

Design and Accuracy of Calipers for Measuring Subcutaneous Tissue Thickness*

BY D. A. W. EDWARDS†

Department of Clinical Research, University College Hospital Medical School

W. H. HAMMOND

Ministry of Health, 23 Savile Row, London, W. 1

M. J. R. HEALY

Rothamsted Experimental Station, Harpenden, Herts

J. M. TANNER AND R. H. WHITEHOUSE

*Sherrington School of Physiology, St Thomas's Hospital Medical School,
London S.E. 1*

(Received 11 August 1954)

A considerable proportion of the body fat lies in the subcutaneous tissue which, in many parts of the body, is only loosely attached to the underlying tissue and can be pulled up between the thumb and forefinger into a fold (Pl. 1). The thickness of this fold of skin and subcutaneous tissues can be measured by applying some form of caliper to either side of it, and this technique has been widely used to give an estimate of body fat by those concerned with nutrition (Sinclair, 1948; Keys & Brožek, 1953), with fat distribution (Edwards, 1950), with child growth (Franzen, 1929; Reynolds, 1950; Tanner, 1953) and with anthropometric surveys (Berry, Cowin & Magee, 1951; Hammond, 1953). However, the observed thickness depends on how the skinfold is picked up and on the design of the caliper with which it is measured; if, for example, the jaws of the caliper are approximated by a spring the reading will depend on the strength of the spring, since the tissue is compressible. Despite these facts, there has been little information published on the best design of skinfold calipers, and no general standardization either of instrument or of technique. Hence the comparison of the results of different observers is usually impossible, and much work in the field is vitiated.

The purpose of the present investigation was (a) to test various designs of caliper and to recommend principles that all calipers should follow, and (b) to introduce, and test the accuracy of, a new skinfold caliper which we believe to be the most satisfactory yet produced. The paper is divided into two parts: Part 1 describes two experiments on caliper design and ends with a set of general recommendations; Part 2 describes the new caliper and gives the results of three experiments made to assess its accuracy in practical use.

* This investigation was carried out at the request of the Medical Research Council's Committee on Growth and Form.

† Working on behalf of the Medical Research Council.

EXPERIMENTAL AND RESULTS

General methods

Different caliper designs can be assessed by the consistency of repeated measurements on the same individual. This assessment has two aspects, namely the consistency of repeated measurements by a single observer and the agreement between the measurements of different observers. The experiments were designed to provide information on both aspects and all followed the same general pattern, whereby a number of observers took duplicate measurements with one or more types of caliper on each of a number of subjects. The total number of observations was kept within practicable limits by the use of incomplete block designs where appropriate. The results of each experiment were treated by standard statistical techniques to provide estimated standard deviations of the differences between duplicate readings taken (*a*) by a single observer or (*b*) by different observers. In (*b*), the actual observers are regarded as a random sample of all possible observers.

Presentation of results. One main difficulty in the design of skinfold calipers is to obtain constant, or nearly constant, pressure at all jaw openings from 2 to 40 mm, the working range. All the calipers of the Franzen (1929) type produce a large increase of pressure as the spring is stretched by increasing the opening of the jaws, and this is to be discouraged, as it leads to difficulties of standardization. The calipers used in Exps. 1 and 2 suffer from this disadvantage, and for this reason we have compared their spring tensions at 10 mm jaw opening only.

The frequency distribution of most skinfold measurements in the general population is considerably skew, with a long tail of high readings. In addition, duplicate measurements at the same site by the same observer agree more closely for small than for large measurements, the measuring error increasing roughly in proportion to the size of the measurement. For both these reasons it is desirable to transform the recorded measurements to a logarithmic scale before analysis. The transformation used in Exps. 1 and 2 was $100 \log_{10}$ (skinfold thickness in mm $- 2.3$). It was chosen after examination of the readings obtained in another study at four sites (over biceps, over triceps, below inferior angle of scapula, and above iliac crest) on some 600 healthy young men aged 18–30, and seemed to make all four distributions approximately Gaussian. In Exps. 1 and 2 of the present series this transformation also made the variability between duplicates roughly constant over the whole scale. The value of 2.3 mm for the constant term is approximately the thickness of the two layers of skin included in the fold when the caliper spring tension is of the order of from 2.5 to 5 g/mm². It was found later that, at higher pressures, the value of the constant has to be decreased, as done in the experiments of Part 2.

Part 1. Caliper design

First experiment. The first experiment was designed to test the effect of different pressures, all high relative to current practice. Four observers took duplicate measurements on twelve subjects at a site at the side of the knee, just over the head of the

fibula. Three Franzen (1929) type calipers were used and measurements were recorded to the nearest 0.1 mm. The standard deviations of the differences between duplicates are given in Table 1 together with the characteristics of the calipers. The 'single-observer' figure is the average for the four observers, who, in this experiment, were all without previous experience of the calipers, except for practice immediately before the experiment. The agreement between single-observer duplicates was best for caliper 3, the pair with the lowest (15 g/mm²) pressure. However, for all calipers the agreement between different observers was worse than that between duplicate readings taken by a single observer. There was little to choose between the three calipers when differences between observers are taken into account. Caliper 2 with 26 g/mm² pressure gave mean readings consistently lower than caliper 3, with 15 g/mm², with an average difference, approximately constant over the whole scale, of 7.3 logarithmic units. Further increase of pressure over 26 g/mm² had very little effect, calipers 1 and 2 giving substantially identical readings.

Table 1. *Exp. 1. Effects on consistency of Franzen-type caliper measurements of different combinations of face area and pressure. Standard deviations of differences between duplicate readings (in logarithmic scale) by four observers, on twelve subjects*

Face area (mm ²)	60	90	90
Pressure at 10 mm jaw opening (g/mm ²)	33	26	15
Standard deviation:			
Measurements by one observer	6.81	6.73	3.41
Measurements by different observers	8.15	9.97	8.57

Second experiment. The second experiment tested all combinations of three pressures (covering a lower range than in Exp. 1) and two face areas. There were six observers, six pairs of calipers and ten subjects. Measurements were taken to the nearest millimetre at sites marked on each subject by ink marks (a) half-way down the back of the arm over the triceps, and (b) immediately below the inferior angle of the left scapula. Some time elapsed and many other readings were taken between duplicates on the same subject so that there was little chance of remembering previous readings. The six calipers used were as follows:

	Pressure at 10 mm jaw opening (g/mm ²)
80	2.3
	9
	19
120	2.3
	9
	19

Each observer used each caliper twice on each of five subjects. The standard deviations of differences between duplicate readings are given in Table 2. The area of the face made little difference to the result, so that the important values in Table 2 are the means for the two face areas. It can be seen that for the triceps measurement the results that could be repeated with the greatest accuracy, both by a single observer and by different observers, were given by the strongest pressure, 19 g/mm². For the

subscapular measurement there was little difference between the results when using 19 and 9 g/mm², but the 2.3 g/mm² caliper was again the worst. As in Exp. 1, there were substantial observer biases and subject-observer interactions for both measurements at all spring tensions, despite marking the skinfold sites. That is to say, some observers habitually gave relatively high, others relatively low, readings, but not to the same extent on each subject. Three of the observers were experienced in taking the measurements and three were not, but there was no significant difference in variability between experienced and inexperienced observers.

Table 2. *Exp. 2. Effects on consistency of Franzen-type caliper measurements of different combinations of face area and pressure. Standard deviations of differences between duplicate readings (in logarithmic scale) by six observers, on ten subjects*

Pressure (g/mm ²)	Standard deviation for triceps at face area of			Standard deviation for subscapular at face area of		
	80 mm ²	120 mm ²	Mean	80 mm ²	120 mm ²	Mean
Measurements by one observer						
2.3	6.95	6.13	6.54	5.22	3.64	4.43
9.0	6.56	7.43	7.00	3.97	4.15	4.06
19.0	4.29	4.20	4.25	4.73	4.57	4.65
Measurements by different observers						
2.3	15.70	8.20	11.95	9.54	8.74	9.14
9.0	8.27	8.12	8.20	6.76	6.94	6.85
19.0	4.41	5.84	5.12	6.64	6.35	6.50

The difference between the readings given by the 9 g/mm² caliper and the 19 g/mm² caliper was constant in the log scale at 11 units. The difference between the 9 g/mm² caliper and the weak caliper giving 2.3 g/mm² pressure averaged about 18 logarithmic units but decreased slightly as the measurement increased.

Conclusions and recommendations. These results indicate that the pressure exerted by the caliper has an important effect not only on the observed thickness of the fold, but also on the consistency with which the measurements can be repeated. The effect of the face area is negligible by comparison. The results of Exp. 1 show that 15 g/mm² is better than 26 or 33 g/mm², and those of Exp. 2 show that there is little to choose between pressures of 9 and 19 g/mm², but that a pressure of 2.3 g/mm² gives results which are not reproducible with accuracy. It seems, therefore, that the best working pressure lies within the range 9–20 g/mm². A change in pressure within this range has a marked effect on the observed thickness of the fold, so that provision of a caliper producing constant, or nearly constant, pressure at all jaw openings becomes a point of first importance. The precise pressure used will depend upon several considerations, one being the discomfort produced by the caliper squeezing the fold.

Since many existing calipers of Franzen (1929) type have face areas of 90 mm² and there seems no reason to depart from this figure, the following recommendations for the design of skinfold calipers are put forward.

(1) The faces of the calipers should be rectangular of size 6 × 15 mm, with well-rounded edges and corners.

(2) The pressure exerted at the faces should not vary by more than 2.0 g/mm² over the range of openings 2–40 mm.

(3) For results which can be reproduced with accuracy, the pressure should lie in the range 9–20 g/mm². The effect on the readings of a change in pressure is smaller at high pressures. A pressure of 15 g/mm² is probably as much as can be tolerated without complaint by the subject, and 10 g/mm² is suggested as a standard value. Some workers (including one of the present authors) feel that lower pressures should be used for small children; if this is done, the pressure should be stated in presenting the results, and, if possible, calibration trials against standard-pressure calipers should be carried out.

(4) The scale of the instrument should be such that readings can be taken at least to the nearest 0.5 mm and preferably to the nearest 0.1 mm. An accuracy of this order in repeated measurements is readily obtainable for readings below about 10 mm.

Part 2. Accuracy of the recommended caliper

The new caliper. At about the time the recommendations above were being formulated a caliper of a design entirely different from the Franzen type was introduced for trial at the Harpenden Growth Study (Tanner, 1953). This caliper, shown in use in Pl. 1, is an adaptation of an instrument used in industry for measuring thickness of pieces of plywood, metal or leather, and is available commercially from British Indicators Ltd, Sutton Road, St Albans, Herts. It has a dial easily read to 0.1 mm and with faces fitted and a suitable alteration made in the type and position of the springs it meets all the recommendations of Part 1. Accordingly, three trials were carried out using this caliper.

Third experiment. Three observers (two experienced) carried out duplicate measurements at each of four sites: those used in Exp. 2 (triceps and subscapular), a point half-way down the front of the upper arm (biceps) and a point in the anterior axillary line immediately above the level of the left iliac crest (suprailiac). The sites were marked, and one caliper, which gave a pressure of 12.5–14.5 g/mm² over the range 2–40 mm was used throughout. The subjects were thirty-five children of both sexes aged 3–18 years.

In this and the following experiments the constant in the transformation was altered to 1.8. It is convenient in practice to use the transformation in the form $100 \log_{10} (\text{reading in } 0.1 \text{ mm} - 18)$ to avoid negative values. A table of this transformation is given in Table 3. The standard deviations of differences between duplicates for the three observers are given in Table 4.

These results represent a great improvement over the figures obtained with the earlier type of caliper; the standard deviations for triceps and subscapular are approximately half of the best values obtained in Exp. 2. The distribution of differences between duplicates for each of the three observers (numbered as in Exp. 4) is shown in Fig. 1 on a millimetre scale. It can be seen that with this caliper readings really can be taken to 0.1 mm; indeed, differences as large as 1.0 mm, which is the minimum difference over zero readable on the Franzen caliper, occur only occasionally.

Fourth experiment. This experiment was undertaken to provide a more comprehensive trial of the new caliper. Thirty men, between the ages of 18 and 45, with physiques varying from very thin to fairly fat, were measured by six observers. Observers 1 and 2 had very little experience of this type of measurement, observers 3 and 4 had con-

Table 3. *Table of log transformation of skinfold measurements*

(Transformation = $100 \log_{10}$ (reading in 0.1 mm - 18))

mm	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	30	48	60	70	78	85	90	95	100	104
3	108	111	115	118	120	123	126	128	130	132
4	134	136	138	140	141	143	145	146	148	149
5	151	152	153	154	156	157	158	159	160	161
6	162	163	164	165	166	167	168	169	170	171
7	172	173	174	175	176	176	176	177	178	179
8	179	180	181	181	182	183	183	184	185	185
9	186	186	187	188	188	189	189	190	190	191
10	191	192	192	193	193	194	194	195	195	196
11	196	197	197	198	198	199	199	200	200	200
12	201	201	202	202	203	203	203	204	204	205
13	205	205	206	206	206	207	207	208	208	208
14	209	209	209	210	210	210	211	211	211	212
15	212	212	213	213	213	214	214	214	215	215
16	215	216	216	216	216	217	217	217	218	218
17	218	218	219	219	219	220	220	220	220	221
18	221	221	221	222	222	222	223	223	223	223
19	224	224	224	224	225	225	225	225	226	226
20	226	226	226	227	227	227	227	228	228	228
21	228	229	229	229	229	229	230	230	230	230
22	231	231	231	231	231	232	232	232	232	232
23	233	233	233	233	233	234	234	234	234	234
24	235	235	235	235	235	236	236	236	236	236
25	237	237	237	237	237	237	238	238	238	238
26	238	239	239	239	239	239	239	240	240	240
27	240	240	240	241	241	241	241	241	241	242
28	242	242	242	242	242	243	243	243	243	243
29	243	244	244	244	244	244	244	245	245	245
30	245	245	245	245	246	246	246	246	246	246
31	247	247	247	247	247	247	248	247	248	248
32	248	248	248	248	249	249	249	249	249	249
33	249	250	250	250	250	250	250	250	251	251
34	251	251	251	251	251	251	252	252	252	252
35	252	252	252	253	253	253	253	253	253	253
36	253	254	254	254	254	254	254	254	254	255
37	255	255	255	255	255	255	255	256	256	256
38	256	256	256	256	256	256	257	257	257	257
39	257	257	257	257	258	258	258	258	258	258

Table 4. *Exp. 3. Consistency of measurements with new dial-type caliper. Standard deviations of differences between duplicate readings (in logarithmic scale) by three observers on thirty-five subjects*

	Biceps	Triceps	Subscapular	Suprailiac
Measurements by one observer	7.61	2.42	2.43	4.85
Measurements by different observers	9.17	3.48	3.50	5.20

siderable experience of skinfold measurements with other calipers but had not previously used the new type, and observers 5 and 6 had experience both with the older and with the new type of caliper. Three pairs of calipers were used, in theory identical, but in practice giving slightly different pressures. The same four sites as in Exp. 3 were marked and measured on each subject. Measurements were recorded to 0.1 mm and transformed before analysis to the form

$$z = 100 \log_{10} (\text{reading in } 0.1 \text{ mm} - 18).$$

Each observer measured ten subjects, taking duplicate measurements at each site with each of the three calipers. The arrangements were such that the possibility of

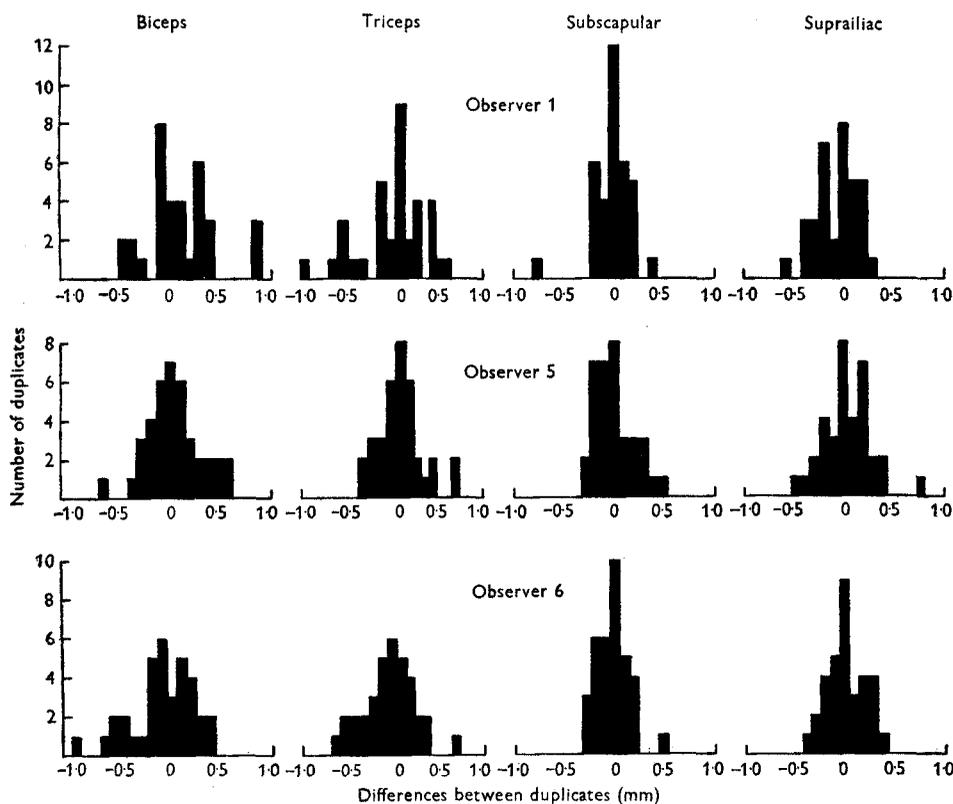


Fig. 1. Distributions of differences between duplicate readings with the recommended caliper of skinfold thickness by three observers at four sites on thirty-five subjects each. The scale is straight millimetres, not logarithmic values.

remembering more than an occasional reading was minimal. The analyses of variance given in Table 5 provide tests of the average difference between calipers and of whether these differences vary with the subject or with the observer. They also test the presence of observer biases, and whether these vary from subject to subject (observer-subject interaction).

The three calipers gave slightly different average readings, but there is no evidence that these differences were affected by the subject or the observer. It is rather difficult, on our present evidence, to give a figure for the relation of average reading to caliper

Table 6 also shows the standard deviations of differences between duplicates for each observer separately, and the mean measurements in the log scale with the corresponding figure in the untransformed scale. The means differ most for the biceps and least for the triceps and subscapular measurements. There appears to be no consistent difference between observers in their standard deviations of duplicates, and the least experienced do not seem to score worse than the others.

Fifth experiment. In view of the radical change in caliper design, an experiment similar to Exp. 2 was carried out to estimate the errors of measurement of the new caliper at three pressures, 6, 10 and 13 g/mm². The four sites used were biceps, triceps and subscapular (as in the previous experiments) and subcostal, at a point immediately under the costal margin, vertically below the nipple. The standard deviations of differences in the log scale are given in Table 7.

Table 7. *Exp. 5. Effects on consistency of dial-type caliper measurements of different pressures exerted at the caliper faces. Standard deviations of differences between duplicate readings (in logarithmic scale) by six observers on ten subjects*

Site	Standard deviation at pressure of		
	6 g/mm ²	10 g/mm ²	13 g/mm ²
Measurements by one observer			
Biceps	12.49	9.26	7.57
Triceps	4.64	3.72	2.55
Subscapular	3.14	3.74	2.16
Subcostal	3.79	3.38	3.11
Measurements by different observers			
Biceps	21.70	22.84	21.14
Triceps	5.74	4.16	4.10
Subscapular	7.97	5.45	4.32
Subcostal	6.71	5.73	6.38

As expected from previous work, the standard deviations decreased with increasing tension, but the difference in accuracy between the 10 and 13 g/mm² calipers was quite small. In this experiment four of the six observers had considerable previous experience of skinfold measurements, three of them with the new caliper, but the other two had little experience of practical anthropometry; the agreement between duplicates was markedly better for the experienced observers.

DISCUSSION

The new caliper measures skinfolds with an accuracy quite sufficient for any presently conceivable purpose. At jaw openings of about 7 mm, the standard deviation of differences between duplicates for a single observer is of the order of 0.3–0.6 mm, and the standard deviation of differences between readings by different observers is of the order of 0.4–1.3 mm. When this caliper is used, the limitations of accuracy are transferred from the instrument to the locating of the skinfold and the manner in which it is picked up. These limitations are analogous to those encountered in

classical skeletal anthropometry and can be overcome only by rigorous specification of technique and training of observers.

Experience has shown that variation from time to time in the average tension exerted by the caliper springs supplied by the manufacturers is unlikely to alter the pressure exerted at the faces by more than 1 g/mm², and this should not affect the readings by more than 0.2 mm except in very fat subjects. For precision work, however, it may be desirable to calibrate the instruments at intervals. This can be done by fixing one arm horizontally in a clamp and suspending a pan with weights from the other. The jaw opening at a series of weights from 900 g upwards in 10 g intervals can then be recorded.

There is no doubt that the accuracy with which a skinfold may be measured varies from one site to another; some folds are more easily picked up, and some vary less in thickness over a small area so that location is less critical. We have made no exhaustive study of this, but our experience suggests that the biceps site is a poor one from this point of view. The triceps, subscapular and supriliac sites seem from the literature and our experience to be fairly good; the supriliac is the poorest of the three. The choice of sites must depend principally on the purpose for which the measurements are undertaken; if it is to classify varieties of physique or of fat distribution, many sites may be necessary, but if it is to estimate the total body-fat content, then possibly a few only would suffice (Brožek & Keys, 1951). Criteria for choice of site based on adequate experimental data are the next requirement in the field of skinfold measurement.

SUMMARY

1. The technique of measuring with calipers the folds of skin plus subcutaneous tissue that can be picked up between the thumb and forefinger and pulled away from the underlying tissue at various sites in man has been much used, but the reading obtained depends on the characteristics of the calipers. These characteristics have never been properly standardized. For this reason experiments were carried out to determine the optimal design of skinfold calipers (Part 1).

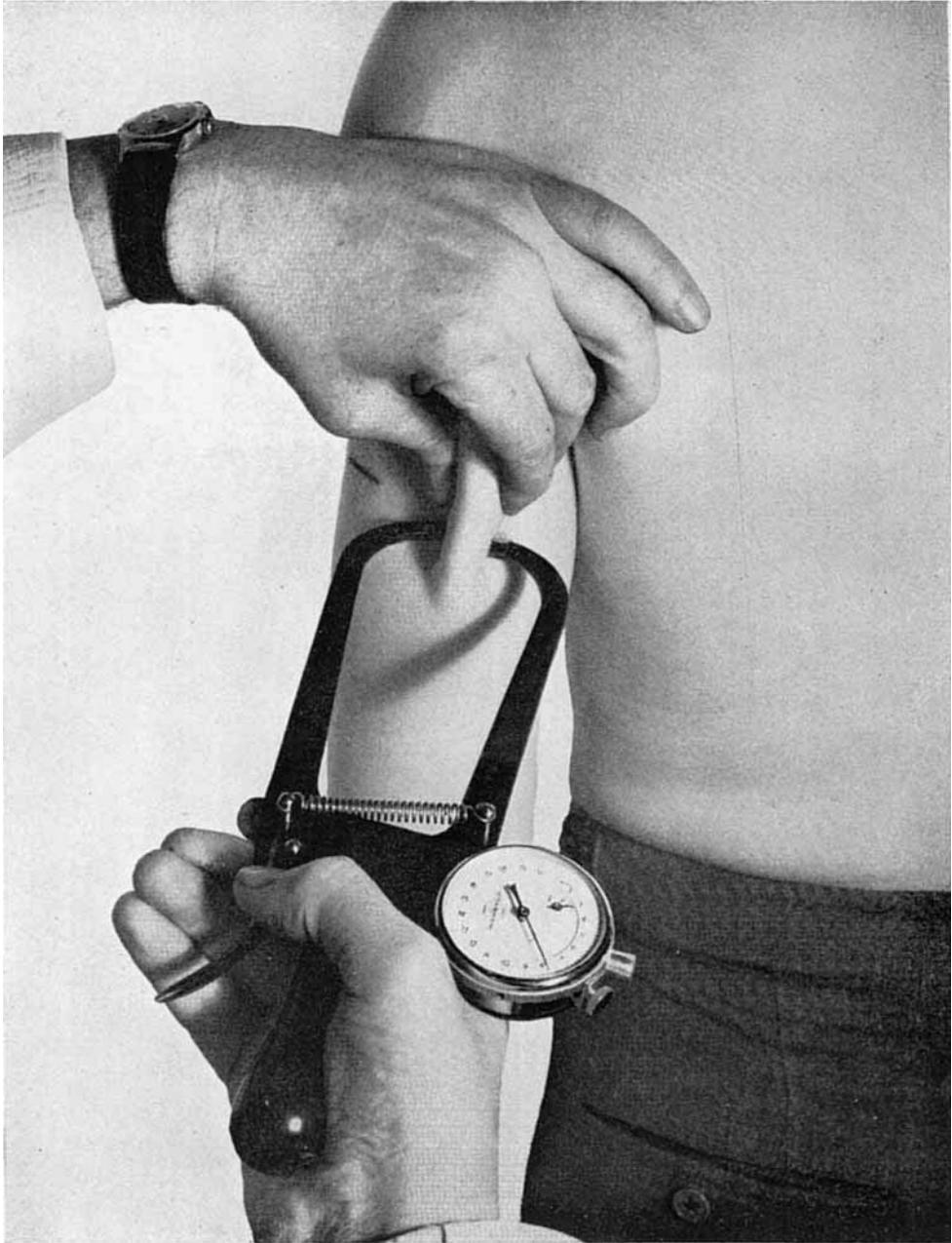
2. As a result of these experiments, recommendations were formulated for a caliper to give maximum consistency between duplicate readings. These were: that the face area of skinfold calipers should be 6 × 15 mm; that the spring pressure exerted over the range of opening 2–40 mm should not vary by more than 2.0 g/mm²; that the pressure should be between the limits of 9–15 g/mm² with a recommended standard value of 10 g/mm²; that the scale of the instrument should be read at least to 0.5 mm and preferably to 0.1 mm.

3. A new type of caliper was introduced (Part 2) which meets these requirements and which in our experience is the most satisfactory caliper yet devised.

4. Since the differences between duplicate readings were proportional to the mean value, the measurements were analysed after transformation to the scale $z = 100 \log_{10}$ (reading in 0.1 mm - 18), a table of which is given for practical use.

5. With the new calipers, the standard deviations of the differences between duplicate measurements taken by a single observer at the triceps, subscapular and

D. A. W. EDWARDS AND OTHERS. DESIGN AND ACCURACY OF CALIPERS FOR MEASURING SUBCUTANEOUS TISSUE THICKNESS *Plate I*



suprailiac sites were 2-4 units in the logarithmic scale, equivalent to 0.3-0.6 mm at a jaw opening of 7 mm. The equivalent figures for different observers were roughly twice these values when the sites for measurement were marked on the body.

6. Observers tended to give slight but consistent differences in readings on any one subject, and these differences varied to a small extent from one subject to another.

7. The choice of sites to be measured for specific problems has not been considered in this paper.

We wish to thank Professor D. V. Davies, Professor J. Z. Young, Colonel J. E. Snow, O.B.E., Dr G. M. Morant and their staffs for their kind co-operation and the provision of facilities; also the Medical Research Council for a grant to defray expenses.

REFERENCES

- Berry, W. T. C., Cowin, P. J. & Magee, H. E. (1951). *Mon. Bull. Minist. Hlth Lab. Serv.* **10**, 155.
Brožek, J. & Keys, A. (1951). *Brit. J. Nutr.* **5**, 194.
Edwards, D. A. W. (1950). *Clin. Sci.* **9**, 259.
Franzen, R. (1929). *Physical Measures of Growth and Nutrition*. New York: American Child Health Association.
Hammond, W. H. (1953). *Hum. Biol.* **25**, 65.
Keys, A. & Brožek, J. (1953). *Physiol. Rev.* **33**, 245.
Reynolds, E. L. (1950). *Monogr. Soc. Res. Child Developm.* **15**, no. 2.
Sinclair, H. M. (1948). *Vitam. & Horm.* **6**, 101.
Tanner, J. M. (1953). *Lect. Sci. Basis Med.* **1**, 308.

EXPLANATION OF PLATE

Measurement of skinfold over triceps with new caliper. The ink mark is placed at a level midway between the tip of the acromion and the head of the radius.