). "Medical Problems of Old Age". Wright Bristol.
Hobson, W., and Pemberton, J. (1955). "The Health of the Elderly at Home". Butterworth, London.
Maddison, J. (1953). *Publ. Hlth*, 66, 186.
Sheldon, J. H. (1948). "The Social Medicine of Old Age". Oxford University Press, London.
Simonds, W. H., and Stewart, A. (1954). *British Journal of Preventive and Social Medicine*, 8, 139.
Thomson, G. H. (1950). "The Factorial Analysis of Human Ability", 4th ed. University of London Press, London.
Thurstone, L. L. (1947). "Multiple-Factor Analysis". University of Chicago Press, Chicago.
Walker, J. V. (1947). *Med. Offr.*, 78, 169.
Wilgram, G. F., Hartroft, W. S., and Best, C. H. (1954). *Brit. med. J.*, 2, 1.
World Health Organization (1948). "Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death". W.H.O., Geneva, and H.M.S.O., London.