RELIABILITY OF THE CLINICAL ASSESSMENT OF "NUTRITIONAL STATE"

BY

E. R. BRANSBY, Ph.D.

AND

W. H. HAMMOND, M.Sc.

(From the Ministry of Health)

To test how far clinicians agree in their assessment of the "nutritional state" two field trials have been made—one at Sheffield in January, 1949, and the other at Tottenham in March, 1950.* As the results of the two trials are substantially the same, the Sheffield data have, to avoid duplication, largely been omitted.

Five clinicians participated at Sheffield—namely, Drs. W. T. C. Berry, P. J. Cowin, R. B. Hawes, E. H. M. Milligan, and John Pemberton. Nine clinicians and a school nurse took part at Tottenham—namely, the above five with the exception of Dr. Milligan, together with Drs. E. W. Adcock, Cecile Asher, R. E. C. Copithorne, D. MacCarthy, Hugh S. Stannus, and Nurse Connor.† At Sheffield two clinicians, and at Tottenham four clinicians, could not participate the whole time.

201 children were examined at Sheffield and 100 at Tottenham. Of these about a quarter were "good" children and one-half "medium" children, as judged by the existing school records; the other quarter were individually chosen by the school medical officers, in association with the teachers, on account of their poor condition. Later in the report these three groups of children are referred to as "selection groups." Boys and girls aged 6 to 13 years were included.

Procedure

In both trials the children were examined by each of the participating clinicians. The order in which the children were examined was randomly arranged—that is, the "good" and "medium" children and those specially selected were mingled. The clinicians did not know beforehand how the children had been selected or the proportions of the three groups in the sample. At Sheffield 49 children and at Tottenham 24 children were, unknown to the clinicians, examined twice.

Each clinician recorded for each child (a) his clinical grade and (b) the incidence of the clinical signs listed in Table III. Certain signs were not noted at Tottenham, as the Sheffield results had shown that they had no particular association with the "nutritional" grades. It is to be noted that the clinicians had no knowledge of the medical history or the home background of the children. Neither had they access to the school medical inspection records.

All but one of the clinicians assessed what they termed "nutritional state"; the other clinician assessed "general condition." Most of the clinicians assessing "nutritional state" stated that it was doubtful if any differentiation

could be made between "nutritional state" and "general condition" or "general fitness." This is the reason for putting the term "nutritional state" in inverted commas. It would prevent confusion if agreement could be reached on the terminology of such assessments. The Ministry of Education has already substituted the assessment of general condition for the assessment of nutritional state.

The trial at Sheffield was organized by Dr. H. M. Cohen, and that at Tottenham by Dr. G. H. Hogben and Dr. P. A. Tyser. We are glad to express our warm appreciation of their valued co-operation.

When used in the report the following terms have the attached definitions:

Incidence: the frequency with which a particular condition occurs.

Accordance: the agreement between the findings of different clinicians from the examination of the same children.

Reliability: the agreement between the findings of one clinician from two examinations of the same children.

Clinical Grades

Incidence (Table I).—The clinicians fall into two groups—namely, those who grade the average or normal child as "good" (Group I) and those who grade the average or normal child as "fair" (Group II). There are marked differences in each group in the number of

TABLE I.—Percentage of 100 Children Graded "Good" (G), "Fair" (F), and "Poor" (P), at Tottenham

	Grade			Twofold Grouping	
	G	F	P	Satisfactory	Unsatisfactory
Group I:					
Clinician A	76	16	8	76	24
" B	78	20	2	78	22
" C	74	24	2	74	26
" D	79	17	4	79	21
" E	64	18	18	64	36
" F	80	18	2	80	20
Group II:					
Clinician G	29	57	14	86	14
" H	23	42	35	65	35
" I	33	51	16	84	16
" J	27	59	14	86	14

NOTE.—Clinician D used a fourfold grading, and grades 2 and 3 were grouped as "Fair."

children graded "good" (G), "fair" (F), and "poor" (P). Thus the percentage of children graded P ranged from 2 to 18 in Group I and from 14 to 35 in Group II.

In view of the differences in meaning attached to grades G, F, and P by the clinicians in Groups I and II, a certain measure of standardization between them, at least on paper, was effected by regrouping the children into two nutritional grades, "satisfactory" and "unsatisfactory." The children graded G by the Group I clinicians were regraded as "satisfactory" and the F and P children as "unsatisfactory"; the children graded G and F by the Group II clinicians were regraded as "satisfactory" and the P children as "unsatisfactory." An example will make this clear. For clinician "A," who fell into Group I, the G (76%) children were regraded as "satisfactory," the F (16%) and P (8%) children as "unsatisfactory"; for clinician "G," who fell into Group II, the G (29%) and F (57%) children were regraded as "satisfactory" and the P (14%) children as "unsatisfactory." With this regrouping the results for

*One reason for these trials was that the clinicians making the nutritional surveys for the Ministry of Health were not satisfied with the methods of assessment then being used.

†Throughout the report Nurse Connor is, for convenience, included in the term "clinician."

different clinicians were more comparable but still varied considerably. Thus, the percentage graded "unsatisfactory" ranged from 14 to 36%.

Accordance.—Table II is based on the standardized classification "satisfactory" or "unsatisfactory." The more clinicians in the trial the less the likelihood of complete agreement. Moreover, the level of agreement depends on the number of children marked down by

TABLE II.—*Accordance: Percentage Agreement Between the "Nutritional" Assessments (Standardized) at Tottenham*

	All Children		Children Graded "Satisfactory"*		Children Graded "Unsatisfactory"*	
	Actual % Agreement	% Chance Expectation	Actual % Agreement	% Chance Expectation	Actual % Agreement	% Chance Expectation
All clinicians in agreement ..	45	12	55	12	5	0
4/5 " " "	74	29	83	30	40	15
2/3 " " "	88	46	92	48	75	17

* The classifications "satisfactory" and "unsatisfactory" are the majority verdicts of the clinicians.

various clinicians. These factors are taken into account by comparing the actual agreement with the agreement expected by chance. Table II shows a considerable measure of disagreement between the clinicians; agreement is better for children graded "satisfactory" than for those graded "unsatisfactory."

Reliability.—The children graded differently at the first and second examinations varied from 8 to 18% at Sheffield and from 4 to 37% at Tottenham for the different clinicians. These differences were on the average about half those expected purely by chance. The differences were mostly of one grade—that is, a child might be graded F on one occasion and G or P on the other—but there were occasionally differences of two grades—that is, where the child was graded G on one occasion and P on the other. The clinicians who down-graded the most children had the most discrepancies, simply because they had a greater possibility of variation than had those clinicians who down-graded fewer.

Clinical Signs

Incidence (Table III).—There are wide variations in the incidence of the different signs as recorded by the several clinicians.

Accordance (Table III).—There was much disagreement between clinicians. This was more marked where the sign was present than where it was absent.

Reliability.—The level of reliability depends to some extent on the level of incidence. For example, if a clinician does not record a particular sign as present in any child there must of necessity be good agreement between the first and second examinations. The actual agreement at the two examinations, compared with the agreement expected by chance, is shown in Table III. On the average the actual agreement is somewhat more than is expected by chance. For some signs—for example, lethargy and poor muscle tone—there was considerable disagreement. On the whole the reliability was not particularly good.

An interesting feature of the results is that the reliability of the nurse and her agreement with the doctors was as good as that of the individual doctors.

Relationship between Clinical Grades and Clinical Signs

Table III also shows the incidence of the clinical signs for the nutritional grades G, F, and P at Tottenham. Both at Sheffield and at Tottenham the incidence of the following signs increased as the nutritional grade worsened: pallor, fatigue rings, mucous membrane pallor, lethargy, poor posture, diminished fat, poor muscle tone, poor muscle development, poor grip, inelastic skin. At Sheffield the incidence of folliculosis, gingivitis, and enlarged tongue papillae varied little according to nutritional grade. Table IV shows that each clinician recorded least clinical signs in his G grade, more in F, and most in P.

There was a high degree of association between the incidence of clinical signs and "nutritional" grade. This was so for all participating clinicians, though some put more emphasis on particular signs. Thus there was for all clinicians a close association between "nutritional" grade and muscular development, while the association with "inelastic skin" was close for only four clinicians. The Group I clinicians tended to down-grade children with fewer clinical signs than did the Group II clinicians. For example, a child with a certain number of signs might be graded P by Group I, but F by Group II. Even with the standardized (satisfactory/unsatisfactory) grading, there is still much variation in the incidence of clinical signs in the two grades as recorded by different clinicians.

TABLE III.—*Clinical Signs as Recorded by the 10 Clinicians at Tottenham*

	Pallor	Fatigue Rings	Mucous Membrane Pallor	Lethargy	Poor Posture	Diminished Fat	Poor Muscle Tone	Poor Muscle Development	Poor Grip	Inelastic Skin
Incidence: range (%)	2-34	15-38	0-8	0-20	21-35	7-42	0-31	3-43	4-24	0-14
Accordance (%):										
Sign absent:										
All clinicians in agreement ..	53	22	79	65	27	59	43	56	56	78
4/5 or more in agreement ..	85	71	99	94	79	79	85	78	91	96
Sign present:*										
All clinicians in agreement ..	0	0	4	0	15	0	0	0	0	0
4/5 or more in agreement ..	23	35	0	0	55	47	0	14	0	0
Reliability:										
Mean % agreement	80	79	97	85	72	81	85	78	90	
Range %	58-100	54-83	75-100	71-100	58-88	71-92	71-96	67-83	75-100	
% agreement expected by chance	58	49	94	80	54	68	75	64	84	
% incidence of clinical signs for grades G, F, and P:										
Group I clinicians:										
Grade G	12	20	1	4	16	4	7	3	4	3
" F	43	54	16	18	57	61	37	57	26	6
" P	57	69	25	27	74	78	45	71	27	16
Group II clinicians:										
Grade G	6	13	0	5	14	0	2	0	3	0
" F	10	32	2	7	21	13	5	4	7	9
" P	28	54	3	21	72	48	34	42	20	27

* Whether the sign was present or absent was determined by the majority verdict of the clinicians.

TABLE IV.—*Clinical Signs Per Child According to "Nutritional" Grade*

	Grade G	Grade F	Grade P	
Group I:				
Clinician A ..	1.0	3.8	4.1	6.3*
" B ..	0.5	2.9	3.0	
" C ..	0.6	2.8	5.5	
" D ..	0.5	3.9	5.4	
" E ..	0.5	3.4	4.9	
" F ..	1.2	3.0	4.0	
Mean ..	0.7	3.5	4.5	
Group II:				
Clinician G ..	0.4	0.7	2.8	
" H ..	1.0	1.5	4.3	
" I ..	0.2	0.9	1.8	
" J ..	0.1	1.2	5.0	
Mean ..	0.4	1.1	3.5	

* This clinician used four grades.

Relationship between Nutritional Grades and Original Selection Groups

As has already been described, the Tottenham children fell into three groups; those who were "good" as judged by existing school records, those who were "medium" as judged by their existing school records, and those selected on grounds of poor nutrition. These are now termed "selection groups" and called groups "A," "B," and "C" respectively. At Tottenham the clinicians graded G a high proportion of the A children, fewer of the B children, and least of the C children (Table V). In considering these results it must be noted that special efforts were made to obtain really

TABLE V.—*"Nutritional" Grade According to Original Selection Group at Tottenham*

"Nutritional" Grade:	Group A			Group B			Group C		
	G	F	P	G	F	P	G	F	P
Group I:									
Average for all clinicians	92	7	1	81	18	1	45	40	15
Range ..	84-100	0-15	0-8	66-87	13-34	0-5	20-58	22-71	6-44
Group II:									
Average for all clinicians	65	29	6	29	62	9	4	46	50
Range ..	54-84	16-43	0-12	24-32	57-66	3-18	0-6	24-68	27-76

poor children for the C group. The Group I and II clinicians behaved somewhat differently in their gradings of the selection groups. The Group I clinicians show little difference between the A and B children, but make a sharp distinction between B and C; while the Group II clinicians differentiate more uniformly between the three groups.

The number of clinical signs per child recorded by each clinician increased from the A to the B children and still further to the C children (Table VI).

TABLE VI.—*Number of Clinical Signs Per Child According to Selection Group at Tottenham*

	Group A	Group B	Group C
Clinician A ..	0.9	1.5	2.5
" B ..	0.4	0.9	1.0
" C ..	0.4	0.8	1.7
" D ..	0.7	0.8	2.6
" E ..	0.4	0.7	1.5
" F ..	0.9	1.6	4.1
" G ..	0.5	0.8	2.1
" H ..	0.9	1.7	2.8
" I ..	1.5	2.1	2.9
" J ..	0.9	1.4	3.0

Consistency of the Grading of Groups of Children

Table VII shows the percentage of children graded "satisfactory" at Sheffield and Tottenham by four clinicians who participated in both trials. The four agree that Tottenham was a worse group than Sheffield. Moreover, their standards in relation to each other are the same in both trials, made at about a year's interval. Comparison between clinicians can also be made for selection

TABLE VII.—*Percentage of Children Graded "Satisfactory" in the Various Selection Groups at Sheffield and Tottenham*

	% of all Children Graded Satisfactory		% of Children Graded Satisfactory in the Three Selection Groups					
	Sheffield	Tottenham	Sheffield Selection Group			Tottenham Selection Group		
			A	B	C	A	B	C
Clinician A						92	87	53
" B	84	78	100	91	63	96	84	58
" C	84	74	98	88	60	96	87	44
" D						92	89	54
" E	79	65	100	87	53	83	82	58
" F						100	87	56
" G						96	87	44
" H	79	64	94	80	44	93	82	24
" I						100	89	69
" J						88	97	72
" K						—	—	—

groups A, B, and C, both at Tottenham and at Sheffield (Table VII). The clinicians record fairly small but consistent differences between groups A and B and fairly large differences between groups B and C. This suggests that not only do they distinguish between groups but they also agree fairly well on the size of the differences. It must be noted that this analysis applies to the standardized (satisfactory/unsatisfactory) grading and not to the original gradings (G, F, and P) used by the clinicians.

Discussion

The clinical assessment of nutrition has two main purposes—the assessment of the individual and the comparison of population groups. For the former, reliability and accordance are paramount; for the latter, individual discrepancies may even out to obtain a correct overall picture. For the individual, the present results confirm those of Huws Jones (1938) in this country, and of Franzen (1929) and Bean (1948) in America—namely, that the clinical assessment is unreliable. Bean (1948) concludes: "The clinical assessment of nutrition is of great complexity. Not only is the interpretation of signs and symptoms a matter for dispute but physicians trained in the discipline of nutrition vary significantly in what they actually see in examining a person."

In relation to the nutrition of population groups, two positive results emerge from the trials. Although clinicians adopt different standards for their nutritional grades, it seems that they maintain these standards relative to one another and can thus distinguish between groups of children and agree on the order of differences. This appears to be so even for examinations made at an interval of more than a year. This does not necessarily mean that clinicians maintain their standards over such a period, but that they maintain their standards relative to one another. Nor does it mean that the findings of different clinicians examining different groups of children are directly comparable.

The second positive result is the relationship between nutritional grades and clinical signs. The present result confirms those already reported by Adcock, Hammond, and Magee (1947)—namely, that there was a high degree of association between certain clinical signs and nutritional grades. The present trials go further and show that clinicians of widely different experience seem to agree substantially on the importance and relevance in relation to nutritional assessment of certain signs which have been called "grading criteria." This suggests that these grading criteria could either be used as separate factors (especially if they could be made more objective) or could be combined into a general grading according to some system of weighting. This opens up the possibility of distinguishing between degrees of severity of signs and whether the signs occur singly or in combination with others.

It must, however, be recognized that it may be as difficult to obtain standardization or objectivity for the clinical signs as for the nutritional assessment. Such an aggregate result based on signs need not, of course, be termed a nutritional assessment. In any case it seems doubtful whether it is possible, under present conditions in this country, to differentiate between nutritional assessment and an assessment of general condition or fitness. It must be emphasized that from the very nature of the clinical method perfection cannot be expected. An attempt should be made to assess its good and bad points against possible alternatives and to decide how the method can best be used and what modifications should be made.

Summary

Two trials were made to test the reliability of the clinical assessment of "nutritional state," the first at Sheffield in January, 1949, the second at Tottenham in March, 1950. Five clinicians took part at Sheffield and nine clinicians and a school nurse at Tottenham.

For each child "nutritional state" was taken to include assessments of "general condition" and the incidence of certain clinical signs.

Both for the clinical grades and for the clinical signs there was a considerable lack of agreement between clinicians in their findings for individual children and for the group of children as a whole. A number of children were examined twice, and there was, for the same clinician, considerable disagreement in the findings of the two examinations.

There was a high degree of association between clinical grade and the incidence of certain clinical signs.

Comparison of the clinical assessments made by those clinicians who took part in both trials and the assessments by the different clinicians of different groups of children at Tottenham shows that the clinicians maintained their standards relative to one another and that they could distinguish between groups and generally agree on the order of the differences.

The high degree of association between nutritional grades and the incidence of certain clinical signs suggest that if the latter could be made more objective they could be used separately for grading purposes or could be combined into a general grading according to some system of weighting.

The trials at Sheffield and Tottenham were entirely dependent on the participation of the clinicians and the school nurse. We are especially glad to acknowledge our indebtedness to them and to express our thanks. We wish to make it clear, however, that this paper represents our interpretation of the data and that only our personal views are expressed. This does not imply that those participating in the trials agree with the interpretation of the

data or with the views expressed. We wish to record again our appreciation of the valued co-operation of Dr. H. M. Cohen, at Sheffield, and Dr. G. H. Hogben and Dr. P. A. Tyser, at Tottenham. We are also glad to thank Dr. H. E. Magee for his interest and advice.

REFERENCES

- Adcock, E. W., Hammond, W. H., and Magee, H. E. (1947). *J. Hyg., Camb.*, **45**, 65.
 Bean, W. B. (1948). *J. appl. Physiol.*, **1**, 458.
 Franzen, R. (1929). *Physical Measurements of Growth and Nutrition*. Amer. Child Hlth Ass., New York.
 Jones, R. H. (1938). *J. R. statist. Soc.*, **101**, 1.

THE PERIPHERAL CIRCULATION DURING GENERAL ANAESTHESIA AND SURGERY

BY

R. B. LYNN, M.D., F.R.C.S.

AND

RALPH SHACKMAN, M.B., B.S., F.R.C.S.

(From the Department of Surgery, Postgraduate Medical School of London)

Changes produced in the peripheral circulation by general anaesthesia may be evident in the flushed warm skin which develops soon after induction, and may be recorded by finger plethysmography (Foster *et al.*, 1945). We have used calf, foot, and hand plethysmographs to measure the circulatory changes which occur when cyclopropane, ether, "trilene," nitrous oxide, oxygen, thiopentone, and gallamine triethiodide ("flaxedil") are used in varying combinations (Tables I and II), and have shown that further readjustment of the circulation develops later during lengthy major operations. The state of the limb circulation can be a considerable aid to the assessment of the general circulatory state of a patient undergoing major surgery under general anaesthesia.

Method of Study

The plethysmographs (Barcroft and Edholm, 1943) were constructed of "perspex" and contained built-in thermostatically controlled heating units. Displacement of water (kept at $32^{\circ} \pm 1^{\circ}$ C. for the hand and foot, and at $34^{\circ} \pm 1^{\circ}$ C. for the calf) was used to measure the limb flows. Snugly fitting diaphragms were bolted to the plethysmograph, and the cuffs, made of thin rubber sheeting, were sealed to the skin with rubber cement. Several turns of self-adhering rubber tape were wound around the cuffs as reinforcement, with care to avoid venous congestion. Thick rubber pressure tubing connected the plethysmograph to a float recorder writing on a mobile bench kymograph.

When calf flows were measured egress of blood from the foot into the calf was interrupted by inflation of an occluding cuff at the ankle above systolic pressure (Barcroft and Edholm, 1943).

Limbs flows were measured, first in the pre-anaesthetic sedated period before the induction of anaesthesia, and then throughout the operative procedure (pre-operative sedation was by morphine, atropine, and hyoscine in varying combinations). Several readings were taken during every five-minute period, and the average value was used as the index of limb flow. The blood pressure and pulse rate were recorded at regular periods throughout the studies.