FACTOR ANALYSIS AS AN AID TO NUTRITIONAL ASSESSMENT

By W. H. HAMMOND

I hope that this article may be of interest in showing how a layman in the field of medicine regards certain medical problems from a different—and it is hoped—suggestive angle.

At the outset I would like to make it clear that the present paper is a suggestion for a programme of research and the examples given are merely illustrative. I hope to follow up some of the statistical investigations suggested and would welcome sources of quantitative data of the kind mentioned.

THE NEED FOR QUANTITATIVE MEASURES IN DIAGNOSIS

The increasing state control for medical services emphasizes both the need and opportunity for obtaining quantitative data concerning the health of the general population. With social insurance in the forefront of attention, with the need for medical examination for national service, and above all when the nation's food policy is determined increasingly by the government, it becomes essential to establish health norms. These are necessary to assess the effects of different food policies and war or labour conditions as they affect the general population. Further, without such norms it is impossible to weigh up the effectiveness of different preventive measures for safeguarding health or provide standards with which to compare individual divergences.

DIAGNOSIS IN GENERAL AND 'SPECIFIC' DISEASES

From the foregoing it is evident that we are concerned with measuring general health conditions rather than 'specific' diseases (in the sense of applying to any particular organ). These two conditions represent completely different problems from the point of view of measurement. In the latter type the symptoms are usually regarded as being all-or-none, e.g. presence or absence of a tumour, particular infection and so on. Here it is sufficient to have a purely qualitative diagnosis. This is ordinarily given in a consultation or examination, possibly with the aid of one or more laboratory tests as in the taking of swabs or urine analyses. Even so every general practitioner has to satisfy himself as to the seriousness of the individual case, and to this extent some form of quantitative assessment is involved. However, the important point is to decide merely whether the pathological condition is present or not and therefore explicit quantitative rating is unnecessary. On the other hand, when we come to problems concerning general health it is the pattern or degree of development of perhaps quite a number of traits which points to the disease. There is no sharp dividing line between the normal and the subnormal and there is no single sign or symptom which is characteristic of either condition alone, such that it can be used to differentiate between them. When dealing with such syndromes all aspects of their development should be considered, each one contributing to the final assessment. This is a complex problem and in my opinion it ought not to be left to more or less superficial intuitive impression. I consider it is expecting too much to attempt to sum up adequately within a single value or rating the effects of these complex interacting influences with no help but past experience. I would not suggest that it is impossible to give reliable and consistent assessments by general impression, but if diagnosis is to be scientific and if results are to be comparable then a more rigorous treatment is necessary. In giving a single rating for an aggregate of symptoms there is a double cause of error. In the first place, without any objective indicator, individual doctors may differ in assessing any one symptom. In addition, each may regard the signs as having widely different importance so that the same ratings for individual characteristics may be combined to produce an entirely different result. This may make a vital difference especially in borderline cases.

MEASUREMENT OF STATE OF NUTRITION IN CHILDREN

Nutritional state accords well with the kind of general condition just mentioned, since its measurement depends upon assessing a number of different signs few of which are capable of objective measurement, and it will perhaps help to clarify some of the issues if we consider an actual study in this field.

As our example we may take R. H. Jones's (1938) Physical indices and clinical assessments of the nutrition of school children.

One aim of the Liverpool survey was to test how far different school medical officers agreed in their assessments of nutrition. Four doctors were asked to rate the same children independently and the results were compared. Each doctor's reliability was checked by repeating the assessments after a few days' interval. Jones's conclusion from the evidence was that 'the doctors showed important disagreement not only with each other, but also with their own assessments of the same population after a short time interval'. He adds, 'the present criticism is directed against the method, not against the doctors concerned' (who, as he points out, knowing the purpose of the inquiry would be likely to take greater care than usual).

In order to see whether the four medical officers were typical the inquiry was repeated in Manchester, Northwich and Breconshire. The Liverpool results were amply confirmed. The same scheme was followed in each case and the fourfold grading as recommended by the Board of Education was adopted. The results were as follows:

Boys placed in the same nutrition grade by all the doctors in each survey: 51 % in Liverpool, 43 % in Manchester, 60 % in Northwich, 19 % in Breconshire (six doctors).

Since there were only four grades into which the subjects could be placed the classification was not a very stringent one. However, to make the division still coarser, the top grades A and B were combined and regarded as satisfactory, whilst C and D were regarded as undernourished. With this twofold division into normal and subnormal, agreement between all examining doctors was as follows:

67 % Liverpool, 68 % Manchester, 74 % Northwich, 48 % Breconshire.

This means that even with the special precautions taken there was some disagreement in the grading of over half the boys, whilst in approximately one-third of them the doctors were not all agreed as to whether they were to be regarded as undernourished or not.

To a psychologist these findings are not altogether unexpected, since similar inconsistency has been shown to exist in assessing examination scripts by this method of general impression (Hartog & Rhodes, 1935). The Ministry of Health's Advisory Committee on Nutrition (1926) endorses these findings when it says that it is unable to recommend any method of assessing nutrition as reliable. It considers the Board of Education's suggestions (as adopted in these surveys) to be the most promising so far.

Obviously here is a problem requiring further research, especially with a view to overcoming the sources of inconsistency mentioned. Fortunately, we have before us the example of the way in which psychologists have overcome similar difficulties in the matter of assessing intelligence—an analogous problem.

The first essential is to define provisionally the concept of nutritional state or general health condition in such a way that different observers will be concerned with the same thing. This is itself no easy task, as the voluminous literature on the subject shows. Psychologists faced with a similar controversy in defining 'intelligence' finally adopted the by no means elegant or logical concept that 'intelligence is what the intelligence test measures'. The procedure was to get together an assortment of intellectual tests and apply these. The results were then analysed to see how far the subtests agreed with each other. On the assumption that 'intelligence' was something that these tests were all attempting to measure, but succeeding with varying degrees of accuracy, we could be fairly sure that some pool of the tests, say, their aggregate, would be likely to provide a better measure than any one test. This aggregate or other pooled result thus becomes the criterion against which to test any individual item. If the subtest agrees closely with results from the whole battery then it may be considered a good test and it should be given prominence in diagnosing or measuring the general state. If, however, the results do not follow those of the whole battery closely then the test may be considered to be measuring something specific to itself. In that case it should be discarded entirely or given a very small influence in determining the pool result. By trying out various tests and by combining them according to their provisionally determined diagnostic importance, it is possible to build up a battery which gives a relatively pure measure of the common 'factor' which it is attempting to establish. The problem of how to combine the tests is a statistical one which need not concern us at the moment so long as the general principle is clear. If we employ such an approach in measuring nutritional state we see that it requires us to break up this concept into its elements. The first stage is for those with special knowledge and experience in this field to decide provisionally what signs of nutritional state to include, e.g. general liveliness, posture and expression, condition of skin, amount and condition of musculature, subcutaneous fat and so on. A representative sample of children should then be given graded assessments in each sign. Here we encounter the difficulty that each scale must be subjectively determined as there are no intrinsically fixed limits.* This problem of measurement is still only partially solved, and all we can do is to employ as many precautions as possible. For example, if we have the subjects assessed by several doctors, the average grade is more likely to be reliable than that of any one. Moreover, since we are at present interested in the tests rather than in the testees we can initially use only those subjects in which the doctors show the greatest unanimity. Again we should use objective tests, e.g. physiological and anthropometric ones, to supplement the clinical diagnosis. By correlating each test with the others we can see how far they are interrelated. The table of intercorrelations can then be analysed, and from the results we can see whether there is any common factor corresponding to a general state of nutrition and further which test symptoms are the most representative of the general condition which the whole battery is attempting to measure. It may seem unnecessary to calculate correlations merely to say which symptom agrees most closely with others, yet by a simple extension it is possible to throw light on the causes operating. If perfect correlation between two variables implies similar causes, and zero correlation indicates the absence of common elements, then partial relation may be regarded as an intermediate condition in which some of the influences are acting on both variables whilst others are specific to one. Where several variables are concerned the factors may operate over all, some, or only one of them. Factorial analysis can be used first to separate, then estimate, these general, group, or individual factors.

* The demarcation of limits need not be entirely arbitrary, for we could assume (unless there was reason to suppose otherwise, e.g. in especially poor districts) that the distribution of nutrition follows that most generally found in nature, i.e. an approximately normal one with most cases in the middle groups and with decreasing numbers as we pass to the extremes. If so, this means that the grading could be chosen to ensure any agreed proportions within the different grades, e.g.

A	B	C	D	
12·5 %	37·5 %	37·5 %	12·5 %	Ratio 1:3:3:1

Cf. Dr Kerr's classification of height, the highest (12·5 %) being 'tall', middle (75 %) 'medium' and lowest (12·5 %) 'small' (Kerr, 1926, p. 834).

https://doi.org/10.1017/S0022172400013140 Published online by Cambridge University Press

By identifying the factors (from the nature of the tests) they can be used for prediction even if they cannot be measured directly. If nutritional condition could be measured as one measures temperature by some single constant effect, there would be no need to invoke correlations or factors at all.

Having tested which symptoms are good indicators of nutritional state, the next problem is to combine the test results so as to give the best measure of nutrition as we have now defined it.

A study somewhat along these lines has already been carried out in America under the American Child Health Association (1929, 1935). 'Emphases' (or diagnostic weights) were calculated for a number of traits and the traits were combined in a regression formula to specify state of nutrition. The results were not, however, conclusive.

We can illustrate the procedures involved by considering the correlations between doctors in the Liverpool survey (Jones, 1938, p. 20) (Table 1). For our purpose it does not matter whether we are analysing persons or tests; much the same principles hold in both cases.

Table 1. *Intercorrelations between doctors*

Doctor	1	2	3	4	
1	(6·61)	0·64	0·58	0·64	
2	0·64	(0·69)	0·62	0·67	
3	0·58	0·62	(0·55)	0·60	
4	0·64	0·67	0·60	(0·66)	
Saturations	0·78	0·83	0·74	0·81	Average 0·79
Rel. weights	1·22	1·62	1·00	1·47	

The 'saturations' are obtained by a procedure of factor analysis known as simple summation. (For details see reference books on Factor Analysis, e.g. Burt, 1939; Spearman, 1927; Thomson, 1939; Thurston, 1937.) They indicate the degree to which each doctor conforms with the combined findings. They are really correlation coefficients between the doctor's results and the weighted average. The figures indicate that the four doctors are all moderately representative and show no great difference as measured by their agreement with the rest. If we were to rely on any one doctor alone we should naturally employ no. 2.* What is the correlation between the team as a whole and the pure factor? (for even the team only estimates this otherwise unknown entity). When each doctor is weighted according to his diagnostic ability we obtain a multiple correlation of 0·94 between the battery and the criterion. However, if we combine the doctors' results without first weighting them differentially the correlation is 0·93. In this case, therefore, the unweighted average gives almost as good a result as a weighted one. This is because there is so little difference between the doctors' diagnostic ability and therefore the weights are very similar.

* We are assuming that no. 2 is the 'best' simply because he conforms with the team most closely. It might, however, be that no. 3 is the best judge of nutrition whilst the rest are influenced by signs which are not really relevant. This might well be the case in a team composed of laymen and one experienced doctor. Methodologically we are concerned with consensus of opinion whether right or wrong.

We have used our analysis technique to discover who is the most representative doctor,* and further what combination gives the best team result, i.e. the best correlation with the factor when this is an entity latent in the assessments. Suppose now that we have some objective criterion against which to check the doctors, e.g. a physiological test of nutrition or an index such as Tuxford's (1942) Index,† how far do the doctors agree with it and how far will a pooled result from all the doctors improve upon any one? We can again illustrate this problem by reference to R. H. Jones's (1938) paper.

P. 22 gives the correlations between the doctors and Tuxford's index which are as follows:

Doctor	1	2	3	4
Correlation	0·53	0·50	0·56	0·40

If these correlations are added to the intercorrelations of Table 1 as a fifth row and column, and the table is analysed as before, we obtain the following correlations with the new factor (i.e. 'saturations'):

Doctor	1	2	3	4	5 (Index)
Saturations	0·80	0·82	0·79	0·76	0·63

The saturation coefficients are for the factor common to the doctors' estimates and the index, and the simple correlation of the battery and the factor is 0·93.

The reason for the low saturation of the index,‡ is that whereas it measures an aspect of physical development, the doctors were asked to ignore body size in their diagnosis. Thus two slightly different but related things, i.e. physical development and clinical symptoms of nutrition, are being pooled.

If, instead of taking the common factor as our criterion,

* The saturation coefficients indicate the doctors' general agreement but if we want information about their grading over a particular range, e.g. borderline cases of nutritional deficiency, we should calculate tetrachoric correlations about that level. Table 2 gives correlations between examiners' scripts of candidates whose abilities were previously adjudged equal (hence having a reduced range of variability).

It should also be noted that correlations between persons are based on deviations about the doctors' own means, thus eliminating differences in their standards. (This incidentally removes one of the difficulties already mentioned, namely, the subjective and individual nature of the doctors' standards.) Standardization also removes differences in their variability, i.e. whether they use extreme grades frequently or not.

† Tuxford index: $\frac{\text{Wt. (kg.)}}{\text{Ht. (cm.)}} \times \frac{381\text{—months}}{54}$, restandardized to $\frac{W}{H} \times \frac{335\text{—age}}{48}$.

‡ Burt (1937) reports that Dr Kerr and himself, working under the L.C.C., obtained closer agreement between various indices of height and weight and nutritional grade than the above; correlation for ht./wt. and nutrition = 0·69 and for weight alone the correlation was 0·54.

we use Tuxford's index as the standard* we can see how far each doctor and how far the team correlates with it. The doctors' correlations range from 0·40 to 0·56, whereas the best weighted team gives a multiple correlation of 0·62 and a simple average of the doctors correlates to the extent of 0·59 with the index. Both team results show a considerable improvement upon the highest individual's correlation. This illustration shows how the validity of subjective estimates can be increased by using a properly weighted team of assessors. In itself this has a limited use in practice, for it would be prohibitively expensive to employ four or more doctors in any general survey, though they might be used in a research to establish standards for general reference later.

The method of multiple correlation is of greater practical importance in testing signs of nutrition rather than the assessors. In this connexion Burt (1937, p. 178) says: 'The ideal method would be to deduce a regression-equation from the partial correlations between nutrition and all the measurable characteristics significantly correlated with it.' This implies, as we have seen, that some external criterion is available, whereas the analogous procedure of factor analysis requires no such condition.

Table 2. *Correlations between examiners*

Examiners	F	A	B	E	D	C
F	—	0·86	0·84	0·82	0·84	0·71
A	0·86	—	0·80	0·74	0·85	0·71
B	0·84	0·80	—	0·80	0·81	0·67
E	0·82	0·74	0·80	—	0·72	0·69
D	0·84	0·85	0·81	0·72	—	0·48
C	0·71	0·71	0·67	0·69	0·48	—
Saturations	0·95	0·92	0·91	0·87	0·84	0·72

Average 0·87

Correlation between weighted battery and the hypothetical 'true' mark = 0·98.
Correlation of unweighted battery and the hypothetical 'true' mark = 0·97.

It is worth while going to some trouble to establish these standards in medicine because the chief criticism against attempting to measure in this field, as in psychology, is that there are no constant standards and no units to make valid comparisons possible between different investigations. In the long run this trouble will be repaid in that once reliable criteria have been established it will be possible to test indices, quick 'foot-rule' methods or readily observable signs. We can then determine what combination will give the highest correlation with the more reliable signs of disease as given, for example, by laboratory tests.

We are provided with an interesting comparison with these results for doctors in the case of an investigation into the agreement between examiners. Burt's memorandum (see Hartog, Rhodes & Burt, 1936), '*The Analysis of Examination Marks*', gives the correlations between the marks of six examiners, each marking the same fifteen scripts for an examination in Latin. The experiment was unusual in that the candidates had all been adjudged equal by another examiner previously. Saturation coefficients for the general factor were calculated (the method being that of simple summation as described in the Appendix to Burt (1939) (see Table 2).

The agreement between the examiners is rather closer than that between the doctors though a direct comparison is hardly possible on account of the selection mentioned above.

* To choose the index as the criterion is merely as an illustration, for one would scarcely use this in preference to the combined result of four doctors: Results from any one doctor might be used equally well. (It is interesting to note in passing that R. H. Jones considers Tuxford's index to be 'at least as satisfactory as a randomly chosen medical officer').

CLASSIFICATION OF MALNUTRITION

We can carry the analysis a stage further than the general factor to establish group similarities. (See Spearman, 1927; Thomson, 1939; Thurston, 1937 and in particular Burt (1939), where full working instructions are reproduced.) Applying these procedures to the correlations in the example of the doctors' nutritional assessments and Tuxford's index we obtain the following grouping and saturation coefficients for the second factor:

Doctor	2	4	1	3	5 (index)
Saturation coeff.	+0·10	+0·27	+0·05	−0·21	−0·21

Thus doctors 2, 4 and 1 showed group similarity with each other in addition to their general level of agreement, whilst no. 3 gave results more akin to those of the index. As with the doctors, so the symptoms themselves may be grouped. For example, in the study of nutrition we might have groups of signs indicating specific vitamin or other dietary deficiencies. Or again we might expect group factors to emerge corresponding to early or recent nutritional defects, e.g. group 1 might contain a factor showing itself in the development of bones (knock-knee, 'pigeon chest' or stunted growth), group 2 factor may affect mainly the soft tissues, mucous membranes and blood. (For clinical evidence see Burt (1937, p. 180), 'Social' and 'Physiological' malnutrition, and Wilkins (1938), 'Nutrition Past or Present'.)

An application in an allied field which promises to be fruitful is in the study of the interrelations of the endocrine glands and the resultant types. Here we have hormones whose action may be local or general corresponding with group or general factors.

SUMMARY

1. *The problem of assessing nutritional state*

This paper emphasizes the need for measurement in diagnosis and recognizes two distinct problems, those of:

(*a*) Diagnosing specific diseases in the sense of distinct pathological conditions.

(*b*) Assessing general malconditions where there is no sharp dividing line between the normal and the subnormal.

It suggests a programme of research designed to overcome some of the difficulties inherent in (*b*) in the case of nutritional assessment.

2. *Establishing general health standards*

The first requirement is the scientific standardization of the elements making up the concept of health and good physique. This is a complex matter requiring all the help which statistical procedures can give. Moreover, we must check the reliability of the assessors and the indicators which they use before we can place any reliance on results purporting to relate to the population studied.

Provisionally we begin by collecting a battery of indicators of general health; each is then given a graded assessment for each child (and the assessment is repeated to obtain its reliability), and by employing the statistical device of factor analysis we can determine what combination of symptoms will give the highest agreement with the criterion of general health abstracted from these individually imperfect measures. We are actually making explicit the stages which are normally telescoped in the method of 'general impression' (which is the method usually adopted in assessing general health, state of nutrition, etc., despite evidence of its unreliability). In this method a number of individual characteristics are fused together without ensuring that different observers will attach equal importance to the signs or even that they will take into account the same ones. The emphasis to be given to each symptom in the final mark is determined statistically by its agreement with the aggregate result. Some indicators may give results so unrelated to the others that they must be eliminated. Other tests, whether physiological, functional or anatomical, may later be added to the battery to reinforce it. A point of practical importance is that some readily assessable symptom or index may be found which gives results in close enough agreement with those of the standard battery to enable it to be used as a substitute, thus shortening the task of diagnosis without appreciably lessening its accuracy.

An alternative criterion could, as we have seen, be the weighted combination of doctors' assessments, for by weighting the results according to each doctor's agreement with his colleagues we were able to increase the agreement between the team result and the hypothetical true mark. However, since the ordinary methods of assessment are unreliable, even the pooled results of the team do not give as good a standard as the standardized battery of tests.

A third standard is sometimes possible in the form of an objective accurately measurable physiological laboratory test or anatomical index. If this is inconvenient to apply generally, any single symptom or pool may be tested by correlation with it to see if the agreement is close enough for practical purposes.

3. *Study of more specific nutritional defects*

Factor analysis can be used to discover and assess more limited similarities than those covering the whole range of observers or test symptoms. For example, the doctors tended to show group affinities (depending possibly on their different emphasis in diagnosis) in addition to their somewhat low general conformity. Applying similar analytic methods to the study of symptoms, we might isolate specific nutritional deficiencies or it is possible that group factors may be revealed corresponding to the effects of recent or early adverse nutritional conditions respectively.

Whilst the examples of analytic techniques have here been applied to illustrative cases in nutritional and general health assessment, they are equally applicable to other fields of medicine, in particular to endocrinology and to the study of predisposing conditions of disease.

My thanks are due to Prof. Burt and Dr E. H. Wilkins for their suggestions for modifying my original draft. However, this in no way commits them to agreement with the views expressed here.

[*Note added in Proof.*] Since writing this article I have had an opportunity of correlating and analysing some results of nutritional surveys involving clinical and biochemical signs of malnutrition. The evidence so far obtained indicates a rather weak general factor for clinical signs identifiable with general nutritional state. I hope to be able to give the full results when the material becomes available for publication.

REFERENCES

AMERICAN CHILD HEALTH ASS. (1929). *Physical Measures of Growth and Nutrition*, Mono. 2 and (1935) *Nutritional Status Indices*. N.Y. Cited with critical appraisement by R. H. Jones (1937), M.A. Thesis, Liverpool University.

BURT, C. (1937). *The Backward Child*. Univ. London Press.

BURT, C. (1939). *Factors of the Mind*. Univ. London Press.

HARTOG, P. & RHODES, E. C. (1935). *Examination of Examinations*. London.

HARTOG, P., RHODES, E. C. & BURT, C. (1936). *Marks of Examiners* with Memorandum by Burt, 'The Analysis of Examination Marks'.

JONES, R. H. (1938). Physical indices and clinical assessments of the nutrition of school children. *J.R. Statist. Soc.* 101 (Part 1) 1–52.

KERR, J. (1926). *Fundamentals of School Health*. London.

MINISTRY OF HEALTH'S ADVISORY COMMITTEE ON NUTRITION (1926). *1st Report*, 1937.

SPEARMAN, C. (1927). *Abilities of Man*. Macmillan.

THOMSON, G. H. (1939). *Factorial Analysis of Human Ability*. Univ. London Press.

THURSTON, L. L. (1937). *Vectors of the Mind*. Univ. Chicago.

TUXFORD, A. W. (1942). Indices of physical development in children. *J. Hyg., Camb.*, **42**, 549.

WILKINS, E. H. (1938). How do we assess nutrition? *Publ. Hlth.* **51**, 293.

(*MS. received for publication* 24. VI. 43.—Ed.)