PHYSICAL MEASUREMENT OF YOUNG CHILDREN

By

Virginia Bergstresser Knott

UNIVERSITY OF IOWA STUDIES: STUDIES IN CHILD WELFARE

VOLUME XVIII (No. 394) NUMBER 3

PHYSICAL MEASUREMENT OF YOUNG CHILDREN: A STUDY OF ANTHROPOMETRIC RELIABILITIES FOR CHILDREN THREE TO SIX YEARS OF AGE

By

VIRGINIA BERGSTRESSER KNOTT, Ph.D.

GEORGE D. STODDARD, Ph.D., Editor

University of Iowa Studies
Studies in Child Welfare

Volume XVIII, No. 3

Published by the University of Iowa Press
Iowa City, Iowa
1941

FOREWORD

Dr. Knott's review of the literature in Chapter II of this volume reveals that, for ages three to six years, a systematic study of the reliability of anthropometric measurements has not been previously undertaken. It is known that this lack of knowledge on reliability has resulted in lessened work efficiency.

The present data are based on thirty-five physical measurements for a good sampling of preschool children. Differences in the observations of two technicians are analyzed and, for some measurements, between two observations by the same technician.

Of special interest are Dr. Knott's estimates of the approximate frequency of measurement considered profitable during the preschool years: stature, ten times a year; hip width, sitting height, leg girth, and total arm length, three to four times a year; head, chest depth and elbow width, every two years; and the remaining measurements chiefly once a year.

The study, as a whole, forms a unit in the anthropometric program being carried on in the Station by Dr. Howard V. Meredith.

George D. Stoddard

Office of the Director
Iowa Child Welfare Research Station
University of Iowa
December 2, 1940

CONTENTS

CHAPTER I

THE RELATION OF ANTHROPOMETRIC RELIABILITIES TO THE SCIENTIFIC STUDY OF PHYSICAL GROWTH

Early studies of prenatal and postnatal growth of the body employed data primarily representing "over-all" measurements: stature, surface area, weight. With the development of research on physical growth an increasing number of segmental measurements was proposed and investigated. Recently, data on as many as seventy different dimensions of the external body were analyzed in a single study (34).

Study of this development indicates that investigators have rather freely developed new measurements to meet their specific research needs and that they have tended not to question new dimensions and procedures recommended by others. Surprisingly few rigorous studies of the techniques themselves, with a view to obtaining objective evidence of the advantages or limitations, have appeared.

The purpose of this investigation has been to determine experimentally the reliability of observations for some thirty different bodily dimensions.

IMPORTANCE OF STUDY OF RELIABILITY

Previous research on the problem of anthropometric reliabilities is assumed to be markedly insufficient. Such an assumption must be substantiated from a critical review of the available research. This will be undertaken in Chapter II.

Here it is pertinent to examine the more fundamental question of why the reliability of physical measurements need be studied at all. It is the contention of the writer that research on anthropometric reliabilities is necessary (1) in order to evaluate certain current yet conflicting assumptions in this area and (2) in order to facilitate the solution of certain research problems relating to the study of the physical growth of the child.

Current Assumptions

Certain assumptions and generalizations warrant critical examination.

1. There appears in much of the literature on physical growth the assumption, usually implicit, that measurements are reliable.

It is not amiss to venture a comment on the question of why more frequent attention has not been given to the subject of reliability. It may be because anthropometry is commonly defined as a quantitative study of the human body. Perchance, the inclusion of the word "quantitative" has been too frequently regarded to connotate high precision. Again, the fact that finely scaled instruments are employed appears to have given rise to the unfounded belief that results obtained by application of these instruments are of necessity almost equally accurate.

2. Certain explicit generalizations, often contradictory to one another, are contained in the literature.

Here, while there is recognition of the fact that complete disregard of the question of reliability is unsound, there is a shift to the opposite extreme of broad generalization.

The statement that "a discrepancy of 10 mm. may easily occur in any measurement upon the living body" (38, p. 25) needs critical evaluation. Small dimensions such as nose width or elbow width, even if taken by unskilled technicians, would rarely, if ever, vary by this amount. Yet according to the above statement, such discrepancies might easily occur.

Obviously, instead of the tacit assumption that anthropometric measurements are either highly reliable or comparatively unreliable, what is needed is a careful experimental approach—an approach that will eventually yield reliabilities, specific for the various proposed procedures and for successive segments of the age scale on the complete series of measurements employed.

Research Needs

Solution of technical problems in the establishment of research programs and the standardization of technique depend, in part, on reliability of findings.

1. In certain research situations there is need for information on the possible effect of measurement errors.

It cannot always be assumed that in cross-sectional studies the effect of measurement errors is relatively negligible. Such an assumption is

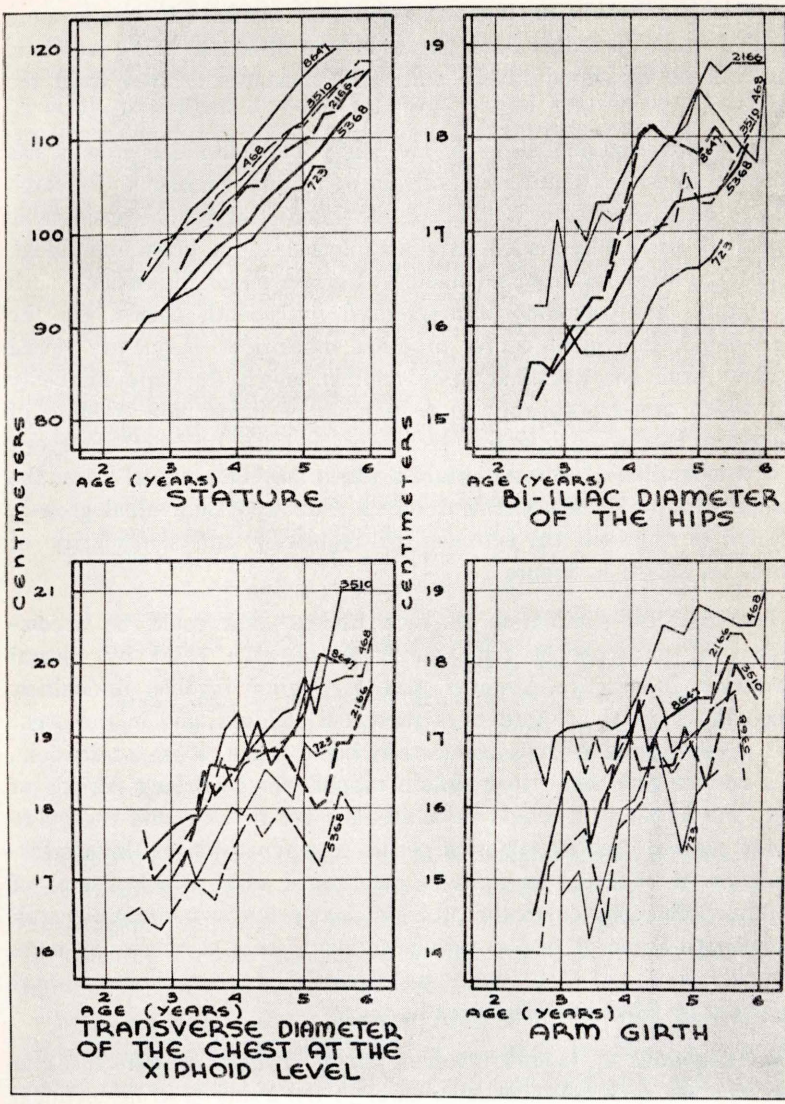

Figure 1. Individual Curves Illustrating Regularity and Irregularity of Trend Lines: Based on Data From the Files of the Iowa Child Welfare Research Station

only valid if the data have been gathered by a well-trained observer, whose errors are completely random. The measures of central tendency and variability of such data on an adequate sample will not be appreciably affected. But if measurements are obtained by one or more persons whose techniques are biased in one direction, this systematic bias will be registered in the statistical constants derived from the data.

The increasing emphasis on the study of individual growth has recently focused attention on certain reliability problems. Research centers have been collecting longitudinal material for a considerable period of time. When such data are plotted for a given individual, results are often far from satisfactory. Certain major dimensions, such as stature, are commonly characterized by smooth trends. On the other hand, individual curves of other dimensions frequently reveal changes from one examination to another which fluctuate markedly and which appear probably as a function of inadequate training of observers, lack of comparable technique between different observers, or a combination of several variables which may be grouped under the general heading of measurement errors. Curves of individual growth, selected to represent the extremes of regularity and irregularity of trend, are shown in Figure 1.

Theoretically, elimination of such useless data could be accomplished by independent remeasurements of each child by several examiners. In practice, however, such a program involves difficulties. When the subjects are infants or preschool children, and in the event that a large series of measurements is being taken, it is not feasible, as a routine procedure, to maintain rapport for such long periods of time. Furthermore, it would be a needless use of time and energy to devote part of each examination period to a repetition of those measurements which could be initially determined with a high degree of accuracy. The solution to the problem clearly lies in investigations on the relative accuracy of the various measurements at successive ages. From the results of such studies, a differential program similar to that suggested by Meredith (26) may be set up.

Investigations of factors involving small changes or differences in size, whether they be undertaken through the study of cross-sectional or longitudinal data, necessitate some mention of, or reference to, reliability. Studies attempting to show the effect of different diets or of different seasons of the year on growth should certainly give some indication that the resulting "statistically significant differences" cannot be attributed to significant differences in technique rather than to valid differences due to the factors under study.

2. In addition to the importance of errors as they relate to various research problems, reliability data are of prime concern in the area of standardization of measurement technique. Investigations on reliability are needed in order to assist: (a) in the choice of anthropometric method and (b) in the training of anthropometrists.

The Geneva Agreement (19) stands as the only international authority concerning the manner in which measurements of the living human body are to be taken. The specifications found therein were decided upon in 1912 by a committee of persons experienced in such work. Today this agreement is acknowledged to be deficient in numerous respects. Before a new agreement is drawn up, studies on the relative reliability of measuring certain portions of the body by alternative techniques should be made.

This need has been indicated recently by Hrdlička (20, p. 293-294).

> "Within the last 10 years the consciousness of needed anthropometric agreements has gradually returned, and new attempts have begun toward regulation of anthropometric procedures. Regrettably the attempts are not as yet fully harmonious; they have not yet reached a definite program; . . . and there is a tendency toward a greater concern for advanced analytical procedures than for the fundamental mechanical and other simpler but valuable practices. . . . In respect to other items, such as nomenclature, the precision of landmarks, etc., it will be equally evident that there will be need of critical studies on ample materials. . . . They will be directly useful . . . to anthropometry in general. Any duplication in such studies here, or here and abroad, will only enhance their value."

The present study, in addition to furnishing some material on relative reliabilities, gives a foundation for additional work in the future. For example, knowing the accuracy with which a given dimension can be taken by a given procedure, alternative procedures may be explored and the accuracy of these compared.

Referring again to the Geneva Agreement, we find here no clue concerning the specific standards of accuracy to which a new technician should be trained. Should the standard be that of exact duplication of the measurements of an experienced worker? Or, would one be sufficiently well-trained if one's discrepancies did not exceed 10 mm., for example, in any measurement? Exact duplication is obviously too rigid a requirement. For some dimensions, reliability to within 10 mm. may be much too lenient, whereas for other dimensions it may be very difficult if not impossible to obtain. Thus, the reasonable expected limit of accuracy may vary with the type of dimension and the age of the subject being measured. The problem then becomes one of discovering what these limits are for the various types of measurements on children of different ages. The present study affords such limits when measurements are made on subjects of preschool age.

SOURCES OF ERROR

The reliability of anthropometric observations may be affected by two general types of factors, one of which might be designated as "accidental" and the other as "technical."

Accidental errors may occur in two ways. The scale of the instrument may be misread, or the value obtained may be recorded incorrectly. One would expect these types of errors to be committed but rarely by careful workers. The present study is concerned, rather, with the technical errors.

Four sources of errors relating to technical difficulties may be distinguished:

First, error may arise from the attempt to follow a definition which is not sufficiently precise in statement concerning the location of landmarks, the position of the subject, or application of the instrument. Even in a reasonably adequate definition, slight differences in interpretation may appear.

Second, given adequate definitions, difficulty sometimes may arise in locating and applying the instruments to the landmark to be used. The location of radiale and stylion ulnare are such examples.

Third, although the landmarks may be located and the instrument applied, slight variations in the slant of the instrument or in the pressure exerted (depending on the dimension being measured) may occur.

Finally, while the position of the subject may be defined very precisely (and the careful technician is concerned with watching to see that this position is maintained in the determination), the position of the subject can scarcely be assumed to be exactly identical each time measurement is attempted.

TERMINOLOGY

Reliability refers to the comparison of independent measurements repeated on the same subject within a short time interval. The interval in this study was approximately one hour.

There are two alternative methods of studying reliability: either the same person may repeat the measurement, if the factor of memory is eliminated, or several different observers may take the measurement.

The latter method, that of using two different observers, was the one primarily used in the present investigation. The term "objectivity" has also been used in reference to this type of study (*23*).

Data collected in this manner may be analyzed in several ways. Coefficients of correlation are commonly used in studies of reliability. Other possibilities are analyses of differences or of errors. "Differences" refer to the actual difference in value between two measurements. This is the type of analysis employed in the present study. "Errors" refer to deviations of measurements from the "true" value for the given dimension, the true value being represented by the mean of the measurements.

In the special case where there are only two determinations for a given dimension, the *error* for each measurement is exactly *one-half* the *difference* between the two measures. It is important to recognize this distinction since both methods of analysis appear in the literature.

CHAPTER II

A REVIEW OF LITERATURE DEALING WITH THE RELIABILITY OF ANTHROPOMETRIC MEASUREMENTS

The literature in physical growth yields two types of material relating to the accuracy of anthropometric observations: (1) statements by authorities and (2) experimental studies.

GENERAL STATEMENTS

Robert's *Manual of Anthropometry* (32) (1878) recognized the occurrence of measurement errors and suggested how the "true" value for a given bodily dimension might be approximated.

> "Whether we make measurements of the human body or of inanimate things, errors of observation are liable to occur; but it is well known that all such errors are subject to a definite law—the law of the frequency of error—and they can be eliminated by making several measurements, and deducing from them an average which will be the nearest approach to the actual measurement required." (32, p. 45)

Montessori (30) (1913) published an insightful discussion of technical problems relating to anthropometric reliabilities. Her relevant statements include the following:

> "The personal mean error is a datum that it is necessary to know in order to give value to any measurements that we may wish to give forth.
>
> "In taking the various test measurements for the purpose of calculating one's personal error, it is well to use the precaution of not taking them twice at the same sitting, but after an interval of time, not only so that all marks will have disappeared that may have been left upon the skin by the instrument in the act of measuring, but also that the preceding figure will have faded from our memory. Accordingly, the measurements should be repeated on successive days and if possible under the same conditions of *time* and *place*." (p. 388)
>
> "The figures are of more value if they have been compared with the results of other observers; it is necessary, after we

> have found our own average error, to select, for the purpose of verifying our results, some other observer, of similar experience to our own, and whose personal error is also known." (p. 389)
>
> "The personal error cannot be calculated in regard to a single measurement and then applied to all the others, but it must be worked out anew for every separate measurement; it oscillates variously, as a matter of fact, in relation to the longer and shorter diameters, the cranial measurements, and the measurements of the trunk and limbs.
>
> "We are sufficiently skilled to take measurements when we have attained for measurements of cranial diameters a mean error of from 1 to 2 mm., for the vertical cranial diameters one of 4 mm., and for the stature, one of from 5 to 6 mm." (p. 390)

Had students of physical growth given serious consideration to these statements, the problem of anthropometric reliabilities would probably have been developed much more soundly and rapidly than has been the case. Too many have failed to give a clear recognition to the fact that reliability constants "cannot be calculated in regard to a single measurement and then applied to all the others." Too many have taken pairs of observations for reliability studies without heeding "the precaution of not taking them twice at the same sitting."

Davenport (10) (1927) in his manual, *Anthropometry and Anthroposcopy*, included some revelant remarks on the reliability of anthropometric measurements.

> "In general, in making bodily measurements, experience shows that those taken on the bony frame are most significant and most accurate. Thus cranial measurements are reliable to within a millimeter. Stature, depending on a complex of conditions and being a large measurement, is subject both to a large absolute and even relative error. The transverse diameters of pelvis are capable of close measurement. The bony chest can be measured fairly accurately, if allowance is made for the changes in size with respiration. The acromial processes are hung in muscles and their position is subject to great changes. A recognition of the limits of precision is essential to the securing of appropriate safeguards." (p. 10-11)
>
> "It is not sufficient merely to get one measurement of a point. . . . It is desirable to check fundamental measures. This is done by repeating measurements; by taking measurements with different instruments, in slightly different ways." (p. 11-12)
>
> "Not all measurements can be readily checked by comparison of different measurements. On this account it would be desirable to measure the subject completely a second time to find the

> extent of reliability of the measures. In the case of growing children that are being measured repeatedly, one set of measures is checked by the preceding one, if any." (p. 12)

These statements indicate the need to recognize the *limits* of precision, and give suggestions in the checking of measurements. The precaution to be noted in checking by the last procedure, (*i.e.*, with preceding measurements) is that one must be sure that the preceding measurements themselves are highly accurate. There also remain the questions of how much variation would be expected from the preceding examination for various measurements, and how much variation could be accepted and regarded as biologically meaningful.

The second and third conference reports issued by the Committee on Child Development of the National Research Council (1927, 1929) carry several statements concerning reliability of measurement.

In the 1927 report, Baldwin (2) discussed variation in reliability for the different dimensions. In part, he stated:

> " . . . The flexibility of the joints of the shoulders makes this measurement [shoulder width] difficult. Width of hips [bitrochanteric diameter] is a more reliable measurement. . . .
>
> "The thickness of the soft parts makes measurements of the circumference of arm, thigh, and abdomen unreliable for statistical comparison. The amount of fat and the tonus of the underlying muscles vary with each individual, making it difficult to regulate the tension of the tape. However, these measurements do give an insight into certain phases of development." (p. 7)

Todd (37), in the discussion which followed the presentation of Baldwin's paper, reported some of his original findings that he considered relevant to the general problem.

> " . . . if I take a reasonable number of measurements and remeasurements on the cadaver and analyze them, I find that the difference of the mean values is about 5 mm. irrespective of whether the measurement is large, say standing height, or small, e.g. the length of the hand. . . . I measure the hand, take an x-ray of the hand and then I find that my measurements on the living body average about 10 mm. out. From other experiments this seems to be true irrespective of the particular part of the body." (p. 21)

Apparently these statements refer not to "reliability" but to the extent to which measurements of the external body approximate measurement of the skeleton. The latter part of the reference undoubtedly related to the difference between direct measurements on the living hand and measurements taken on a skeletal X-ray of the hand. The first part, on cadavers, probably refers to measurements on cadavers before and after the removal of tissue covering the skeleton. Otherwise, it would be exceedingly difficult to account for the consistent difference in values of means determined on the same sample. While results on this type of experiment are pertinent in certain considerations of validity, they are clearly not applicable to the question of anthropometric reliabilities.

Scammon (33), in the same discussion, made reference to certain original work on reliability:

> "Some time ago we became interested in the matter of sitting height and standing height, and we made a number of experiments on determining these measurements in the living and the dead body. . . . We found that the sitting height . . . was less reliable in the living child than in the dead child. . . . Ordinary disturbances in the room seemed to have little effect on the accuracy of the measurements but if there was an unusual disturbance, it immediately appeared in the measurements. . . . Fatigue seemed to have no effect. . . . We were interested to find that the reliability of hip breadth as a transverse measure is much better than shoulder breadth." (p. 23)

Baldwin (2) made the comment that

> "In regard to accuracy of measurement, our degree of inaccuracy is not on the average 1 cm., and it is not 2 or 3 mm." (p. 25)

This statement by Baldwin might be interpreted to mean that the degree of inaccuracy in taking physical measurements is even less than 2 or 3 mm., and it has been so regarded by some (17). However, in the light of other discussion on the question by Baldwin, it appears more consistent to the writer to interpret the statement as meaning that no one number can be given which will apply to all dimensions; the degree of inaccuracy varies with the dimension considered.

Dearborn (13) indicated the method used in the Harvard Growth Study to minimize the obtaining of unreliable data.

> "In any series of physical measurements we have been accustomed to have two examiners measure independently and when they disagree more than ½ per cent the third examiner measures. Even this does not guarantee accuracy under ½ per cent, for there is the matter of chance, and while practically the measurements agree to a mm. we do not guarantee that degree of precision." (p. 26)

Todd, in a paper read at the third conference (38), gave some summary statements.

> "A critical survey of anthropometric technique and possible errors in deduction has been made by Todd and Lindala. . . .

> According to these authors a discrepancy of 10 mm. may easily occur in any measurement upon the living body and while such errors may cancel out in statistical reduction they render dubious a close adherence to actual individual determinations. Scammon has shown that such discrepancies are not the result of fatigue or lack of attention . . . while Todd and Pyle . . . have suggested, upon the basis of carefully checked evidence, that subconscious retention of figures may be characteristic of certain trained minds. Edith Boyd . . . has taken up the problem of repeated measurement of the child and doubts the advisability of too close a dependence upon the records of monthly measurements. None of these investigators however would detract from the value of assembled and reduced data as evidence of trends in a population." (p. 25)

Thus, according to this authority, reliability of physical measurements is relatively low and to obtain anything more than crude trends of individual growth would, at best, be a "dubious" possibility. Wilson (44), in a paper given later in this conference, referred to the above statement.

> "Dr. Todd in his paper suggests a discrepancy in measurements of ten millimeters. In our work at Merrill-Palmer we have found a probable error of one-sixteenth of an inch or a little over a millimeter, when one person makes the measurement. However, if we compare the work of two different persons in the same laboratory and using the same technique, the probable error is doubled. Even in these groups there would be occasional discrepancies as great as ten millimeters, and if we compared our work with work outside the error would be much greater than I have mentioned." (p. 255)

It is not possible to determine whether this statement applies to only a few measurements or to a large series of various measurements.

Martin (24) (1928) indicated that permissible errors in taking the major head measurements (presumably, head length and breadth) amount to between .5 and 1 mm., for the head height 2 mm., for most body measurements 3 to 5 mm., and for stature and span 10 mm.

Gray and Ayres (17) (1931) stated that when they worked in pairs, they were not satisfied with differences greater than the following: 1 to 2 mm. for head length and breadth, face breadth, nose breadth, and bicristal diameter; 2 to 4 mm. for stature, sitting height, and biacromial diameter; and 2 to 10 mm. for chest measurements. (p. 15)

A preliminary report of the White House Conference series, *Section on Growth and Development: Determination of Physical Status* (43), included a discussion of measurement technique. Brief mention of reliability was made in regard to shoulder measurement and girth measurements.

> "*Bi-acromial or bi-deltoid.* One or the other of these alternative measurements is being increasingly favored over the transverse diameter of the chest as a measurement of the upper part of the trunk. The measurements of the transverse diameter seem to be less reliable than either of the shoulder measurements in question." (p. 7)
>
> "*Girths.* The reliability of measurements of the various circumferences has been debated." (p. 7)

On the one hand, there is an unsupported assertion that one diameter of the trunk is superior to another diameter, and, on the other hand, the need for investigation of reliability of taking girth measurements is indicated.

A recent discussion on the types of errors in anthropometry, is one by Davenport in 1937 (11). In this paper there was a rather thorough discussion of sources of error, as well as suggestions for taking them into account, or eliminating them. In part, he says,

> "To determine different personal errors, it is desirable that persons work in pairs upon a number of the same subjects in immediate succession. Thus there will be revealed surprising differences in the measurements of the same dimensions made by equally conscientious and careful workers. Every anthropometrist should know his personal error, in order that he may make allowance for differences between his work and the work of others.
>
> "It is especially to be observed that errors of instruments and the personal errors are not eliminated but rather exaggerated by large numbers." (p. 95)
>
> "Owing to the great error due to variations in posture of the subject I think it is fair to say that no single measurement is entirely satisfactory. The reason for this is that while one has a number to be sure, one has no means of evaluating the accuracy and hence significance of this number. Two measurements of the same dimension give a means of judging of the significance of either one." (p. 96)
>
> " . . . A single set of measurements is made by techniques that are not standardized, between landmarks which are vaguely defined and difficult to determine; with subjects in varying and unrecorded postures and without checking of measurements by repeated readings. Very little wonder then that it becomes so difficult to compare different results of different anthropometrists. . . . Instances of such differences, chiefly due to differences in techniques might be multiplied.

"Thus the whole study of racial differences and the genetics of such differences is badly handicapped, and research on the physical development of children can hardly be said to be on a scientific basis. It is hoped that a frank recognition of the errors made in anthropometry may lead to improvement of measurements on the living human body." (p. 98-99)

From the preceding quotations, we find statements and suggestions that may be grouped together in the following summary.

1. Measurement errors may be detected and eliminated by various procedures.
   a. One or more persons may make several determinations. (Davenport, Dearborn, Montessori, Roberts)
   b. Observations may be made by different techniques and compared. (Davenport)
   c. Subsequent measurements may be checked with the preceding one. (Davenport)
2. Certain precautions should be taken in the determination of the magnitude of measurement errors.
   a. Observations should not be taken in immediate succession—there should be a reorientation of the subject and a replacement of the instruments. (Montessori)
   b. The possibility of memory of previous observations must be adequately controlled. (Montessori)
3. The magnitude of measurement errors *varies* according to the dimension considered.
   a. Specific limits for different dimensions are indicated. (Montessori, Martin, and Gray and Ayres)
   b. Sitting height, in general, is reported to be less reliable than stature. (Scammon)
   c. The reliability of hip breadth is regarded as much higher than that for shoulder breadth. (Scammon, Baldwin)
   d. Measurement of transverse diameter of the chest is stated to be less reliable than either bi-deltoid diameter or bi-acromial diameter of the shoulder. (White House Conference)
   e. Measurements of girth of arm, thigh, and abdomen are relatively unreliable. (Baldwin)
   f. The reliability of the various girth measurements is considered an open question. (White House Conference)
4. The magnitude of measurement errors *does not vary* according to the dimension considered.
   a. The mean values of a number of measurements and remeasurements differ about 5 mm. irrespective of the size of the measurement. (Todd)

b. A discrepancy of 10 mm. may easily occur in *any* measurement upon the living body. (Todd)

The second point concerning the pitfalls to avoid in the experimental determination of measurement errors will be found useful in the evaluation of the available studies on reliability—as well as in evaluation of the present investigation.

The third and fourth points include statements almost all of which are open to evaluation in terms of experimental data.

EXPERIMENTAL STUDIES ON RELIABILITY

Anthropometric reliability has been studied in a variety of ways and with subjects widely divergent in age. The available investigations will be reviewed according to the age of subjects. This appears advisable since, from the theoretical viewpoint, errors of measurement of one age level cannot be assumed to correspond to those at another age level and since, in practice, there is fairly general recognition that errors tend to be larger for, say, an individual aged eighteen months in contrast to the individual aged eighteen years.

Fetal Period

Schultz (35) (1929) took a series of measurements on one preserved (14th week) human fetus. Three sets of observations were made for each measurement with an interval of one month between successive sets. The maximum differences for the three observations are shown in the following tabulation.

Dimension	Maximum Difference (Tenths of Millimeter)
Foot length	0
Hip breadth, thigh length, leg length, forearm length	1
Trunk height, arm length, hand length, total face height	2
Shoulder breadth	3
Head length, head breadth	4
Sitting height, chest circumference	5
Head height	6

It may be noted that the maximum difference between two observations is less than one millimeter in every instance. Since the most finely scaled instrument stated to have been used was one graduated in half millimeters, it is difficult to see how the majority of the measurements could have been determined to within differences less than this amount. It seems probable that the observations as given by Schultz are not single observations but represent the average of a number of observations.

Scammon and Calkins (34) (1929) reported a study on reliability for crown-heel and crown-rump length in the fetal period and at birth. Three observers co-operated in the study. One held the specimen in position, while another alternately measured crown-heel and crown-rump lengths. The third observer read the scale and recorded the measurement. For different subjects, from forty-two to 346 seriatim observations were taken for each dimension. Rest periods were taken at fifteen-minute intervals.

Data were given on each of the two dimensions for eight single-subject experiments. These included one living newborn infant, three fresh specimens (one of which was used twice), one preserved specimen, the same preserved specimen after it had been set in plaster, and finally, the plaster cast of the preserved specimen.

The findings are presented through the medium of the following statistical constants: the mean, maximum, and minimum for the two dimensions in each experiment; the standard deviation of the observation and the coefficient of variation; the absolute and percentage range.

The absolute range of variation was found to be greater for crown-rump length than for crown-heel length except in the case of the living newborn infant and one preserved specimen. The relative range of variation was always greater for crown-rump length, as was the coefficient of variation.

Comparison between the eight experiments reveals, as would be expected, that the range of measurements obtained on the living infant is greater than the ranges found on nonliving specimens. The range, or maximum difference, on the newborn for crown-rump length was 18 mm., and for crown-heel length 21 mm.

Infancy

Bakwin and Bakwin (1) (1931) studied six dimensions of the body: total body length, sitting height, bimalar diameter of the face, bicristal diameter of the hips, biacromial diameter of the shoulders, and circumference of the thorax at the nipple level. Between eleven and twenty different infants were used and approximately ten determinations were made for each dimension on each infant. The age or age range of the infants is not given. One person took the measurements while another read the scale.

Their data were treated in the following manner. "The average for each series was calculated, and deviations from the average recorded. All the deviations on all the infants were then grouped and the standard deviations computed." (1, p. 373) A table is given showing the average of all the measurements, the standard deviations (of the deviations or errors) and the coefficient of variation. While the deviations for each infant are found as deviations from the mean of the ten observations for a given dimension, they are referred to the mean of the entire series on all children for the determination of the coefficient of variation. There is no statement as to whether the deviations were taken as positive and negative or whether just the absolute value of the deviations was considered. Without this information it is impossible to interpret meaningfully either the standard deviations or the coefficients of variation. Finally, it is not stated whether or not the determinations were strictly independent.

Bayley and Davis (4) (1935), as one estimate of the reliability of their data, correlated measurements taken at one examination period with those taken at the next period. The interval between examinations varied from one to six months, depending upon the age of the child. Judging by this method, the nine dimensions so studied vary from "highly stable" to "relatively unreliable" in this order: weight, total body length, head circumference, stem length, hip width, chest circumference, shoulder width, chest width, and chest depth. This method of analysis was admitted by the authors to give only a rough estimate and was supplemented by a second approach. Fifteen infants between nine days and seven months of age were examined two times, with an interval of only one week between measurements. Ten measurements were taken on each occasion.

In addition to means and standard deviations of the ten dimensions in the two series, mean differences between the two series and standard deviations for these differences are presented. Also, the mean difference is divided by the mean of the series. The following is taken from Table 2 (4, p. 35). The values given in centimeters have been changed to millimeters and the standard deviations rounded to tenths of millimeters.

Dimension	Mean Difference, Millimeters	Standard Deviation of Difference (Tenths of Millimeters)	*Mean Difference* M_1
Length	7.5	4.0	.0139
Stem length	4.6	4.6	.0129
Shoulder width	4.4	2.7	.0285
Chest width	4.1	3.0	.0361
Hip width	1.8	2.1	.0163
Chest depth	2.9	2.6	.0289
Head circumference	2.1	1.8	.0055
Chest circumference	6.5	5.5	.0181
Transverse head	.5	.5	.0050
Antero-posterior head	.4	.5	.0031

The last column gives an index of variability in relation to the absolute size of the dimension. Thus while total length of the body shows the greatest mean difference, other dimensions such as shoulder width, chest width, and chest depth show greater relative variability. The ranges of the differences are not given, but these may be approximated by assuming the lower limit to be zero and estimating the upper limit as three standard deviations above the mean difference. This yields estimates of approximately 19 mm. as the maximum difference for length and 18 mm. for stem length. It is recognized that this is not a rigorous statistical procedure in that a distribution of differences is not "normal". However, it may be assumed that the upper range would not differ markedly from this estimate.

Bayley (3) (1936) in her study of the cephalic index during the first three years of life again employed the technique of correlating measurements at two adjacent age levels. The coefficients, based on from thirty-seven to fifty-seven cases, range between .83 and .97. Averaging the coefficients for the first thirty-six months, she finds that for head circumference it is .92, for head width .93, and for head length .88.

Cates and Goodwin (9) (1936) made a study of the twelve-day-old infant and incorporated data on precision for one observer. The procedure was as follows: "Each investigator recorded thirty measurements of a single dimension under consideration on one subject. This was done at one sitting and an independent observer read the instrument and noted the result each time. The process was repeated for each dimension using a new subject each time . . ." (9, p. 435) A table is presented showing the maximum, minimum, mean, probable error of the mean, the standard deviation, its error, and the coefficient of variation of the observations. The maximum differences obtained for thirty-one dimensions may be derived from this table. The maximum difference between any two determinations on total body length and on crown-rump length was found to be 7 mm.

The following is a tabulation of corresponding maximum differences in millimeters as found in three studies for infants:

Dimension	Scammon and Calkins (34)	Bayley and Davis (4) (Estimated)	Cates and Goodwin (9)
Crown-heel	21	19	7
Crown-rump	18	18	7

There is a marked discrepancy apparent in the values of Cates and Goodwin (9) when these values are contrasted with those of the two other studies. It seems probable that the strikingly smaller ranges of differences found in the study by Cates and Goodwin may be explained in part by the procedure followed. The observations were made "at one sitting". It has been pointed out previously that differences found in immediately successive observations do not completely represent all factors which may affect the reliability of anthropometric observations.

One other point may be noted with regard to the studies reviewed thus far. Comparison has been made in all instances between observations made by the same investigator. That is, the data give information on the reliability of measurements taken by one observer but not on the reliability of measurements taken by several observers. It is reasonable to expect that the latter might be appreciably greater.

Preschool Period

The preschool period is here taken to include the age range from between two or three years of age to six years of age. Examination of the literature on physical growth revealed only two studies of reliability in which the subjects were children within this age interval. Boyd (5) (1929) made a study of errors in measuring stature and body stem by several techniques. Hejinian and Hatt (18) (1929) in a study on *The Stem-Length: Recumbent-Length Ratio* included some data on reliability of taking these two dimensions.

The procedures investigated by Boyd (5) were six. Total body length was taken in the erect position and in recumbent position. Measurements of the body stem include crown-rump length taken on the board, sitting height with the legs straight, sitting height with legs hanging, and the stem length with knees flexed.

The subjects, ranging in age from two to five years, consisted of eighteen girls and seventeen boys who were examined five or more times. Three consecutive readings were usually taken and their mean used as the monthly observed value.

Analysis was made of errors by use of consecutive and progressive measures. "Consecutive" refers "to measures taken in rapid sequences with no appreciable time interval"; while "progressive" refers to "measures taken at appreciable time intervals, such as days or months". (p. 391)

Progressive measures, separated by an interval of one month, were studied by several methods: regularity or irregularity of the individual curves, correlations between different observations of the same dimension, variability of trends fitted to the observed values, and deviations (average and root mean square) from fitted rectilinear and curvilinear trends. The reader is referred directly to Boyd's paper for the results of these analyses. It may be noted here that the curve-fitting method of analysis has certain shortcomings. In the first place, there is the difficulty of choosing the type of curve to be used. Secondly, when the choice has been made, it is fitted to the data in such a way that the observations show a *minimum* deviation from the curve. Finally, the mathematical curve may not coincide with the *true* curve of growth. This last difficulty is revealed in the curves used by Boyd in the study of progressive measurements on body stem. "Those for the four measurements of body stem are much more diverse in type, many indicating decrease in the size of the dimension, and only 60 per cent are either straight lines or approximate them." (p. 408)

Errors of the "consecutive" measures were studied by determining "the mean deviation and mean percentage deviation of the three readings and then averaging these for each measurement on each child, giving a total mean deviation and a total mean percentage deviation." (p. 400-401) Unfortunately this method in which double averaging is involved obscures the range and variation of the individual errors. While, for example, 75 per cent of the total mean deviations are less than 1 mm. for standing height, it cannot be concluded that 75 per cent of all deviations were within this range.

Next, it may be pointed out that the errors for stem length apply to a single observer, immediately repeating the measurement, while the child remains in "a constant position, whether correct or incorrect, during the short time taken to make three immediately successive readings." (p. 410) The three readings for total length were "interrupted." It is not stated whether the child was moved from the board, or whether the child remained in the same position on the board while several other dimensions were measured, before length was repeated.

Hence, for the stem length measures, and possibly for the total lengths, the factor of memory was not shown to be adequately controlled. And finally, in these consecutive measures, the error due to slight variations in the placing of the subject, as Boyd herself indicates, "is not measured by variability of readings taken in rapid succession. . ." (p. 397)

Hejinian and Hatt (18) studied the reliability in measuring stem length and recumbent length, using as subjects three children of different body types. They found that for either of two different technicians "the chances were 957 in 1,000 (3 PE) that a single measurement would not deviate by more than .2 cm. from the mean of 25 measurements taken in sequence on the same child." (p. 299) This is exactly what would be expected since it is stated in a footnote to the above sentence that "when a difference of .5 cm. or more was noted in the measurements of H and B, the results were discarded and another pair of measurements were made." It must be concluded, then, that this part of the study does not relate to "random" reliability, since all differences greater than 5 mm. have been ignored. It was noted that the two technicians did not differ from each other consistently in the same direction.

Mention is made of the fact that measurements of two other laboratory technicians did differ by as much as 17.5 mm.

In another series of measurements on over 100 children, three paired technicians obtained median differences ranging between 1.60 and 2.75 mm. It was shown that on stem length for one pair of observers 58 per cent of the differences were less than 4 mm.; for another pair 75 per cent of the differences were less than 3.3 mm.; and for the third pair 65 per cent were less than 4 mm. For recumbent length 56 per cent of the differences found by the first pair of technicians were less than 4.3 mm.; 82 per cent of the differences were less than 3.2 mm.; and for the third pair 76 per cent of the differences were less than 3.6 mm. The maximum differences were not reported.

No studies were found on reliability of measurements other than those on total body length and length of the body stem using as subjects children between two and six years of age.

Above Six Years: Cadavers

Todd (36) (1925) stated, "It is generally recognized among Anthropologists that measurements made on the living body are reliable to within about 5 mm. and this is the standard to which beginners are trained. We have no information however to show just how much discrepancy may be expected as a result of instrumental method, apart altogether from the deviations due to different observers." (p. 275)

Thus the reliability investigated was that of so-called "instrumental method." This was done by comparing means of bispinous diameter on 100 whites and 100 Negroes (all male), before and after stripping off the overlying soft tissues. The difference between these two means on male white cadavers was found to be 4.7 mm. For male Negroes, it was 2.8 mm. As Todd indicates, since the individual differences were not analyzed, "it may of course be much greater in a single determination." (36, p. 279) In this same article, however, the following statement appears: "This [5 mm.] is the approximate amount by which two successive determinations of bispinous diameter differ from each other." (36, p. 279) Standing alone, this statement would obviously be misleading, and misrepresent the findings. It should read that 5 mm. is the approximate amount by which two determinations of *mean* bispinous diameter (the first *with* the overlying present, the second *without* tissue) differ from each other.

On 100 white male cadavers, several measurements were made: namely, acromiale to dactylion, acromiale to stylion, acromiale to radiale, radiale to stylion, stylion to dactylion, acromiale to floor, and dactylion to floor. It was found that when arm without hand and hand length were summated, they deviated on the average by 6.05 mm. from the direct measurement of arm length. When the middle finger height from the ground was subtracted from the shoulder height, the average deviation from the arm length was 7.66 mm. Addition of upper arm length, forearm length, and hand length yielded an average deviation from arm length equal to 10.76 mm. From these findings, Todd makes the following broad generalization. "Computed dimensions of bodily parts based upon the summation of directly measured factorial components will be subject to this instrumental error of 5 mm. for each of (n-1) components." (36, p. 279) It seems very likely that this would not apply to summation of various other parts of the body; for example, facial dimensions. It is sound only as a generalization applying to the *arm* (and even then only roughly) until substantiated by further data.

Todd and Lindala (39) (1928) published a paper which, though presenting no new material on the problem of reliability, gave a discussion of the problem that warrants inclusion. In this article, they state: "In order that the same point be used in the various measurements we are in the habit of cutting a small cross with a knife through the outer skin over the point as defined by Martin. With this precaution, successive determinations on the same subject are on the average about 5 mm. in error." (p. 40-41) Following this statement, there is a reference to the study reviewed above. As was pointed out, there is no substantiation for this generalization. The study showed only that the differences in means for one dimension (bispinous diameter) on two groups of 100 cadavers were 4.7 and 2.8 mm. when measurements were made before and after removal of the overlying soft tissues. It did not incorporate errors of successive determinations where the points are marked by a cut in the skin and the soft tissues are present for the two determinations.

Immediately following the above statement, Todd made an additional assertion. To quote: "It is my considered opinion that the corresponding error in the living does not exceed but comes close to 10 mm. While many of my anthropological colleagues claim a precision greater than this I am here giving the results of experiments undertaken by myself on school children who are co-operating to the best of their ability but who are naturally subject to slight changes in posture during the examination." (39, p. 40-41) No experimental data are referred to in support of this claim that successive determinations when made on a living subject are *on the average* about 10 mm. in error. This is an unreasonably large average error, except possibly for a very few measurements which are exceptionally difficult to determine. Moreover, if it should refer to actual differences rather than errors from the "true" measurement, average differences of this magnitude would indicate seriously crude technique. Ten millimeters might more nearly approach a *maximum* error for many measurements, rather than an average error.

Todd and Pyle (40) (1928) in a study of the effect of maceration and drying on the vertebral column give additional data on the precision of measurement. Pyle measured composite height of four columns by two different methods. Then after an interval in which she measured some sixty other columns, she remeasured these same four. By the first method, the differences for the four columns on the two successive determinations were 2.0, 2.0, 1.0, and 1.0 mm. By the second method, the differences were 2.0, 0.0, 2.0, and .5 mm. It is then stated that this "evidence which looks so satisfactory is precisely the kind of evidence which gives a false sense of security for it is based on the results of a trained observer consciously checking herself." (p. 309) Two problems immediately present themselves. First, should any person who is not reasonably well-trained attempt, or be permitted, to make anthropometric observations for research use? Secondly, despite a record of years of experience, should any anthropometrist ever lose the sense of making a conscious effort to measure with the greatest amount of accuracy possible? That both questions must be answered in the negative appears self-evident. Moreover, one can scarcely attribute the close agreement obtained in the study to memory, since that factor seems to have been reasonably well controlled.

When the stature of two skeletons was measured by two different observers, the differences for the two determinations were 8.0 and 4.5 mm., thus giving an average difference of 6.3 mm. However, by a different method, differences of 1.0 and 2.0 mm. were found, giving an average of 1.5 mm. Concerning these, it is stated: "We believe that the former difference of 6.3 mm. is a closer probable value and that the difference of 1.5 mm. is a fortunate accident." (p. 309) Might it not be just as reasonable to believe (especially in light of the eight differences found by Pyle, all of which were less than 2 mm.) that the difference of 6.3 mm. was an "unfortunate accident"? The conclusion that one method of determining stature was found to be more reliable than the other method could have been made with more justification. While drawn from too few differences to be accepted as valid, this conclusion would have pointed to a problem awaiting verification by further research.

Above Six Years: Living Subjects

Of eleven studies on reliability based on measurement of subjects over six years of age, five report data on only one or two dimensions. These will be reviewed first.

Gray (15) (1922), in a study of two methods of taking chest girth, gives some findings on the reliability. Using forty male adults and 114 private school boys, the chest girths were taken at the nipple level. For each individual, the "resting" girth and the mean girth between inspiration and expiration were determined. In ten children, the first measurement was found to differ from the average of three successive measurements by 5.7 per cent for the resting girth, and by

4.9 per cent for the midway girth. These differences, it may be noted, are very large, amounting (if only 50 cms. are used as an approximate figure for chest circumference) to at least 2.5 cms.

Gray (16) (1922) took measurements of sitting height and "stem length" (as defined by Dreyer) on 114 boys. He found that the difference between the two measurements was inconstant, ranging from zero to 5 cms. Sitting height was equal to or greater than stem length, the average difference being 1.2 cms. Although no data are given concerning repeated measurements of these two dimensions on the same children, the following statement is made. "The stem-length was found easier to measure accurately, *i.e.*, with more constant agreement on repeated measurements." (p. 418)

On the same problem, Faber (14) (1923) asserts, "Checked by the stem length in the series of sixty-five children, errors of 5 cms. in measuring sitting height were found, and Gray has recently reported the same discrepancies. . . . The sitting height, therefore, is a dangerously inaccurate measure. It can be easily supplanted by the stem length of Dreyer, which appears to have many advantages of accuracy." (p. 340) It should be noted that differences of as high as 5 cms. between sitting height and stem length give no information whatsoever on the relative reliability of these two measures, unless the unwarranted assumption is made that stem length is an ideal measure against which to check other measurements of length of the body stem. Neither Gray nor Faber present any unequivocal data in support of the claim that stem length is the easier to measure accurately.

Gray and Ayres (17) (1931) give information on differences obtained by three observers in measurements of face length and nose height. The subjects were twenty-five adult males. The average differences between the three observers for face length were 1.9, 1.4, and 1.2 mm. Comparable differences for nose height were 1.6, 1.5, and .9 mm. None of the means for the face length of twenty-five subjects as obtained by the three different observers was significantly different. The greatest difference in the means was 1.3 mm.

Whitacre (42) (1934) studies the comparative reliability of two techniques of measuring stature. Stature was taken with the subject against an upright measuring board and against a fixed, perpendicular measuring rod. One hundred and fifty-nine children, in groups of five at a time, were each measured at the board by each observer, and then after 15 to 30 minutes remeasured. Ninety-seven per cent of paired measurements of one observer agreed to within 5 mm. or less;

99 per cent of the pairs obtained by the other observer agreed to within this limit. Only one difference was as great as 10 mm. Comparison of these measurements between the two observers showed that for 81 per cent of the subjects, the four stature measurements (two by each observer) differed by 5 mm. or less, and for 98.6 per cent of the subjects, the difference was 8 mm. or less.

Twenty-five children were measured one day and remeasured after an interval of several days, both at the board and against the rod. The same observer took all the measurements. It was found that for twenty-four of the twenty-five children, the stature measurements at the board differed by only 5 mm. or less, the remaining difference being 8 mm. Against the measuring rod, there was lower agreement. Only fourteen of the twenty-five children had differences of 5 mm. or less, while for three children the difference was 8 mm. and for the remaining children the differences were between 10 and 15 mm. Thus these data show that measurement of stature against a measuring board is more reliable than that taken against a fixed anthropometer. Also, they give clear-cut evidence that successive determinations of stature on school children, made by trained observers, check within an average difference of less than 5 mm.

Lincoln (22) (1930) reported one of the first studies which carried findings relating to a number of different dimensions. The measurements were taken "by graduate students who were carefully trained." (p. 446) The subjects were ninety-four seventh grade pupils. Differences were computed between two observations, approximately half of which were made by the same operator and about half by a different operator. The range, median, and the first and third quartiles of the differences were presented. The median and maximum differences for each measurement (in millimeters) are given in the following tabulation:

Dimension	Maximum Difference	Median
Standing height	39	2.9
Porion height	18	4.8
Sternal height	25	5.8
Dactyl height	33	8.0
Sitting height	43	3.4
Head width	40	1.7
Head length	21	1.7
Chest depth	59	10.0
Chest width	41	6.0

From these findings it was concluded that a more careful technique was needed. In a later check-up on over 100 subjects when a more careful technique was being used, differences between observations were found to be much smaller. They are shown below.

Dimension	Maximum Difference	Median (Millimeters)
Standing height	12	3.2
Sternal height	20	3.7
Sitting height	16	3.7
Iliac width	13	3.0
Chest depth	16	3.8
Chest width	19	3.8
Derived measures		
Leg length	16	3.8
Trunk length	20	5.4
Standing-sternal	14	4.5

Davenport, Steggerda, and Drager (12) (1934) presented an extensive study of reliability. The first part of the study consists primarily of an analysis of repeated measurements of one observer, using one subject. Observer MS measured an adult female, RM, fifty consecutive times at different times of the day through a number of days. The mean of the fifty determinations, together with the standard deviation and probable error of the distributions and the coefficient of variation, was presented for each of forty-nine dimensions. Certain of the dimensions were determined in different ways, so that in all, data are given for sixty-five items.

The following tabulation presents findings on selected measurements.

Dimension	Mean	Standard Deviation	4 x Standard Deviation (Estimated Range)
		Millimeters	
Stature	1672.23	3.10	12
Head length	194.24	.43	2
Head breadth	146.30	.46	2
Head girth	572.84	1.86	7
Sitting height	905.08	5.36	21
Chest girth	852.32	7.78	31
Lower arm girth	231.70	2.21	9
Intercristal	301.48	3.13	12
Chest depth	200.76	2.40	10
Acromio-radiale	294.04	3.64	15
Chest width	271.08	3.50	14
Radiale-stylion	227.96	3.26	13
Calf girth	325.52	4.71	19
Hand length-tracing	181.24	2.70	11
Acromion breadth	353.56	5.46	22
Upper arm girth	273.96	6.15	25

The range of observations for each dimension is not given. This has been estimated by multiplying the standard deviations given by four. Four standard deviations (instead of six) were used here, since the range seldom is as great as six standard deviations when the number of cases is not very large. This procedure reveals that there were probably some rather large differences between the minimum and maximum observations.

In Part II of the study by Davenport, Steggerda, and Drager, additional findings on reliability are given. This section reports the results from measuring eleven adults, five males and six females. The only clothing worn was a bathing suit. Two observers, MS and CBD, each measured each subject twice. The first and second observations by MS were averaged for each dimension on each subject. The same was done for the values obtained by CBD. From the average for MS was subtracted the average obtained by CBD. These differences were then averaged algebraically for the eleven subjects. Unfortunately, by thus taking into account the direction of the difference, and averaging plusses and minuses as such, the magnitude of the differences is misrepresented unless it may so happen that all the differences are in the same direction. There is no statement to the effect that this is the case.

For fifteen dimensions, the averages of the differences for the two observations made by the same technician were given. Since no statement to the contrary is made, it is assumed that the subjects were the eleven adults referred to above. No measure of variability of the differences is presented. Also, given in the table showing the average findings for CBD and MS are similar results obtained by the Dahlbergs (G. D. and S. D.). These average differences are shown in the following tabulation.

Dimensions	Average of the Difference, Millimeters			
	CBD	MS	GD	SD
Head length	1.55	.64	.51	.40
Head breadth	1.18	.18	.54	.54
Frontal minimum diameter	1.91	.73	1.38	1.25
Face breadth	1.36	.64	.69	.44
Bigonial diameter	1.64	.64	.65	.73
Face length	3.70	1.00	1.52	1.36
Stature	5.09	5.91	3.56	4.57
Sternal height	8.55	9.18	4.25	5.39
Height of symphysis			4.53	8.06
Height of acromion	10.55	4.55	4.61	6.00
Height of tip of middle finger	9.55	7.27	5.09	6.38
Length of arm	4.90	3.73	4.03	6.41
Length of trunk	3.36	6.09	5.83	9.09
Biacromial diameter	5.73	5.00	3.31	2.99
Bicristal diameter	9.73	4.73	1.64	1.81

Carey (1936) gives mean differences obtained by one observer on six different dimensions. Measurements on thirty boys were repeated on successive days. The boys used in the study were between eleven and eighteen years of age. His results are given in the following tabulation.

Dimension	Mean Difference	Standard Deviation Difference	Estimated Range 4 x Standard Deviation
		Millimeters	
Standing height	3.2	1.6	6
Shoulder height	5.7	3.7	15
Leg length	2.5	1.4	6
Shoulder width	2.4	1.7	7
Hip width	1.7	1.0	4
Span	3.1	2.5	10

The estimated maximum difference (or range) column shows considerably smaller values than those found by Davenport, Steggerda, and Drager. This fact might be accounted for, in part, by the age of the subjects used.

The two most recent studies on reliability were done on eight- and nine-year-old males. The first study was by Meredith (1936). The second study, employing the same group of children as the first, was by Marshall (1937).

The subjects were twenty-five boys, ranging in age between seven years, eight months and ten years, three months. Each boy was measured monthly for approximately seven months. At each examination period each boy was measured three times: first by HM, then by EM, and finally by HM.

Meredith's study (captioned a study of "reliability") presented an analysis of the data obtained from the pairs of measurements taken by HM. Marshall's study gave an analysis of the differences obtained between the two different observers, HM and EM. This study was designated a study of "objectivity," to distinguish it from the former.

In both studies, the actual distributions of differences were presented. Also, the median, the 90th percentile, and the maximum difference were given for each dimension. A series of fifteen measurements was studied.

The following tabulation gives the results (in millimeters) obtained by each author.

Dimension	Meredith Median Difference	Meredith 90th Percentile	Meredith Maximum	Marshall Median Difference	Marshall 90th Percentile	Marshall Maximum
Stature	2.0	7.5	12	3.6	9.2	19
Sitting height	2.7	7.8	15	3.1	9.5	20
Thoracic circumference	2.4	9.1	20	2.8	11.0	24
Biacromial diameter	3.5	10.3	24	3.0	12.5	24
Arm girth	.9	3.9	7	1.9	5.3	15
Forearm girth	.8	3.2	7	1.3	4.1	9
Thigh girth	3.2	7.8	12	4.0	9.5	17
Leg girth	.8	2.6	5	1.4	4.2	8
Bi-trochanteric diameter	.8	3.3	7	1.5	5.1	10
Bi-iliac diameter	.4	1.5	3	.9	2.3	4
Breadth of elbow	.6	1.7	2	.9	2.2	3
Breadth of knee	.3	1.3	2	.8	1.8	5
Thickness of skin and subcutaneous tissue						
At arm back	.6	1.5	3	.8	2.1	3
At thorax back	.3	1.3	3	.4	1.3	2
Above iliac crest	.4	1.7	4	.9	3.0	6

Supplementary reliability findings were incorporated in each study by relating the median difference and the maximum difference for a given dimension to the mean size of the dimension. Coefficients of correlation for each dimension were also given. These ranged in size from .82 to .99. Finally, on the assumption that "seriatim observations are only profitably made at such intervals as the mean increment of growth equals or exceeds 90 per cent of the reliability differences", a table of estimated maximum frequency for remeasurement was given.

The estimated frequency of measurement for eight- and nine-year-old boys is as follows:

Dimension	Meredith	Marshall
Stature	Bimonthly	Bimonthly (?)
Bi-iliac diameter	Quarterly	Quarterly (?)
Girth of leg	Quarterly	Semiannually
Bi-trochanteric diameter	Quarterly (?)	Semiannually
Sitting height	Semiannually	Semiannually
Thoracic circumference	Semiannually	Semiannually
Girth of thigh	Semiannually	Semiannually (?)
Forearm girth	Semiannually	Annually
Arm girth	Semiannually	Annually
Bi-condylar diameter of femur	Annually	Annually
Bi-acromial diameter	Annually	Annually
Bi-condylar diameter of the humerus	Biannually	Biannually

In summary of the available experimental data on reliability the following notations appear pertinent.

1. The criticism most commonly applicable is that the range of errors or differences is not given in most studies. This is true of the studies by Bakwin and Bakwin; Bayley and Davis; Boyd; Carey;

Davenport, Steggerda, and Drager; Gray and Ayres; Hejinian and Hatt; and Todd. In some of these studies the standard deviation is included, but this affords only a rough estimate of the range.

2. There appears no clear indication in several studies as to whether the subject remained in one position while repeated measurements were made or whether measurements were strictly independent. This shortcoming characterizes the studies of Bakwin and Bakwin, Cates and Goodwin, and a portion of the studies of Hejinian and Hatt, and Boyd. It is clearly stated by Boyd that for some of the dimensions studied, the subject did not move around in the interim between measurements.

3. Other specific deficiencies are:

a. Boyd's study indicates no attempt to control the factor of memory.

b. Bakwin and Bakwin give insufficient description of the ages of the subjects used.

c. Hejinian and Hatt discarded differences greater than 5 mm. in part of their study, thereby invalidating the results in that portion.

d. Faber, Gray, and Todd give generalizations inadequately supported by experimental data.

CHAPTER III

AIMS, MATERIAL, METHOD OF ANALYSIS

With specific reference to the age period from three to six years, the review of previous research on reliability has revealed that:

1. Only two studies, one by Boyd and the other by Hejinian and Hatt, have been reported.

2. These studies covered but two dimensions of the body, stature and sitting height.

3. One of the studies, that by Boyd, used immediately successive determinations of the same observer. The procedure was not sufficiently rigorous on two counts. First, it did not adequately control the factor of memory. Second, the subject was not oriented anew for each determination, and, consequently, errors due to minor fluctuations in posture (possibly the greatest single source of error) were excluded.

4. In the other study, that by Hejinian and Hatt, differences in one experiment which were greater than 5 mm. were discarded, thus invalidating this section. In the other experiments undertaken, the complete range of differences was not presented.

The present study includes data designed to contribute to the scientific study of anthropometric reliabilities by adherence to the following general aims:

1. To make observations on a wide sampling of subjects between the ages of three and six years

2. To investigate a large number and variety of measurements

3. To follow a rigorously independent procedure, a procedure that justifies the assumption (a) that memory has been adequately controlled and (b) that errors due to undetectable fluctuations in the position of the subject as he is oriented and reoriented for measurement have not been eliminated

4. To give a reasonably complete analysis of the data by inclusion of the maximum range of differences as well as other indices of the variability and central tendency

SUBJECTS

The subjects were children between the ages of three years, zero months and six years, zero months. One hundred twelve of these children were enrolled in the Iowa Child Welfare Research Station preschool laboratories or in the University of Iowa elementary school. Nineteen were orphanage children in residence at the Iowa Soldiers' Orphans' Home, Davenport, Iowa.

Sixty-six, or 50 per cent, of the children had never been measured before. Forty-seven (36 per cent) had been measured in preschool the previous year. The remaining eighteen (14 per cent) had not been measured the year immediately preceding, but had been measured at an earlier time at preschool or in the infant laboratory maintained by the Iowa Child Welfare Research Station.

Distribution of the subjects according to age and sex is given in the following tabulation.

Age, Years	Males	Females	Total
3	22	16	38
4	22	18	40
5	30	23	53
Total	74	57	131

The use of 131 different children, selected only in terms of age and availability, assures a wide sampling of build. Actual examination of the children gave ample evidence that no weighting of the sample with builds relatively easy to measure was present.

COLLECTION OF DATA

The data were all collected by four persons, who will be designated as K, M, R, and S. All had measured during the previous year according to the technique employed at the Iowa Child Welfare Research Station, and were trained to work with preschool children. The two observers, K and M, had longer experience in such work. K had been actively engaged in the work for over two years and M for over six years.

Before measurements were taken, as much time was spent in establishing rapport with the child as was necessary. The attempt was made to maintain rapport equal to that necessary for administering a good mental examination. In only four cases did the attempt to establish or maintain rapport fail. No measurements were taken on a crying or fretful child, nor any child who showed fear of the instruments or procedure.

Reliability values were obtained by two procedures. The major procedure was that of obtaining the difference between observations made by two different technicians. This was employed with all the measurements studied. A supplementary procedure was used for five of the measurements. Here the differences between two independent observations by the same technician were secured. The five measurements were chest width, chest depth, hip width, knee width, and shoulder width.

The observations made by the two technicians were completely independent. The person taking the second measurement had no knowledge of the values obtained by the first person. Hence, any attempt for one person to duplicate or approach the measurement obtained by the other person was entirely eliminated in this part of the study.

The question of memory appears only in the small portion of the study in which one observer repeated five measurements. This question may be considered after a description of specific procedure used when these data were collected.

Measurements Taken During July and August, 1937

The measurements in this series were all made by the two persons, K and M. Children enrolled in the preschool laboratories and elementary school for the summer term served as subjects. The procedure in each group was as follows:

1. *Junior Primary (Five-year-olds)*.—The children were examined in pairs. K took ten different measurements on child A, while M obtained a like series on child B. Then K took the ten measurements on child B while M measured child A. Finally, K repeated five of the series on A and then on B, while M served as recorder.

2. *Preschool Group III (Four-year-olds)*.—Fifteen children were measured on two successive days. On the first day, K took the five measurements on each of six children. After the children had an interval of play, M took the measurements on the children and K remeasured the children in random order. On the second day, the the procedure was repeated on a group of nine other children.

3. *Preschool Group II (Three-year-olds)*.—The measurements were taken on two days. On one day, the same procedure as that used in Group III was followed for a group of nine children. On the other day, the children were taken two at a time, and while K measured child A, M measured child B. Then the children changed places. After this, K remeasured child A and then B. Only five measurements were taken by each examiner.

4. *Preschool Group I (Three-year-olds)*.—Only two children of this group (which was primarily composed of children under the age limits

covered in this study) were used. Five measurements were taken by K and M on each child. K did not remeasure the children as was the case in the other three groups.

This first set of data includes all measurements which were repeated by the same technician. We may now take up the question of memory. In all cases but one, it is clear that there is reasonable justification for the claim that memory did not affect the results. The one possible exception is the second day of measurement in Group II. Here there were only five intervening measurements before K repeated the measurements on a given child. Examination of the data collected on this day, however, revealed that the differences were equally great in all five dimensions as those on days when memory was more satisfactorily controlled. It is thus believed safe to conclude that the results of the repeated measurements by the same person are not vitiated by the factor of memory.

Measurements Taken Between October, 1937 and February, 1938

Throughout these five months, the regular measurement program in the preschools was in progress and at this time the major portion of the data was collected. Thirty measurements were taken on each of sixty-seven children by two different observers. For a majority of the series of measurements there was a recorder. When this was so, one person took all thirty measurements and called the values to a recorder. When these were completed the second technician entered the room and measured the children.

When no recorder was available, the first person measuring recorded for himself and then for the second technician.

If the last measurement taken by the first anthropometrist was a recumbent measurement, the child was not left in position for the second anthropometrist. The same was true if the last measurement was stature or sitting height erect. The time required for the two complete sets of measurements averaged about one hour.

Since some of the children examined during the summer session (July and August) also attended school during this next year, these children were used again. This accounts for the fact that while just 131 different children were used, paired observations on some dimensions total 150, and while only fifty-seven different females and seventy-four different males were used, there are as many as seventy-five pairs of observations for each sex.

The technicians worked in the pairs: K and M, K and R, and K and S.

Measurements Taken March 12, 1938

At this time, technicians K, M, and R, each with a recorder, gathered observations at the Iowa Soldiers' Orphans' Home.

For fourteen children, twenty-six measurements were taken on each child by each observer. For five other children, a series of seven measurements was made by each person.

The use of three different observers in this set of data makes possible the intercomparison of three different pairs of measurement.

Measurements Taken March 23, 1938

Observers K, M, and R each independently measured head length, head breadth, hand length, stature, and sitting height on children in Preschool Group III. The head measurements had not been included in the fall series. Stature and sitting height were being taken in connection with another study.

Measurements in April, 1938

Two observers, K and M, took head breadth, head length, and hand length measurements on children in Junior Primary and in Preschool Groups II and III.

Measurements

Thirty-three measurements were made directly and in addition two were derived. These two were the measurements of length of the lower extremity, erect and supine.

Of these thirty-five measurements, twenty-nine were different measurements, while the remaining six were of dimensions included in the group of twenty-nine but were taken with the subject in a different position. That is, stature, sitting height, length of the lower extremity, chest width, chest depth, and chest girth were taken with the subject standing and also with the subject in recumbent position.

The measurements, together with the number of observations for each, are given in the following tabulation.

Measurement	Paired Observations Girls	Boys	Total
Lengths			
Stature			
Erect	50	50	100
Recumbent	25	23	48
Sitting height			
Erect	50	50	100
Recumbent	23	22	45
Upper extremity			
Arm length	50	50	100
Forearm length	50	50	100
Hand length	50	50	100
Length of upper extremity	50	50	100
Lower extremity			
Stature minus sitting height, erect	50	50	100
Stature minus sitting height, recumbent	23	22	45
Transverse and antero-posterior diameters			
Of the head			
Length of head	66	84	150
Breadth of head	66	84	150
Of the trunk			
Shoulder width—biacromial	75	75	150
Chest width, erect	75	75	150
Chest width, recumbent	50	50	100
Chest depth, erect	75	75	150
Chest depth, recumbent	50	50	100
Abdomen depth	50	50	100
Hip width—bi-iliac	75	75	150
Of the extremities			
Elbow width	50	50	100
Knee width	75	75	150
Girth measurements			
Head girth	50	50	100
On the trunk			
Abdomen	50	50	100
Chest—supine	34	21	55
Chest—erect	50	50	100
Upper extremity			
Arm	50	50	100
Arm, with plunger	50	50	100
Forearm	50	50	100
Lower extremity			
Thigh	50	50	100
Leg	50	50	100
Measurements of thickness of skin and subcutaneous tissue			
At chest front	50	50	100
At chest back	50	50	100
Above iliac crest	50	50	100
At abdomen	50	50	100
At arm back	50	50	100

All measurements were taken on the nude body. Precise definitions of measurement technique appear in the appropriate sections of the succeeding four chapters in which the results are presented.

METHOD OF ANALYSIS

The basic data then consist of 3,643 pairs of measurements taken by two different observers and also 221 pairs of measurements taken by the same observer.

The difference between each pair of measurements was computed and the resulting figures tabulated for each dimension.

Final analyses were made for each dimension with the three ages and both sexes combined. Although there was little reason to expect differences in reliability in measuring the two sexes at these ages, the distributions were first made for each sex. No differences sufficient to require separate treatment were found. The following tabulation shows the distribution for males and females for six of the dimensions.

Difference Interval, Millimeters	Per cent					
	Girls	Boys	Girls	Boys	Girls	Boys
	Hip Width		Fat, Arm Back		Fat, Chest Front	
0 to 1	84.0	82.7	72.0	78.0	84.0	94.0
2 to 5	14.7	16.0	28.0	22.0	16.0	6.0
6 to 10	1.3	1.3				
Number	75	75	50	50	50	50
	Arm Girth		Forearm Girth		Head Girth	
0 to 1	36.0	48.0	58.0	74.0	32.0	50.0
2 to 5	64.0	48.0	42.0	26.0	46.0	36.0
6 to 10		4.0			18.0	12.0
11 to 15					4.0	2.0
Number	50	50	50	50	50	50

There is possibly some reason to expect that measurements would be markedly more difficult to take on three-year-olds than on five-year-olds. This was not, however, found to be the case. The following tabulation gives the distributions of differences for the three ages in eight dimensions.

Difference Interval, Millimeters	Per Cent					
	Age, Years			Age, Years		
	3	4	5	3	4	5
	Stature			Sitting Height		
0 to 1	28.0	25.0	33.3	38.5	31.3	34.6
2 to 5	60.0	58.3	33.3	42.3	39.6	50.0
6 to 10	12.0	16.7	33.3	15.4	29.2	15.4
11 to 15				3.8		
Number	25	48	27	26	48	26

Difference Interval, Millimeters	Per Cent: Age, Years 3	4	5	Age, Years 3	4	5
	Chest Depth			Shoulder Width		
0 to 1	38.2	33.9	43.3	25.0	29.6	30.8
2 to 5	55.9	60.7	50.0	56.8	55.6	36.5
6 to 10	5.9	5.4	5.0	15.9	11.1	21.2
11 to 15			1.7	2.3	3.7	7.7
16 to 25						3.8
Number	34	56	60	44	54	52
	Thigh Girth			Leg Girth		
0 to 1	14.3	33.3	21.4	67.9	73.3	66.7
2 to 5	60.7	43.3	54.8	32.1	26.7	33.3
6 to 10	21.4	20.0	21.4			
11 to 15	3.6	3.3	2.4			
Number	28	30	42	28	30	42
	Chest Girth			Abdomen Girth		
0 to 1	17.4	31.0	14.6	21.4	20.0	14.3
2 to 5	56.5	44.8	52.1	35.7	30.0	23.8
6 to 10	26.1	13.8	29.2	25.0	30.0	26.2
11 to 15		10.3	4.2	7.1	16.7	16.7
16 to 25				10.7	3.3	16.7
26						2.4
Number	23	29	48	28	30	42

The differences obtained between the various pairs of observers are analyzed, all being considered positive. The median and the 90th percentile were computed for each distribution. This choice of indices of central tendency and variability was made, since the distributions dealt with were not "normal." (See distributions in following chapters.)

The differences found between the observations of K and M were also tabulated separately, since these two persons had had longer experience in measurement technique. Also, differences between paired observations of one observer were treated separately. These distributions could then be compared to those obtained using all four persons. In only a few dimensions were there indications of closer agreement between K and M than for all pairs of technicians. By analysis of differences tabulated according to direction of the difference, it was possible to note any consistent bias in technique between anthropometrists K and M. Findings on these two points will be given along with those for the total group.

Six dimensions were taken in two ways: (1) with the child erect and (2) with the child recumbent. It is well established (31) (41, p. 24, 26) that these two methods yield reliably different values for stature. In the present study, the differences in value between erect and recumbent stature are tabulated for each individual. Similarly, this was done for sitting height, length of lower extremity, chest width, chest depth, and chest girth. The individual values are taken as the mean of the paired observations of two technicians. These differences may be used to facilitate comparison between data based on one method with that based on the other.

The four following chapters (IV, V, VI, and VII) present results on the four groups of measurements: (1) lengths, (2) transverse and antero-posterior diameters, (3) girths, and (4) measurements of skin and subcutaneous tissue.

Correlations between two observations on the same child were computed for each dimension. The correlations were done over the whole age range rather than for each separate age group, since the overlapping of distributions was great. For stature, the dimension in which there was the least amount of overlapping of ranges, the correlations were computed separately for each age as well as for the total group. There was no significant difference found even in this dimension. The correlation in height determinations for the three-year-olds was .996 ± .002, for four-year-olds .994 ± .002, for five-year-olds .984 ± .007, and for the total group .997 ± .001.

The median and maximum differences found for each dimension are related to the mean magnitude of that dimension on the children examined, yielding reliability findings which may be directly compared among the various dimensions.

Finally, reliability constants are related to mean annual increments of growth, as a means of estimating approximate frequency at which it may be desirable to repeat measurements on given individuals.

These findings, *i.e.,* correlation coefficients, the relation of differences to the size of the dimension and to the annual increment in the dimension, are presented in Chapter VIII.

CHAPTER IV

RELIABILITY OF MEASUREMENTS: LENGTHS

The present chapter and the succeeding three chapters give parallel analyses for length measurements, breadth measurements, girth measurements, and measurements of skin and subcutaneous tissue, respectively. In each chapter, tabular presentation and discussion of results is preceded by detailed definitions of the measurements included.

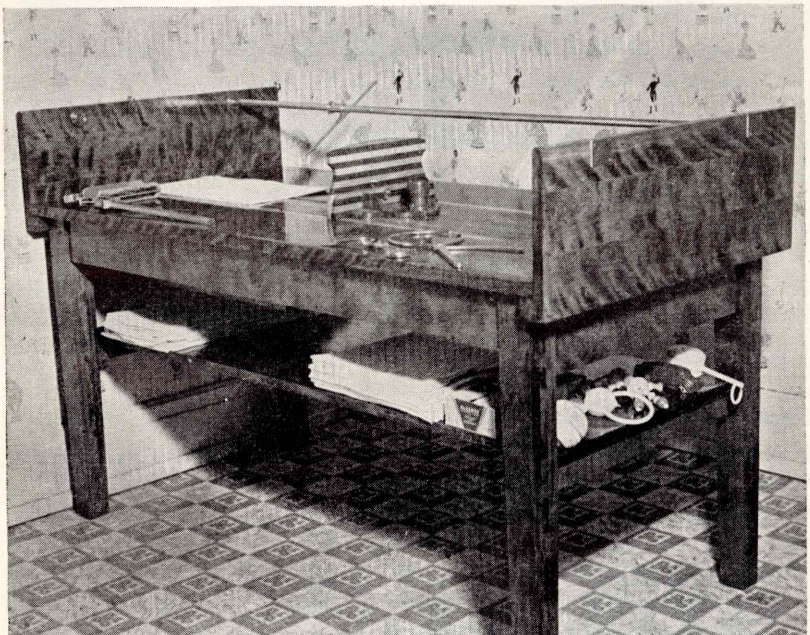

Figure 2. Anthropometric Table

DEFINITIONS

Stature: Erect

Instrument.—Baldwin Paper Measuring Scale and Square

Definition of Measurement.—Erect body length from the soles of the feet to the vertex

Landmark.—Vertex—the most superior point of the head when head is oriented so that the tragionorbitale plane is at right angles to the long axis of the body

Posture.—The subject stood erect with heels almost touching each other and in firm contact with the floor. Heels, buttocks, upper part of the back, and rear of head were in contact with the wall to which the scale is attached. The arms were permitted to hang at the sides in a natural manner.

Technique.—When the subject was in position, the anthropometrist took the square and placed it above the subject's head so that its one face was against the scale and its other face was horizontal with the floor. The square was then brought down until the horizontal face crushed the subject's hair and made firm contact with the vertex.

Stature: Recumbent

Instrument.—Meredith-Knott Anthropometric Table (See Figure 2.)

Definition of Measurement.—Recumbent body length from the soles of the feet to the vertex

Posture.—The subject reclined on the table in a position of dorsal recumbency. The vertex made firm contact with the vertical head board, and the body axis paralleled the millimeter scales which are embedded in the table top at right angles to the head board.

Technique.—The anthropometrist oriented the subject with the aid of an assistant. The assistant then continued to hold the head in position while the anthropometrist placed one hand above the subject's knees to prevent flexion and with the other hand brought the sliding foot board into firm contact with the soles of the feet.

Sitting Height: Erect

Instruments.—Baldwin Paper Measuring Scale, Square, and a horizontal walnut bench thirty centimeters in height

Definition of Measurement.—Distance from the surface of the bench to the vertex with the subject in an erect sitting position

Posture.—The subject sat on the bench with his knees flexed and hanging free or (if his legs reached the floor) flexed and spread apart. In this latter case, the ankles were crossed before the knees were spread apart. The subject's hands were placed on his thighs, and the posterior aspect of his trunk made contact with the scale both at the sacral region and at the upper thoracic region.

Technique.—The square was brought down firmly on the vertex and the sitting height recorded as the value which was read from the scale minus the height of the bench.

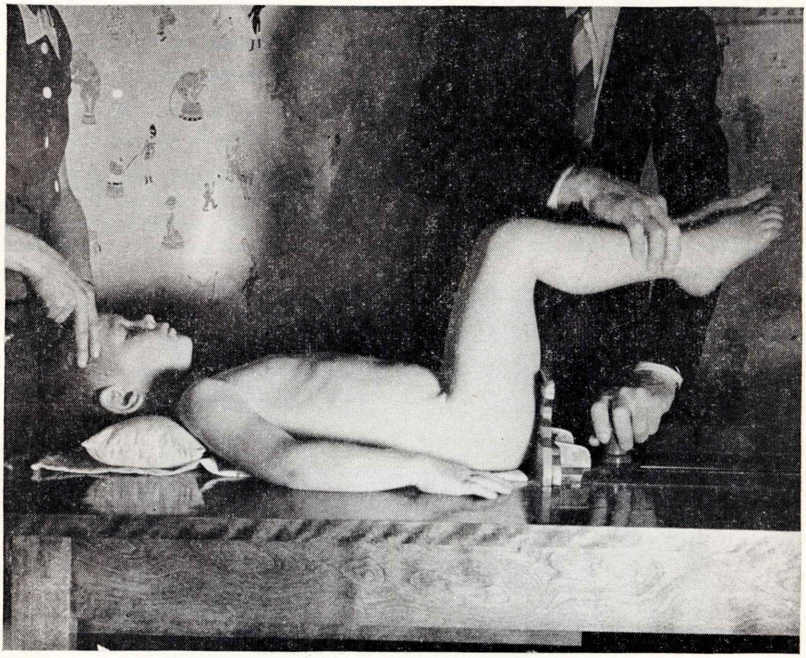

Figure 3. Sitting Height: Recumbent

Sitting Height: Recumbent

Instrument.—Meredith-Knott Anthropometric Table

Definition of Measurement.—Distance from the vertex to the most caudal projection of the (compressed) soft tissues inferior to the ischia, with the subject in a supine sitting position

Posture.—The subject reclined on the table with the vertex firmly contacting the head board and body axis parallel to the millimeter scales. The lower extremities were then put into a position analogous to that for erect sitting height: *i.e.*, the thighs were perpendicular to the table top and the legs at right angles to the thighs. (See Figure 3)

Technique.—The anthropometrist, while holding the lower extremities in position with one hand, brought the sliding board into firm contact with the buttocks. As in the case of recumbent stature, an assistant was responsible for maintaining the head position during measurement.

Length of Lower Extremities: Erect

Definition.—Sitting height erect subtracted from stature erect.

Note: Used interchangeably with the term "erect leg length"

Length of Lower Extremities: Recumbent

Definition.—Sitting height recumbent subtracted from stature recumbent
Note: Also designated as "recumbent leg length"

Arm Length

Instrument.—The Sliding Calipers (small)

Definition of Measurement.—Length of arm from acromion to radiale

Landmarks.—Acromion—the most lateral point on the acromion process of the scapula. Radiale—the most proximal point on the lateral side of the capitulum of the radius

Posture.—The subject stood with the upper extremities hanging in an extended position at the sides of the body.

Technique.—The calipers were applied with the shaft paralleling the axis of the left arm and the inner surfaces of the branches at the level of the landmarks.

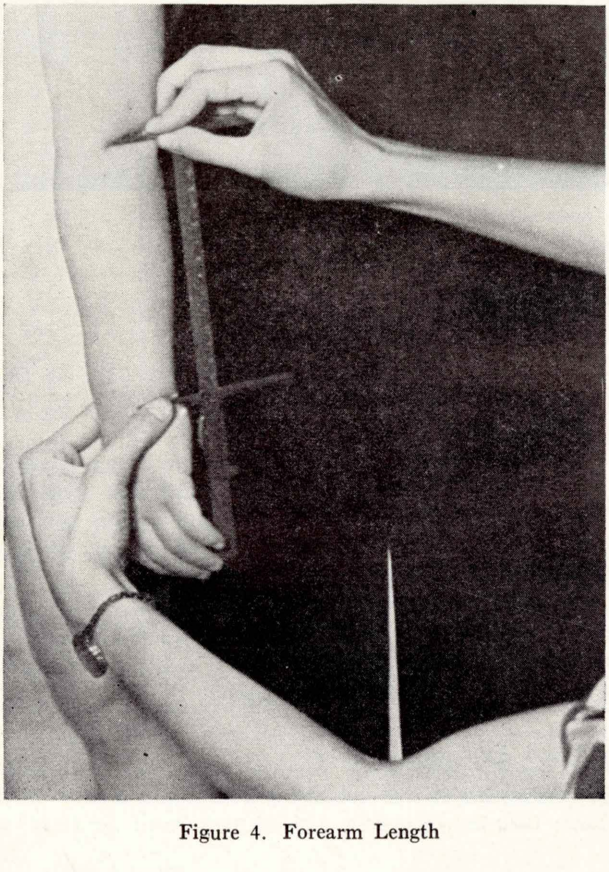

Figure 4. Forearm Length

Forearm Length

Instrument.—The Sliding Calipers (small)

Definition of Measurement.—Length of forearm from radiale to stylion ulnare

Landmarks.—Radiale. Stylion ulnare—most distal point on the medial side of the styloid process of the ulna

Posture.—The subject stood with upper extremities hanging in an extended position at the sides of the body. The left forearm was rotated so that at its distal end the position of the ulna was lateral to that of the radius.

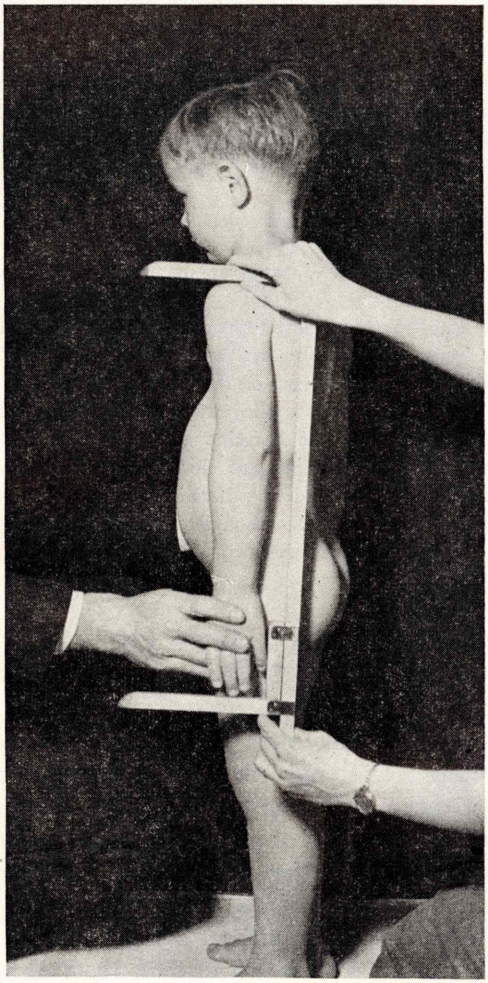

Figure 5. Length of Upper Extremity

Technique.—With the shaft of the instrument parallel to the forearm axis, the branches were applied to the two landmarks. (See Figure 4)

Hand Length

Instrument.—The Sliding Calipers, ruled paper, and pencil

Definition of Measurement.—Length of hand from stylion ulnare to dactylion

Landmarks.—Stylion ulnare. Dactylion—the most distal point of the middle finger

Posture.—The subject stood with his left arm in ninety degrees abduction, his left forearm and hand in the same plane as the arm but flexed at right angles to it and resting on a table, with hand pronated and digits moderately fanned.

Technique.—The hand rested on a sheet of ruled paper so that the forearm-hand axis coincided with a ruled line. Marks were made on the paper immediately below the landmarks, stylion ulnare and dactylion. The hand was removed from the table and, with the calipers so placed on the paper that the shaft paralleled a ruled line, the distance between these two marks was measured.

Length of Upper Extremity

Instrument.—Wooden Calipers (Wooden Shoulder Breadth Calipers, Narragansett)

Definition of Measurement.—Length of upper extremity from acromion to dactylion

Posture.—The subject stood with upper extremities in extended position at the sides of the body. The left forearm was rotated so that the palmar surface of the hand faced the lateral surface of the thigh.

Technique.—The calipers were applied with the shaft paralleling the left arm axis and the inner surfaces of the branches at the level of the landmarks. Care was taken to see that the child did not elevate his right shoulder and tilt his left. The child's attempt to extend or otherwise accommodate to the instrument frequently resulted in the tilt of the shoulder axis. (See Figure 5)

Note: Also referred to as "total arm length"

ANAYLSIS OF TOTAL DATA ON ERECT AND RECUMBENT LENGTHS

The distribution of differences between independent observations for stature, sitting height, and length of lower extremity are given in Table 1. Below each distribution the median, the 90th percentile, and the maximum difference are shown. Selected findings from Table 1 are:

1. The median differences for erect stature, recumbent stature, and erect sitting height fall between 2 and 3 mm.; those for recumbent sitting height, and erect and recumbent leg length between 4 and 5 mm. These values are within a millimeter of those reported by (a) Hejinian and Hatt (18) on stature and sitting height recumbent on children of

preschool age, (b) by Lincoln (22) for erect stature, sitting height, and derived leg length on seventh grade children, (c) by Carey (8) for erect stature of boys over twelve years of age, and (d) by Meredith and Marshall (23) for erect stature and sitting height of eight- and nine-year-old boys. They are between 2 and 3 mm. less than those given by Davenport, Steggerda, and Drager (12) for erect measurements on adults. The mean differences given by Bayley and Davis for infants are 5 mm. larger for recumbent stature, but identical with the median difference of the present study for recumbent sitting height.

Table 1

Reliability Findings for Erect and Recumbent Body Lengths: Based on Differences Between Independent Observations of Two Technicians

Difference Interval, Millimeters	Stature		Sitting Height		Length of Lower Extremity	
	Number	Per Cent	Number	Per Cent	Number	Per Cent
Erect Measurements						
0 to 1	28	28	34	34	20	20
2 to 3	40	40	25	25	25	25
4 to 5	12	12	18	18	23	23
6 to 7	12	12	14	14	16	16
8 to 9	6	6	7	7	8	8
10 to 11	2	2	1	1	5	5
12 to 13			1	1	1	1
14 to 15					0	0
16 to 17					2	2
Number	100		100		100	
Median	2.6		2.4		3.9	
90th percentile	6.5		7.4		8.5	
Maximum	10		12		17	
Recumbent Measurements						
0 to 1	16	33.3	10	22.2	9	20.0
2 to 3	16	33.3	18	17.6	11	24.4
4 to 5	13	27.1	11	24.4	8	17.8
6 to 7	1	2.1	10	22.2	10	22.2
8 to 9	2	4.2	2	4.4	4	8.9
10 to 11			2	4.4	2	4.4
12 to 13			2	4.4	0	
14 to 15					1	2.2
Number	48		45		45	
Median	2.2		4.6		4.0	
90th percentile	5.2		9.0		8.3	
Maximum	9		13		15	

2. The 90th percentile is smallest for recumbent stature and largest for recumbent sitting height. For stature and sitting height erect, the 90th percentiles are 6.5 and 7.4 millimeters, respectively. These values correspond roughly to those found by Meredith (26) on eight- and nine-year-old boys, where the same observer made pairs of independent measurements. They are below those found by Marshall on the same children, where differences obtained between two different observers were employed.

3. Maximum values are smallest for the stature and greatest for the derived measurements of the lower extremity. The maximum differences for the stature measurements (10 mm.) are about one-half the size of those found by Scammon and Calkins (34) on a newborn infant and by Marshall (23) on eight- and nine-year-old boys. They more nearly correspond to the results of Whitacre (42) on children of school age.

4. No marked tendency for one method of measurement of a given dimension to involve smaller absolute differences than the other was found. The distributions for the leg measurements derived by the two methods are very similar. There is a tendency for the differences to be somewhat smaller for recumbent stature and erect sitting height, when compared respectively to erect stature and recumbent sitting height.

ANALYSIS OF THE DATA ON ERECT AND RECUMBENT LENGTHS BY K AND M

The data analyzed above consisted of differences obtained by four pairings of observers: K-M, K-R, K-S, and M-R. The pair, K-M, as indicated in the previous chapter, were the more experienced technicians. The following is a tabulation of the findings on differences obtaining between these two observers:

Dimension	Number	Median	90th Percentile	Maximum
		Erect		
Stature	56	2.8	7.6	10
Sitting height	58	2.3	7.6	10
Leg length	55	4.3	8.8	17
		Recumbent		
Stature	23	2.0	6.2	9
Sitting height	22	4.0	7.4	10
Leg length	22	4.0	7.9	9

Only four of these values found for the K-M pairing are more than a millimeter smaller than those found for the total group. The maximum difference for K-M was 2 mm. less for erect sitting height, 3 mm. less for recumbent sitting height, and 6 mm. less for recumbent length of the lower extremity than the maximum differences for all pairs of observations. The 90th percentile for recumbent sitting height, based on the differences between K and M, is 1.6 mm. less than that for the total group. It is clear that these composite results are not invalidated by the inclusion observations made by the two technicians having a lesser amount of experience than was the case for K and M.

An investigation was made to see if there was any systematic bias between the observations of K and M, or whether the differences were normally distributed about zero as a midpoint. In order to contrast the observations made by these two technicians, the medians of the differences were computed when the direction of the difference was taken into account. That is, the difference was considered positive when the observation of K was greater than that of M and negative when the opposite was true. The medians computed for such distributions were as follows:

	Median, Milli-meters
Erect stature	1.5
Erect sitting height	– .8
Erect leg length	2.3
Recumbent stature	– .8
Recumbent sitting height	–1.5
Recumbent leg length	.5

Comparison of these medians with their corresponding standard errors revealed that four of these values did not differ significantly from zero. The differences approached significance for erect stature and leg length. The ratio of the median for erect stature to its standard error was 2.25, which is interpreted to signify that there are 988 chances in 1,000 that the median of 1.5 mm. cannot be attributed to chance. The ratio for erect leg length was 2.77, corresponding to 997 chances in 1,000 for a true difference from a zero midpoint. Thus, it is seen that K tended to obtain slightly larger measurements for erect stature and leg length than did M. A difference of this magnitude might possibly be explained on the basis of a rather consistent variation in the firmness with which the square was brought down on the vertex.

THE COMPARABILITY OF ERECT AND RECUMBENT LENGTHS

Supplementary analysis on this portion of the data was made for the difference in absolute magnitude between a given dimension taken with the subject in the erect position and the same dimension taken with the subject in dorsal recumbency. The results are shown in the following tabulation:

Difference, Millimeters	Stature Recumbent Minus Stature Erect	Sitting Height Recumbent Minus Sitting Height Erect	Leg Length Recumbent Minus Leg Length Erect
-20 to -16			1
-15 to -11			2
-10 to - 6			3
- 5 to - 1		2	10
0 to - 4	14	10	15
5 to 9	16	18	6
10 to 14	9	6	2
15 to 19	1	1	
20 to 24	1	2	
Number	41	39	39
Median	7.4	6.8	.6
Range	0 to 21	–3 to 20	–19 to 13

The central tendency of these differences is smaller than that generally reported. Boyd (5) states that body length is from 15 to 20 mm. greater than standing height. The discrepancy is probably due to differences in technique, the measurements of the present study being made with firm pressure against the feet.

ANALYSIS OF DATA ON LENGTHS OF UPPER EXTREMITY

Results for the independent observations for the four length measurements made on the upper extremity are shown in Table 2. Findings from this table are:

1. The magnitude of the median difference increases from 2.6 to 5.6 mm. for four dimensions studied in the following order: hand length, total arm length, forearm length, and arm length. Davenport, Steggerda, and Drager report average differences for length of arm for the same observers checking themselves. These range from 3.6 to 5.9 mm., using adult subjects. The corresponding median difference here is 4.3 mm.

2. The four lengths maintain the same ranking on the 90 percentiles as that shown in the medians. These percentiles are 10 to 12 mm.

for each measurement with the exception of hand length where the value is 7 mm.

3. The maximum difference is 17 mm. for arm, forearm, and hand length. For length of the upper extremity, it is 22 mm. The study referred to above does not give a measure of variability in connection with the means.

Table 2

Reliability Findings for Arm Lengths: Based on Differences Between Independent Observations of Two Technicians

Difference Interval, Millimeters	Arm Length	Forearm Length	Hand Length	Length of Upper Extremity
	Number and Per Cent	Number and Per Cent	Number and Per Cent	Number and Per Cent
0 to 1	15	12	30	23
2 to 3	18	20	29	19
4 to 5	16	20	20	22
6 to 7	14	24	12	13
8 to 9	19	11	4	9
10 to 11	5	6	3	8
12 to 13	8	5	1	1
14 to 15	3	1	0	2
16 to 17	2	1	1	1
18 to 19				1
20 to 21				0
22 to 23				1
Number	100	100	100	100
Median	5.6	5.3	2.6	4.3
90th percentile	12.1	10.3	7.3	10.2
Maximum	17	17	17	22

Since measurements of lengths on the upper extremity had not been taken during recent years as part of the routine anthropometric examination but were specifically introduced for purposes of this study, it is especially pertinent to examine the findings for the pair of technicians, K-M:

Dimension	Number	Median	90th Percentile	Maximum
Arm length	55	4.7	12.3	17
Forearm length	56	4.5	9.0	10
Hand length	54	2.2	5.0	10
Length of upper extremity	58	4.5	9.7	12

The medians shown here are all within 1 mm. of those for all pairs together. However, there is a marked contrast in three of the maximum values. While the maximum differences for forearm and hand length are 17 mm. for the total paired observations, here they are only 10 mm. Again, the maximum difference between all observations on length of the upper extremity is 22 mm., but for K-M it is 12 mm. These smaller maximum deviations represent a higher standard of accuracy, and hence one toward which technicians should aim.

Examination was made of the direction of the differences between K and M. The differences were computed as the value obtained by K minus the value obtained by M. The median and its standard error was found for the distribution for each of the four length measurements made on the arm. Results were as follows:

Dimension	Number	Median Difference	Median Difference / Standard Error, median
Arm length	55	4.2	4.62
Forearm length	56	4.2	6.09
Hand length	54	-0.5	.86
Length of upper extremity	58	2.0	2.35

For hand length, no tendency for one technician to obtain either systematically higher or lower values is revealed. A consistent bias is clearly indicated for arm and forearm length, since the ratios signify that the chances are very high that the magnitude of the medians cannot be ascribed to chance factors alone. For the length of the upper extremity, the chances are 991 in 1,000 that the median differs significantly from zero. This consistent directional difference in the observations of K and M warrants further attention. If careful exploration of the methods used by these two technicians should reveal the reason or reasons for this phenomenon, perhaps more precise definition of the measurements could be made and the reliability improved.

CHAPTER V

RELIABILITY OF MEASUREMENTS: TRANSVERSE AND ANTERO-POSTERIOR DIAMETERS

The diameters that were studied include two on the head, seven on the trunk, and two on the extremities.

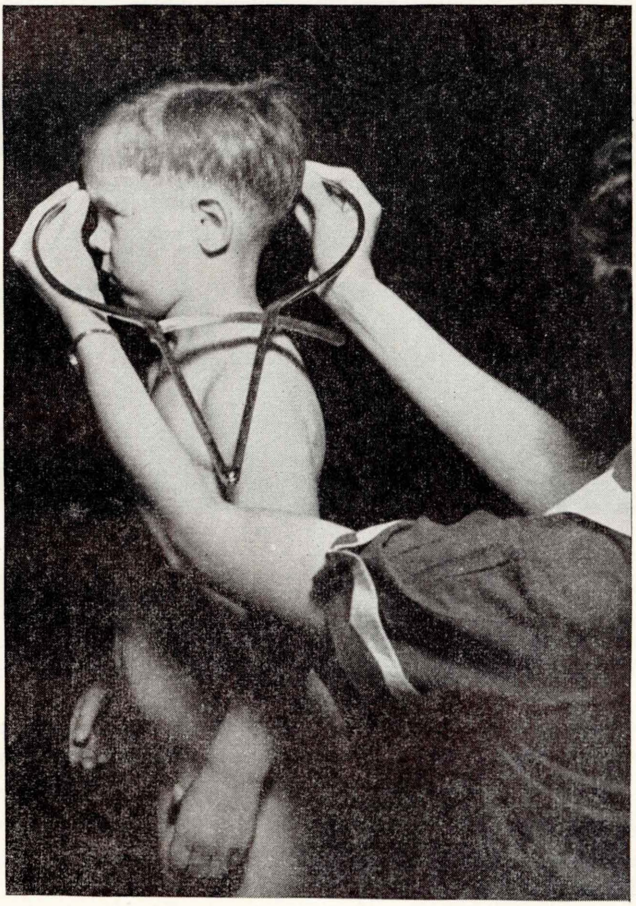

Figure 6. Head Length

DEFINITIONS

Head Length

Instrument.—The Hrdlička Spreading Calipers

Definition of Measurement.—The glabella ad maximum diameter

Landmarks.—Glabella: the point in the midline of the forehead at the level of the superciliary arches. The point in the median plane of the occiput farthest from the glabella

Technique.—The left branch of the calipers was applied to the anterior landmark, the thumb and two succeeding fingers of the left hand being so placed about the tip of this branch as to form a ball-and-socket arrangement while holding the tip in position. The right branch of the calipers was moved up and down the back of the head until the maximum distance was encountered. (Figure 6)

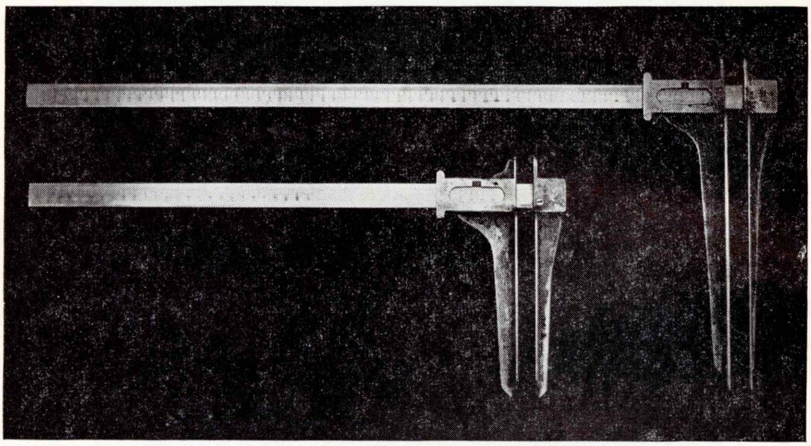

Figure 7. Hrdlička Sliding Calipers

Head Breadth

Instrument.—The Hrdlička Spreading Calipers

Definition of Measurement.—Maximum transverse diameter

Landmarks.—The point at either end of the greatest transverse diameter of the head above the level of the zygomatic arch and supramastoid crest

Technique.—The tips of the spreading calipers were placed against the sides of the head and moved around in a horizontal plane until the greatest width was found.

Shoulder Breadth

Instrument.—The Hrdlička Sliding Calipers (modified for use on young children)[1] (Figure 7)

[1]The shaft of the Hrdlička Large Sliding Calipers was cut off at the 43 cm. mark and the branches were reduced to about ⅔ of the original size.

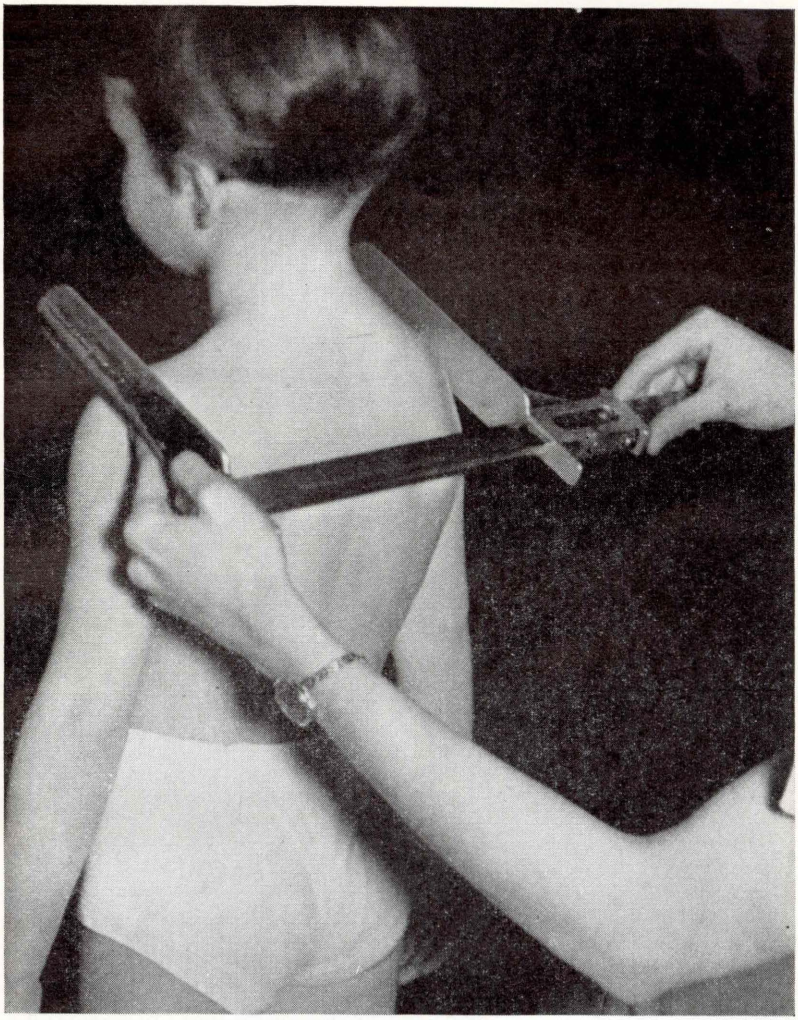

Figure 8. Shoulder Breadth

Definition of Measurement.—Biacromial diameter

Landmarks.—Acromion: the lateralmost point of the acromion process of the scapula

Posture.—The subject stood erect with upper extremities in contact with the sides of the body, but not rigidly so. The shoulders were neither "slumped forward" nor "thrown back." An attempt was made to have the subject in his natural erect position.

Technique.—The anthropometrist stood at the rear of the subject with one of the broad flat branches of the calipers resting on the back of each hand and projecting forward in a plane between the thumb and second digit. The branches were extended to the approximate breadth of the subject's shoulders and, after each landmark had been located with a free finger, they were brought against the acromion points. Moderate pressure was applied. (Figure 8)

Chest Width: Erect

Instrument.—The Hrdlička Large Sliding Calipers (modified)

Definition of Measurement.—Transverse diameter of the thorax at the level of the xiphoid cartilage

Posture.—The subject stood erect, though not tensely erect. His arms were relaxed and held slightly away from the sides of the body to permit the passing of the instrument between the arms and the lateral walls of the thorax.

Technique.—The anthropometrist stood in front of the subject and applied the calipers to the thorax in such a manner that the shaft lay directly in the plane of the xiphoid cartilage and the branches were perpendicular to the vertebral column. The contact of the tips of the branches with the lateral aspects of the thorax was as light as possible. The scale was watched carefully and, where possible, two or three respiratory excursions of the chest were followed. The record was made as the median diameter during the normal inspiration-expiration oscillations.

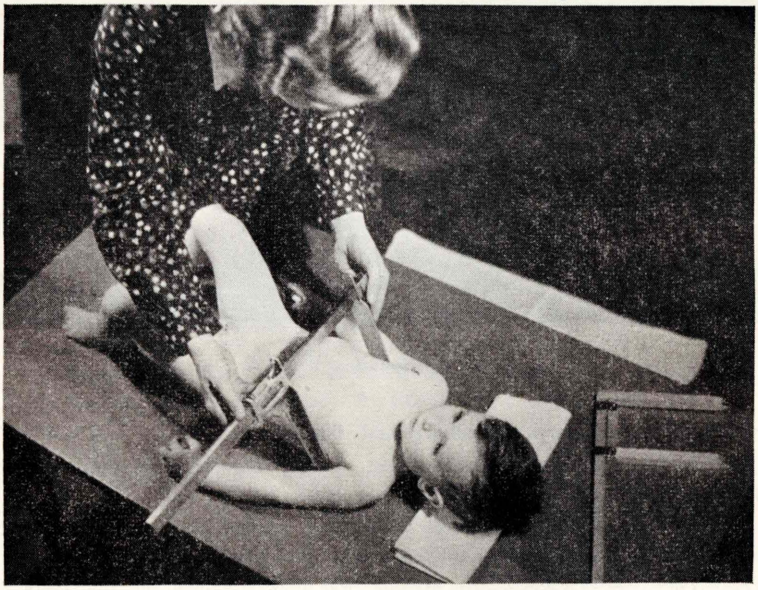

Figure 9. Chest Width: Supine

Chest Width: Supine

As above, except that the subject took a position of dorsal recumbency

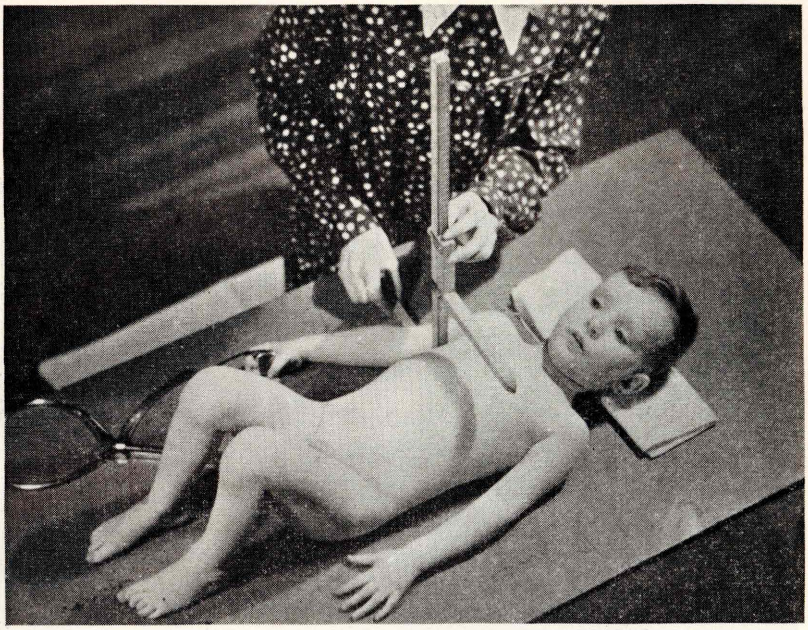

Figure 10. Chest Depth: Supine

Chest Depth: Erect

Instrument.—The Hrdlička Large Sliding Calipers (modified)

Definition of Measurement.—Antero-posterior diameter of the thorax at the level of the xiphoid cartilage

Posture.—As for erect chest width

Technique.—The anthropometrist stood on the right side of the subject. The shaft of the instrument was passed between the arm of the subject and his right lateral chest wall. With the shaft oriented perpendicular to the vertebral column, the fixed branch was brought against the back at the level below the inferior angles of the scapulae while the moving branch was placed in a horizontal plane crossing the xiphoid cartilage. The measurement was taken to represent the intermediate stage between inspiration and expiration. (Figure 9)

Chest Depth: Supine

Instrument.—Wooden Calipers (Wooden Shoulder Breadth Calipers, Narragansett)

Definition of Measurement.—Vertical distance from the surface upon which the subject reclines to the anterior chest wall at the level of the xiphoid cartilage

Posture.—The subject reclined on a table in a position of dorsal recumbency with knees flexed at approximately ninety degrees and soles of feet resting on the table.

Technique.—The movable branch of the calipers was removed, turned through one hundred and eighty degrees and replaced. The fixed branch of the calipers was placed on the table at the level of the xiphoid cartilage. The subject's right arm rested across the fixed branch, the shaft stood perpendicular to the table top and close to the lateral chest wall, and the movable branch extended over

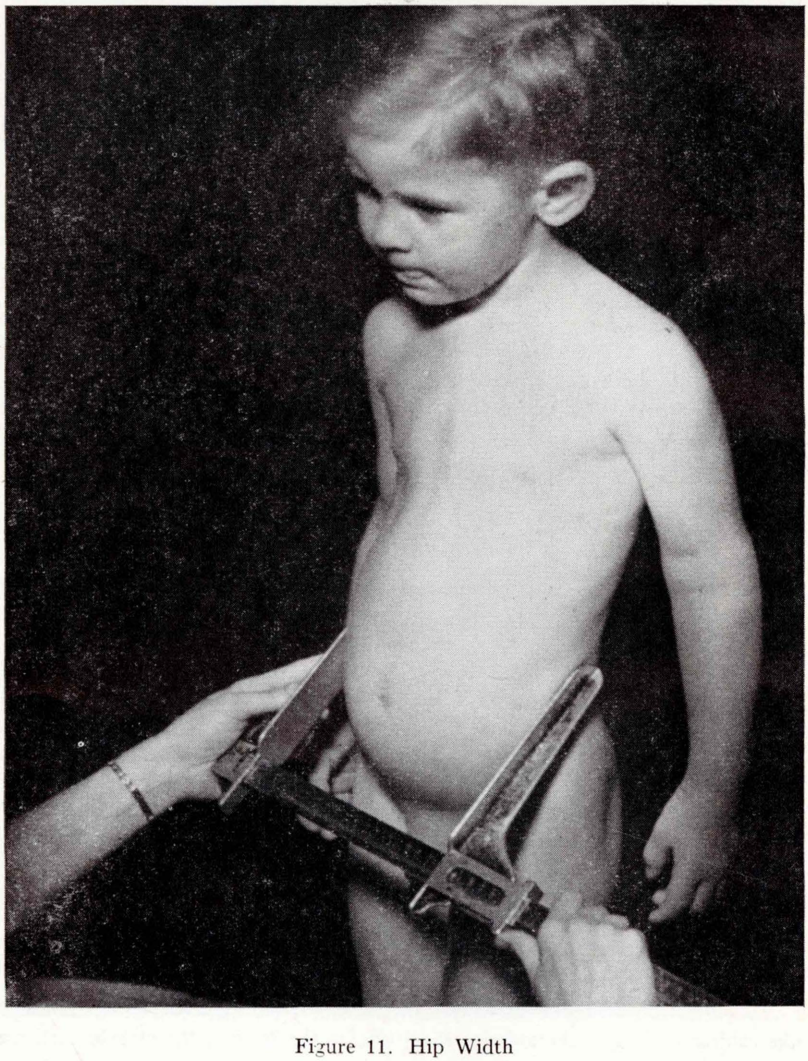

Figure 11. Hip Width

the anterior chest wall at the xiphoid level. The sliding branch was then moved downward until light contact was made with the anterior wall of the chest. (Figure 10)

Abdomen Depth: Supine

Instrument.—Wooden Calipers

Definition of Measurement.—Distance from the surface upon which the subject reclines to the anterior abdominal wall at the level of the umbilicus

Posture.—As for chest depth supine

Technique.—The calipers were adjusted as for chest depth supine. The fixed branch of the calipers was placed under the right forearm of the subject, so that the shaft was vertical and the movable branch extended over the abdomen of the subject at the level of the umbilicus. The sliding branch was then moved downward until light contact was made with the abdomen in maximum distension during normal respiration.

Hip Width

Instrument.—The Hrdlička Large Sliding Calipers (modified)

Definition of Measurement.—Bi-iliac (intercristal) diameter

Landmarks.—Iliocristale: the lateralmost point of the crest of the ilium.

Posture.—The subject assumed an erect position.

Technique.—The anthropometrist stood in front of the subject and brought the face of each branch of the calipers squarely in contact with one of the landmarks. The maximum pressure was applied that could be exerted without pain to the subject. In the event the subject appeared to turn his hips as the pressure was applied, the measurement was checked. (See Figure 11)

Elbow Width

Instrument.—The Hrdlička Large Sliding Calipers (modified)

Definition of Measurement.—Bi-condylar diameter of the distal extremity of the left humerus

Landmarks.—The most lateral point of the lateral epicondyle of the humerus and the most medial point of the medial epicondyle of the same bone

Posture.—The left arm of the subject was raised forward to approximately the level of the shoulder and the forearm was flexed upward at right angles to the arm.

Technique.—The branches of the calipers were applied against the epicondyles of the humerus in such a manner as to bisect the angle of the elbow and to lie in the same plane as the arm and forearm. Heavy pressure was used.

Knee Width

Instrument.—The Hrdlička Large Sliding Calipers (modified)

Definition of Measurement.—Bi-condylar diameter of the distal extremity of the left femur

Landmarks.—The points on either epicondyle of the lower extremity of the femur most lateral to the median plane of the bone

Posture.—The left knee of the subject was flexed sufficiently to relax and largely remove the musculature at the lateral aspects of the epicondyles.

Technique.—The anthropometrist stood in front of the subject and applied the branches of the calipers to the bony prominences in such a manner that they bisected the thigh-leg angle. With the calipers bisecting the knee angle and the branches lying in a plane parallel to the thigh and leg, heavy pressure was applied and the maximum diameter determined.

Table 3

Reliability Findings for Transverse and Antero-Posterior Diameters: Based on Differences Between Independent Observations of Two Technicians

Difference Interval, Millimeters	Number	Per Cent	Number	Per Cent	Number	Per Cent	Number and Per Cent
	Chest Width Erect		Chest Width Supine		Shoulders		Abdomen Depth
0 to 1	65	43.3	49	49	43	28.7	31
2 to 3	59	39.3	36	36	43	28.7	22
4 to 5	16	10.7	9	9	31	20.7	25
6 to 7	6	4.0	5	5	16	10.7	9
8 to 9	3	2.0			6	4.0	6
10 to 11			1	1	5	3.3	4
12 to 13	1	.7			3	2.0	3
14 to 20					3	2.0	
Number	150		100		150		100
Median	1.8		1.5		2.8		3.1
90th percentile	4.5		4.3		8.2		8.5
Maximum	12		11		20		13
	Chest Depth Erect		Chest Depth Supine		Hip Width		
0 to 1	58	38.7	41	41	125	83.3	
2 to 3	57	38.0	35	35	22	14.7	
4 to 5	26	17.3	20	20	2	1.3	
6 to 7	7	4.7	4	4	1	.7	
8 to 11	2	1.4					
Number	150		100		150		
Median	2.1		1.9		.8		
90th percentile	4.8		4.9		2.1		
Maximum	11		6		6		
	Head Length		Head Breadth		Knee Width		Elbow Width
0 to 1	86	57.3	125	83.3	132	88.0	90
2 to 3	55	36.7	25	16.7	17	11.3	9
4 to 5	8	5.3			1	.7	1
6 to 7	1	.7					
Number	150		150		150		100
Median	1.2		.6		.7		.5
90th percentile	3.1		2.0		1.7		1.5
Maximum	6		3		4		4

ANALYSIS OF DATA ON HEAD DIAMETERS

The total data for all eleven diameters to be considered in this chapter are given in Table 3. The lower portion of this table carries the results relating the head length and head breadth. Selected findings are:

1. The median difference for paired observations on head length is 1.2 mm., while on head breadth it is .6 mm.

Bayley and Davis (4), on fifteen infants, found a mean difference of .4 mm. for head length and .5 for head breadth. These lower differences found on infants might be explained by the fact that children of that age have much less hair obscuring the posterior landmark. Davenport, Steggerda, and Drager reported average differences on eleven adults. For one observer, the mean difference was .6 for head length and .2 for head breadth; for the other observer, the values were 1.6 for head length and 1.2 for head breadth.

2. The 90th percentile falls at approximately 3 mm. for head length and at 2 mm. for head breadth.

3. The maximum difference in paired observations of head length is 6 mm. and for head breadth 3 mm.

4. The findings of the present study indicate that head breadth is more accurately taken than is head length.

The differences found between the observations of K and M were compared with the total series of paired observations. For head breadth, the median, the 90th percentile, and the maximum of the K-M differences were each smaller than the corresponding constants for the total data. In the measurement of head length K-M showed somewhat better agreement. The median difference for this pair of technicians was .9 mm., the 90th percentile 2.2 mm., and the maximum difference 3 mm. From Table 3 it may be seen that the maximum for the total observations was 6 mm., or twice that for K-M.

The K-M differences for head length and breadth, considered with regard to sign, showed no tendency for either observer to deviate systematically from the other.

ANALYSIS OF TOTAL DATA ON TRUNK MEASUREMENTS

Seven diameters of the trunk were studied. They include four measurements of the thorax, one of the shoulder girdle, one of the abdomen, and one of the pelvic girdle. Results on the paired independent observations for these dimensions are shown in the two upper sections of Table 3. Selected findings are:

1. When ranked on the basis of the magnitude of the median difference, the dimensions take the following order: hip width (.8 mm.), supine chest width, erect chest width, supine chest depth, erect chest depth, shoulder width, and abdomen depth (3.1 mm.).

These median differences, in general, fall below mean differences reported by other investigators: namely, Bayley and Davis, Carey, and Lincoln. Davenport, Steggerda, and Drager found average differences markedly greater than those in the present study. Their values were 5 and 6 mm. for shoulder (biacromial) width, and 5 and 10 mm. for hip (bicristal) width, in sharp contrast with the corresponding medians of 2.8 and .8 mm. shown in Table 3. It may be noted that Davenport *et al* measured the intercristal width in such a way that the "skin was hardly indented." From the standpoint of reliability, probably the technique of using maximum pressure is superior. In the studies of Meredith and Marshall where the same technique as that of the present study was used, the differences found were very close to those found here: less than 1 mm. in both instances.

2. A ranking of the measurements of the trunk, based on 90th percentile values, gives the same order as obtained for median differences.

3. Maximum differences range from 6 mm. for hip width and supine chest depth to 20 mm. for shoulder width.

4. Comparison of the distributions and the statistical constants for erect chest width and supine chest width clearly reveals that there is no significant difference in reliability for the two methods. The values for the subject in the erect position are practically identical with those for the subject in the supine position.

5. While the medians and 90th percentiles for chest depth erect and chest depth recumbent are found to be in close agreement, the maximum difference is larger for erect than for supine measurement of the dimension. The maximum values are 11 and 6 mm., respectively. Although only a small percentage of the differences for erect chest depth were greater than the maximum for supine chest depth, if by taking chest depth with the subject in supine position the range of difference is reduced, it may be worth-while to consider the supine method as the more desirable procedure. The explanation for the lower range of difference probably lies in the fact that the instrument is placed with regard to only one body landmark, whereas in erect chest depth the placing of the instrument involves two areas of contact with the body.

ANALYSIS OF DATA FOR MEASUREMENTS OF THE TRUNK COLLECTED BY K AND M

The total data on erect chest width and depth, shoulder width, and hip width are heavily weighted with differences between K and M. This is due to the fact that these measurements were four of the five investigated in July and August, 1937, when K and M were the only technicians making observations. In consequence, as might be expected, the differences for the pairings of K and M are almost exactly the same as the differences reported for the total group. The maximum is the same in each case, and the medians and 90th percentiles vary only by tenths of millimeters.

For the supine measurements of chest and the abdomen depth, again, there is no indication of higher reliability between the observations of K and M than between those for all pairing considered together.

With regard to any tendency for one of the technicians, K or M, to obtain higher or lower readings for a given measurement, reference may be made to the following listing of medians computed with regard to the direction of the differences.

Dimension	Median Differences, Millimeters
Chest width erect	0.1
Chest width supine	0.0
Shoulder width	1.7
Abdomen depth	−0.5
Chest depth erect	1.0
Chest depth supine	0.5
Hip width	0.5

Four of these medians—those for erect chest width, supine chest width, abdomen depth, and supine chest depth—do not differ significantly from zero. However, the medians for shoulder width, erect chest depth, and width of hips, when related to their standard error, yield ratios all greater than 3; that is, they represent a central tendency which differs reliably from zero. The positive median difference in the hip width is explained by a slightly greater pressure exerted by M, resulting on the whole in smaller observed values. The factor or factors accounting for the systematic bias in the observations on shoulder width and erect chest depth probably include differences in orientation of subjects as well as pressure differences.

ANALYSIS OF PAIRED DIFFERENCES FOR TRUNK MEASUREMENTS TAKEN BY K

For four of the measurements—erect chest width, erect chest depth, shoulder width, and hip width—two independent observations were made by one observer, K. The following tabulation presents the data (in millimeters) for these observations.

Dimension	Number	Median	90th Percentile	Maximum
Chest width	42	1.5	4.2	6
Shoulder width	45	1.9	5.0	11
Chest depth	45	1.4	3.3	4
Hip width	44	.9	2.3	4

The values of this tabulation may be compared with corresponding values (based exclusively on independent observations for different observers) shown in Table 3. Shoulder width and chest depth constants are found to be considerably smaller for paired observations of one observer than for the paired observations of different observers. For chest width and hip width, the medians and percentiles are in close agreement, but the maximum is smaller for the single observer. In general, then, it is found that two observations by the same technician agree more closely than do two observations by different technicians. This is also confirmed by a comparison of the findings of Meredith on "reliability" and Marshall on "objectivity."

THE COMPARABILITY OF CHEST DIAMETERS TAKEN ERECT AND SUPINE

The difference in magnitude of measurements on the chest as taken by the two methods is distributed in the following manner:

Difference Interval, Millimeters	Chest Width Supine Minus Chest Width Erect	Chest Depth Supine Minus Chest Depth Erect
−12 to −10		2
− 9 to − 7		11
− 6 to − 4		14
− 3 to − 1		28
0 to 2	9	17
3 to 5	35	5
6 to 8	28	1
9 to 11	5	
Number	77	78
Median	5.0	− 2.6
Range	1 to 11	−12 to 7

The thorax becomes somewhat flattened when the subject changes from the erect position to the position of dorsal recumbency. Specifically, on assumption of the supine position, chest width is found to show a median increase of 5 mm. and chest depth a decrease of approximately one-half this amount.

ANALYSIS OF TOTAL DATA ON BREADTH OF THE EXTREMITIES

The last two dimensions to be considered in this chapter are the two measurements, knee width and elbow width. The last section of Table 3 gives the data on these measurements from which the following findings have been selected:

1. The median difference for elbow width was .5 mm., and for knee width .7 mm.
2. The 90th percentile is 1.5 mm. for elbow breadth and 1.7 mm. for knee breadth.
3. The maximum in each case is 4 mm.

These results for knee width lie midway between the results reported by Meredith and those reported by Marshall.

The median, the 90th percentile, and the maximum value for the paired observations of K and M on knee width are identical with those found on the total data. For elbow width, the maximum difference in 55 pairs of observations between K and M was 2 mm., while that for all pairs of technicians was 4 mm.

Forty-five paired observations of knee breadth made by a single observer, K, gave a median difference of .4 mm., a 90th percentile equal to 1.5 mm., and a maximum difference of 2 mm.

CHAPTER VI

RELIABILITY OF MEASUREMENTS: GIRTHS

Nine circumference measurements were studied: one of the head, three of the trunk, and five of the extremities.

DEFINITIONS

Head Girth

Instrument.—Steel millimeter tape

Definition of Measurement.—Maximum girth through glabella and opisthocranion

Landmarks.—Anteriorly, the tape crossed the glabella and lay superior to the supraorbital ridges. Posteriorly, it lay at the level of the occiput which gave the maximum circumference.

Technique.—The anthropometrist extended the tape and, standing in front of the subject, placed it around the head of the subject. The tape was sufficiently extended so that when drawn around the head from back to front its zero end could be brought in contact with the other end in the vicinity of the glabella. When the tape was oriented anteriorly, it was moved up and down posteriorly until the level of the maximum girth was ascertained. The measurement was read after the anthropometrist had noted that the level of the tape was the same on both sides of the head and had applied sufficient tension to the tape to crush the hair against the head.

Chest Girth: Erect

Instrument.—Steel millimeter tape

Definition of Measurement.—Girth of the thorax at the level of the xiphoid cartilage and in a plane at right angles to the vertebral column, with the subject in erect position

Landmarks.—Anteriorly, the xiphoid cartilage of the sternum, or, in individuals lacking this cartilage, just below the inferior extremity of the body of the sternum

Posture.—The subject stood in a natural manner with head erect and with arms relaxed and held slightly away from the sides of the body in order to permit the passing of the tape around the thorax. (It was necessary to note that the position of the head was maintained, since the child tended to look down at the tape.)

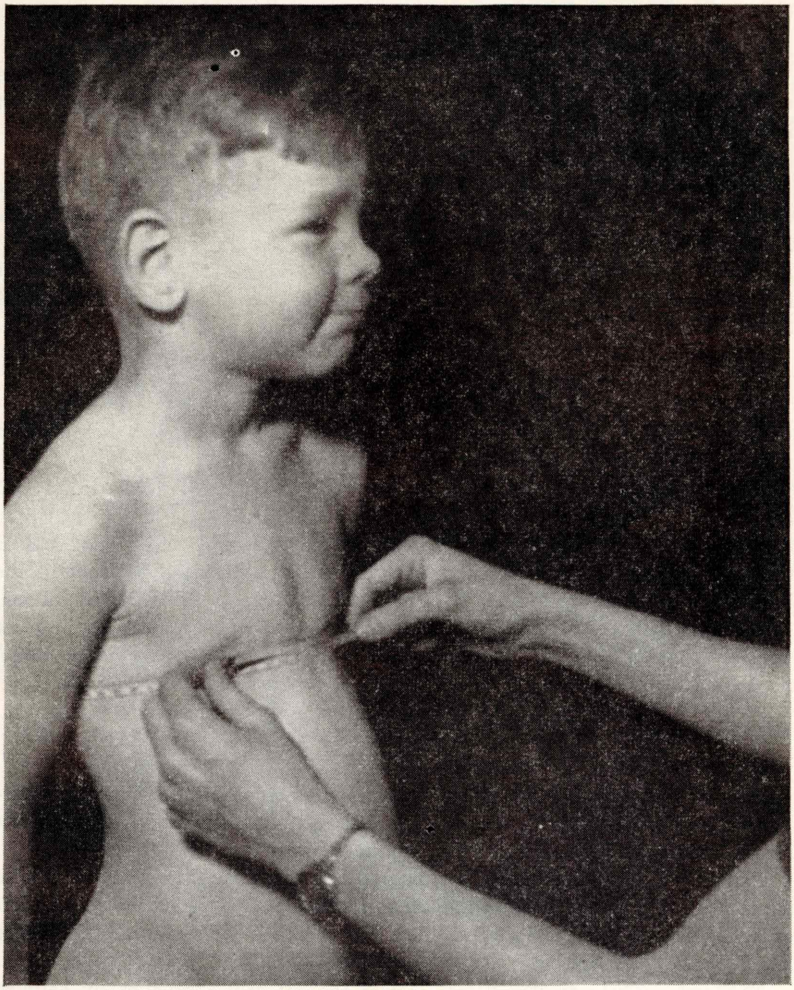

Figure 12. Chest Girth: Erect

Technique.—The anthropometrist stood in front of the subject and passed the tape around the thorax so that it crossed in front at the level of the xiphoid cartilage and lay in a plane at right angles to the vertebral column. Posteriorly, the tape always rested below the inferior angles of the scapulae. The tension applied was only sufficient to enable the tape to rest against the perimeter of the thorax without slipping. The measurement recorded was the median value during normal respiration. (Figure 12)

Chest Girth: Supine

Instrument.—Steel millimeter tape

Definition of Measurement.—Girth of the thorax at the level of the xiphoid cartilage and in a plane at right angles to the vertebral column, with the subject in recumbent position

Landmarks.—As for erect chest girth

Posture.—The subject sat on the table while the tape was placed around the thorax as for erect chest girth. Then while the anthropometrist held the tape in position, the subject lay back on the table.

Technique.—After noting that the tape circumscribed the thorax at the correct level, the anthropometrist took the measurement in the same manner as for erect chest girth.

Abdomen Girth

Instrument.—Steel millimeter tape

Definition of Measurement.—Girth of the abdomen at the level of the umbilicus and in a plane at right angles to the vertebral column

Posture.—The subject stood erect in a natural manner with head erect and with the upper extremities held slightly away from the body.

Technique.—The anthropometrist stood in front of the subject and passed the tape around the abdomen of the subject so that it crossed in front at the level of the umbilicus and lay in a plane at right angles to the vertebral column. The tension applied was only sufficient to enable the tape to remain in place without slipping. The measurement recorded was the median value during normal respiration.

Arm Girth

Instrument.—Steel millimeter tape

Definition of Measurement.—Maximum girth of the left arm below the insertion of the deltoid muscle and in a plane which crosses the "bulge" of the biceps muscle and is at right angles to the line of the humerus

Posture.—The subject stood in a natural manner with the upper extremities hanging in an extended and relaxed condition at the sides of the body. The left extremity was slightly abducted so that the arm was not in contact with the lateral wall of the thorax.

Technique.—The tape was passed around the left arm approximately halfway between the acromion and radiale. This area was explored, with the tape always at right angles to the line of the humerus, until the greatest girth was located. The record was taken with the tape in sufficiently light contact with the skin to avoid compression of the tissues.

Arm Girth: Compressed

Instruments.—Steel millimeter tape and Gulick spring attachment

Definition of Measurement.—Girth of the left arm at the same level as for arm girth, but with the spring attachment, and the end of the tape to which the attachment is applied, extended so the markers register a constant physical pressure.

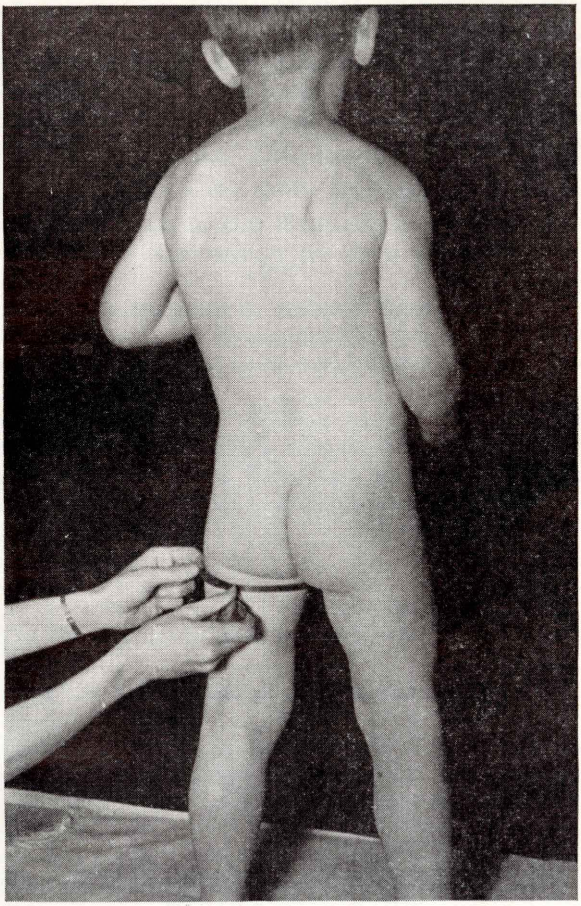

Figure 13. Thigh Girth

Posture.—As for arm girth

Technique.—As for arm girth, except that the Gulick spring attachment was fitted onto the free end of the tape and tension was then applied to this attachment until its two markers coincided.

Forearm Girth

Instrument.—Steel millimeter tape

Definition of Measurement.—Maximum girth of the left forearm below the elbow joint and the region of radiale

Posture.—The position of the subject and the orientation of the limb were the same as for arm girth. Care was taken to see that the hand was "open," that the forearm extended downward in line with the arm, and that the musculature of the forearm and hand relaxed.

Technique.—The tape was placed around the forearm immediately below the elbow-joint level and the horizontal plane of greatest girth located. The measurement was read with the tension of the tape only sufficient to contact the skin.

Thigh Girth

Instrument.—Steel millimeter tape

Definition of Measurement.—Maximum girth of the left thigh at a level immediately below the gluteal sulcus and at right angles to the long axis of the thigh

Posture.—The subject stood with feet spread about nine inches apart (so that the medial surface of one thigh was not in contact with the medial surface of the other) and with his weight equally distributed through both lower extremities.

Tehnique.—The anthropometrist moved to the left-rear of the subject and passed the tape around the left thigh at a level below the gluteal fold. On the medial aspect of the thigh, the tape passed directly over the maximum bulge in thigh contour. The tape was adjusted so that its contact was not sufficient to indent the skin. (Figure 13)

Leg Girth

Instrument.—Steel millimeter tape

Definition of Measurement.—Maximum girth of the left calf at right angles to its long axis

Posture.—The subject maintained his position as for the measurement of thigh girth.

Technique.—The tape was passed around the leg and the region of the calf explored until the largest reading at right angles to the long axis of leg was obtained. The tape was at "light contact" tension.

ANALYSIS OF RELIABILITY ON GIRTHS OF HEAD AND TRUNK

The reliability findings for the circumference of the head, chest, and abdomen are presented in Table 4. From this table it may be noted that:

1. The median difference is least for head girth (2.1 mm.) and greatest for abdomen girth (5.9 mm.). For the erect and recumbent chest girths, the medians are 3.6 and 2.9 mm.

2. A similar order is found for the 90th percentiles. The values increase in size from 6.5 mm. for head girth, through 8.7 mm. for the two chest measurements, to 16.5 mm. for abdomen girth.

3. Maximum differences obtained were 13 mm. for head girth, 15 and 16 mm. for the chest girths, and 26 mm. for abdomen girth.

4. No significant difference was found in the reliability for chest girth with the subject erect and with the subject supine.

Bayley and Davis (4) reported mean differences for observations of head girth and chest girth on infants. For head circumference, the mean difference was 2.1 mm. This figure is identical to the median found in this study. For chest girth, the mean difference obtained on

Table 4

Reliability Findings for Girths of Head and Trunk: Based on Differences Between Independent Observations of Two Technicians

Difference Interval, Millimeters	Head Girth	Chest Girth Erect	Chest Girth Supine		Abdomen Girth
	Number and Per Cent	Number and Per Cent	Number	Per Cent	Number and Per Cent
0 to 1	41	20	11	20.0	18
2 to 3	25	29	21	38.2	15
4 to 5	16	22	10	18.2	14
6 to 7	10	12	4	7.3	11
8 to 9	4	11	6	10.9	9
10 to 11	3	2	1	1.8	14
12 to 13	1	2			7
14 to 15		2			
16 to 20			2	3.6	9
21 to 25					2
26					1
Number	100	100	55		100
Median	2.1	3.6	2.9		5.9
90th percentile	6.5	8.7	8.7		16.5
Maximum	13	15	16		26

infants was 6.5 mm., or approximately double the median values reported in Table 4 for chest girth measurements on children of preschool age.

Findings are presented by Meredith (25) and by Marshall (23) on eight- and nine-year-old boys for erect chest girth. In each of these studies, the median difference is roughly 1 mm. less than the median shown in Table 4. The 90th percentile and the maximum given by Meredith are 9.1 and 20 mm. Corresponding values found in Marshall's study are 11 and 24 mm. These values found on older children are somewhat higher than those of the present study.

With respect to differences between the paired observations of K and M, it may be stated that these differences were similar to those presented in Table 4. When analyzed with regard to sign, the K-M differences showed no systematic bias for head girth, recumbent chest girth, or abdomen girth. For erect chest girth, the ratio of the median (–1.7 mm.) to its standard error was 2.2, indicating a tendency for K to obtain smaller chest girths than M.

THE COMPARABILITY OF ERECT AND RECUMBENT CHEST GIRTHS

The following tabulation shows the difference in values obtained by the two methods of taking chest girth.

Difference Interval, Millimeters	Chest Girth Supine Minus Chest Girth Erect
5 to 9	1
10 to 15	10
16 to 20	3
Number	14
Median	11
Range	6 to 20

Although there were only fourteen subjects on which the chest girth had been observed in both ways, in ten of these cases the chest girth supine was between 10 and 12 mm. greater than the erect chest girth.

ANALYSIS OF DATA ON GIRTHS OF THE EXTREMITIES

Three girth measurements were taken on the upper extremity and two on the lower extremity. Table 5 presents the data for these dimensions. The more important findings are:

Table 5

Reliability Findings for Girths of Extremities: Based on Differences Between Independent Observations of Two Technicians

Difference Interval, Millimeters	Arm Girth	Arm Girth Compressed	Forearm Girth	Thigh Girth	Leg Girth
	Number and Per Cent	Number and Per Cent	Number and Per Cent	Number and Per Cent	Number and Per Cent
0 to 1	42	44	66	23	69
2 to 3	40	40	27	34	30
4 to 5	16	8	7	19	1
6 to 7	1	7		13	
8 to 9	1	1		7	
10 to 14				4	
Number	100	100	100	100	100
Median	1.8	1.7	1.1	3.0	1.1
90th percentile	4.2	4.8	3.1	7.8	2.9
Maximum	8	9	4	14	4

1. The forearm and leg girths, judged by absolute differences, are the more reliable measurements. In each instance, the median difference is 1.1 mm. and the maximum difference is 4 mm.

2. The two arm girths—the first taken with the tape just contacting the skin, and the second taken with greater tension—show very similar distributions. The median differences are 1.8 and 1.7, the 90th percentiles are 4.2 and 4.8, and the maximums are 8 and 9 mm.

3. For girth of the thigh, the median difference is 3.0 mm., the 90th percentile 7.8, and the maximum is 14 mm.

With reference to the measurement of arm girth, it may be noted that the use of a standard spring tension does not increase the reliability of the observations. However, with regard to certain research problems (e.g., the estimation of physical status) the two methods may serve as supplementary rather than alternative procedures.

Meredith (26) presented reliability findings for a single technician in taking the arm, forearm, thigh, and leg girth on eight- and nine-year-old boys. In general, his reliability constants are equal to or slightly less than those reported here. Marshall's reliability constants for differences between two technicians on this group of children are all higher than those found in this study. This is especially true for the maximum differences observed. For arm, forearm, and leg girth his maximum values are approximately twice the magnitude of the corresponding values found here on preschool children.

For forearm, thigh, and leg girth, the differences in approximately fifty observations made by K and M are practically identical to those found for all pairings taken together. The medians and the 90th percentiles in no instance differ more than .3 mm. and the maximums are the same as for the total data. Observations on the arm girth yield a median difference identical with that found for the total group, but the maximum difference is 2 mm. less. For arm girth compressed, the median and maximum are the same for K-M as for all pairs of technicians, while the 90th percentile is 1 mm. less.

No systematic bias was found between the observations of K and M for the arm girths, or for leg girths. A tendency was found for M to obtain larger values for thigh girth and smaller values for forearm girth. For thigh, the median difference, taking account of the direction of the differences, was −1.3 mm. The ratio of this median to its standard error is 1.78 and, consequently, is not statistically significant. However, the forearm median of .8 mm. is significantly different from zero, the ration being 3.24.

CHAPTER VII

RELIABILITY OF MEASUREMENTS: SKIN AND SUBCUTANEOUS TISSUE

The reliability of five measurements of skin and subcutaneous tissue was investigated. Four of these measurements were made on the trunk and one on the upper extremity.

DEFINITIONS

Thickness of Skin and Subcutaneous Tissue

Instrument.—Small spreading calipers especially devised for this purpose by the American Child Health Association

Technique.—The objective was to measure the thickness of a complete double layer of skin and subcutaneous tissue without including any underlying muscle tissue. The anthropometrist placed the thumb and forefinger of his left hand over the region to be measured in such a manner that the two digits were in an opposable relationship and the distance between them was approximately 30 to 40 mm. For the measurements on the chest front and chest back and above the iliac crest, the interdigit plane was at right angles to the long axis of the body of the subject. In the case of the measurement at arm back, the interdigit plane was at right angles to the long axis of the hanging extremity. For the measurement on the abdomen, the interdigit plane was parallel to the long axis of the body.

After the left thumb and forefinger of the anthropometrist had been oriented in accordance with these specifications, the two digits were moved directly toward each other so that a complete double layer of skin and subcutaneous tissue was grasped between them. This fold was held with moderate firmness while the instrument was applied to it with the right hand and the record read off.

Location of Measurements

Arm Back.—This measurement was taken on the left arm midway between the acromion and radiale over the triceps muscle. The arm was extended downward in a relaxed conditon.

Thorax Front.—The calipers were applied at the level of the xiphoid cartilage immediately below the left thelium.

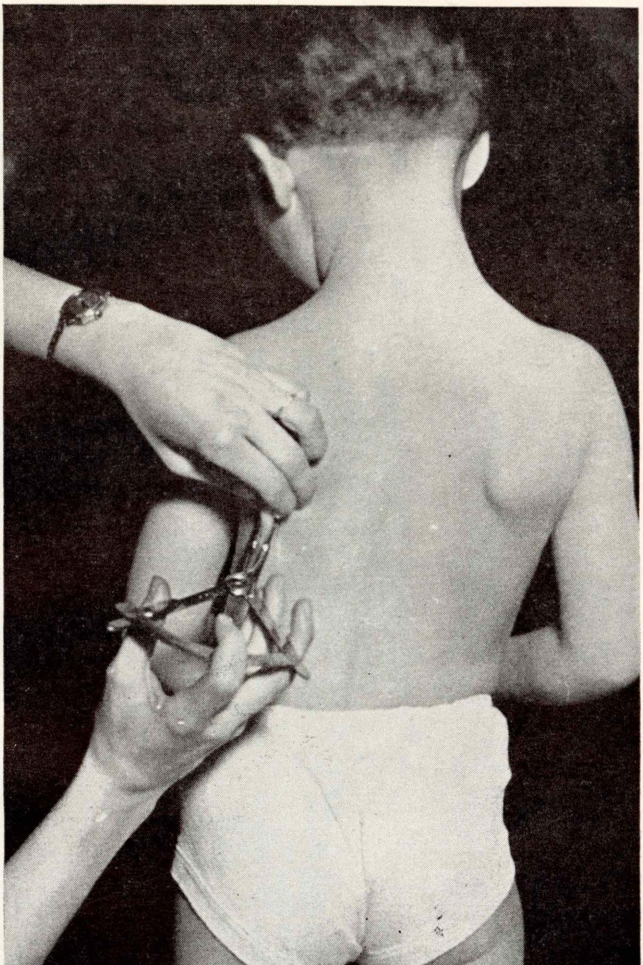

Figure 14. Thickness of Skin and Subcutaneous Tissue: Thorax Back

Thorax Back.—This measurement was taken at the region below and slightly lateral to the inferior angle of the left scapula and in the transverse plane of the xiphoid cartilage. (Figure 14)

Supra-iliac.—The calipers were applied immediately superior to the crest of the left ilium in a line vertical with the left axilla.

Abdomen.—This measurement was taken at the front of the abdomen, slightly above and about 2 cms. to the left of the umbilicus. The double layer was measured as a transverse fold, at right angles to the direction for the other four measurements. (Figure 15)

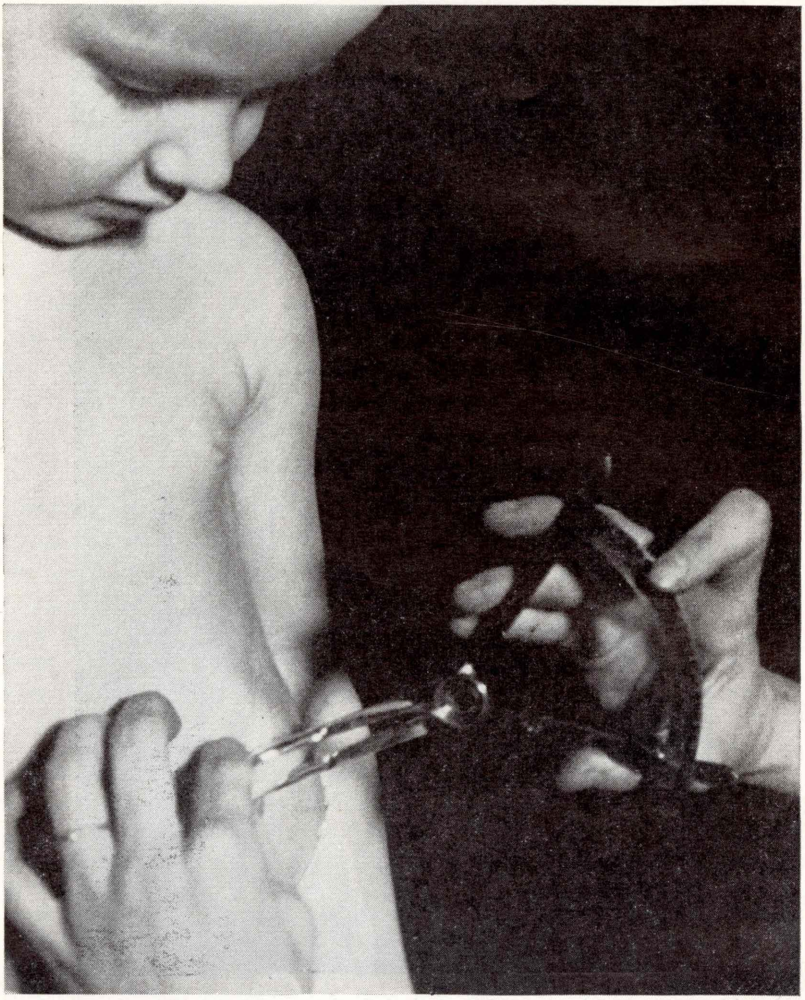

Figure 15. Thickness of Skin and Subcutaneous Tissue: Abdomen

ANALYSIS OF DATA ON MEASURES OF SKIN AND SUBCUTANEOUS TISSUE

The distributions and reliability constants for the differences between observations by all pairs of observers are shown in Table 6. Selected findings on these measurements are:

Table 6

Reliability Findings for Measurements of Skin and Subcutaneous Tissue: Based on Differences Between Independent Observations of Two Technicans

Difference Interval, Millimeters	Thickness of Skin and Subcutaneous Tissue*				
	At Arm Back	At Thorax Front	At Thorax Back	Above the Iliac Crest	At Abdomen
	Number and Per Cent	Number and Per Cent	Number and Per Cent	Number and Per Cent	Number and Per Cent
0	36	50	34	26	49
1	39	39	47	37	40
2	19	7	19	21	10
3	4	4		12	
4	2			1	1
5 to 9				3	
Number	100	100	100	100	100
Median	.9	.5	.8	1.2	.5
90th percentile	2.3	1.6	2.0	3.0	1.6
Maximum	4	3	2	9	4

*Values in this table are based on measurement of a double layer of tissue. Differences in the estimation of a single layer would be equal to one-half these values.

1. The medians range in size from .5 mm. for measures at chest front and abdomen to 1.2 mm. for supra-iliac.

2. The 90th percentile for supra-iliac tissue measurement is 3 mm., approximately 2 mm. for chest back and arm back, and 1.6 mm. for the two other measurements.

3. Not one of the differences of the one hundred paired observations for chest back measure was greater than 2 mm. The maximum for chest front was 3 mm.; for abdomen and arm back 4 mm. The largest difference in observations for supra-iliac was over twice that for any of the other four measurements; thus in all comparisons, it stands as the least reliable of the measures of skin and subcutaneous tissue studied.

The study by Meredith (25) and the study by Marshall (23) include data for measurements taken at arm back, thorax back, and supra-iliac. In general, the constants found in these two studies are somewhat below these reported here. Findings of each of these investigators reveal that the measurement above the iliac crest was the least accurately determined and the measurements taken at the back of the thorax were the most accurate of the three. This ranking corresponds to the order of absolute reliability found in this study.

With regard to measurements made at the chest front, chest back, abdomen, and above the iliac crest, the differences based on observations of K and M are closely similar to those for the total pairings of technicians. In the case of measurement at arm back, each of the constants was slightly less for K-M than for all pairs together: the median was .7 mm., the 90th percentile 1.8 mm., and the maximum 3 mm.

No systematic bias was revealed between measurements taken by K and M at the front and back of the thorax or on the abdomen. K tended to obtain larger readings on the arm back. The critical ratio of the median algebraic difference to its standard error was 3.40. The median difference of –.8 mm. for supra-iliac measurements divided by its error gave a ratio of 2.17—indicating a tendency, not definitely reliable, for M to make observations which were larger than those made by K.

CHAPTER VIII

RELIABILITY OF MEASUREMENTS IN RELATION TO THEIR MAGNITUDE AND RATE OF GROWTH

Presentation of results in the preceding chapters has been entirely in terms of absolute differences. It is also important to study differences obtained for a given measurement in relation to the size of the dimension and to the rate of growth of the dimension.

Table 7

Reliability of Measurements Relative to Mean Size

Dimension	Median Difference, Millimeters	Maximum Difference, Millimeters	Mean Size, Millimeters	Median Difference / Mean Size x 100	Maximum Difference / Mean Size x 100
Stature	2.6	10	1068	.24	.94
Sitting height	2.4	12	605	.40	1.98
Head girth	2.1	13	502	.42	2.59
Head breadth	.6	3	138	.43	2.17
Hip width	.8	6	172	.47	3.49
Leg girth	1.1	4	226	.49	1.77
Forearm girth	1.1	4	171	.64	2.34
Chest girth	3.6	15	538	.67	2.79
Head length	1.2	6	173	.69	2.37
Total leg length	3.9	17	463	.84	3.67
Thigh girth	3.0	14	335	.90	4.18
Chest width	1.8	12	187	.96	6.42
Total arm length	4.3	22	444	.97	4.95
Knee width	.7	4	70	1.00	5.71
Arm girth	1.8	8	171	1.05	4.68
Abdomen girth	5.9	26	541	1.09	4.81
Elbow width	.5	4	45	1.11	8.89
Shoulder width	2.8	20	239	1.17	8.37
Chest depth	2.1	11	137	1.53	8.03
Hand length	2.6	17	125	2.08	13.60
Abdomen depth	3.1	13	132	2.35	9.85
Arm length	5.6	17	181	3.09	9.39
Forearm length	5.3	17	138	3.84	12.32
Tissue:					
At abdomen	.5	4	7.5	6.67	53.33
At arm back	.9	4	13.2	6.82	30.30
At chest front	.5	3	7.1	7.04	42.25
At chest back	.8	2	8.5	9.41	23.53
Above iliac crest	1.2	9	10.6	11.32	84.91

RELIABILITY OF MEASUREMENTS RELATIVE TO MEAN SIZE

The mean size of each dimension was taken as a value midway between the mean of the initial observations and the mean of the second observations for the entire group of children used. The obtained mean for each of the twenty-eight different dimensions studied is given in the third column of Table 7. This table also reproduces for each dimension the median and maximum difference between independent observations by two technicians, and shows both the median and maximum differences expressed as percentages of the mean size of the dimension to which they apply.

Selected findings relative to the mean magnitude of the measurements are:

1. The median difference for repeated observations of stature is less than one-fourth of 1 per cent of the average size of this dimension, while the median difference for thickness of skin and subcutaneous tissue above the iliac crest is more than 11 per cent of its mean size.

2. One-half of the twenty-eight dimensions gave a median difference which was 1 per cent or less of the size of the dimension.

3. The measures of skin and subcutaneous tissue are shown to be the least reliable when the median differences are related to the magnitude of these measures. In contrast, the most reliable measurements include stature, sitting height, head girth, head breadth, hip width, and leg girth.

4. When maximum differences of each dimension are expressed as percentages of their respective dimensions, stature again yields the lowest percentage. The maximum difference of 10 mm. is less than 1 per cent of the mean magnitude of stature for the entire group of children.

5. Approximately one-half of the maximum differences are more than 5 per cent of the magnitude of the corresponding dimensions.

6. For the measures of thickness of skin and subcutaneous tissue, the maximum differences are all more than 20 per cent of the mean size of these measures.

Table 7 may be directly compared with similar tables given in the studies by Meredith (25) and by Marshall (23). It is found that the values of the present study are slightly above those of Meredith and close to those of Marshall. It might be expected that the values of the present study would exceed those found on elementary school children since the mean size of the dimensions is smaller for preschool children. However, comparison of the values in this study with those of Marshall (where the observations were made by two different technicians, and are thus strictly comparable with those reported here) shows approximately one-half of the ratios of the median difference to mean size to be slightly smaller and approximately one-half slightly larger. In general, the order of relative reliability for preschool children is similar to that for eight- and nine-year-old boys.

COEFFICIENTS OF CORRELATION

Product-moment correlation coefficients between the paired observations for each dimension were computed. The number of cases for each dimension is the same as the number of differences studied. The following tabulation gives the coefficients obtained:

Dimension	Correlation	Dimension	Correlation
Stature	.997	Head girth	.955
Leg girth	.993	Shoulder width	.952
Sitting height	.987	Abdomen girth	.949
Hip width	.987	Elbow width	.945
Thigh girth	.986	Tissue at abdomen	.942
Total leg length	.985	Chest depth	.928
Forearm girth	.984	Tissue at chest front	.906
Total arm length	.981	Abdomen depth	.906
Knee width	.980	Arm length	.905
Arm girth	.979	Forearm length	.888
Head breadth	.979	Tissue at chest back	.884
Chest girth	.971	Tissue above iliac crest	.873
Head length	.967	Tissue at arm back	.857
Chest width	.958	Hand length	.848

While these coefficients are not considered essential or directly pertinent to the present problem, they are included as supplementary information which might be of interest to other investigators.

RELIABILITY OF MEASUREMENTS RELATIVE TO GROWTH INCREMENTS

Baldwin (1927) stated,

> "The interval between measurements is a question of vital importance . . . If the child has reached the age when the increment of growth is small from month to month, nothing is gained by measuring that trait at short intervals . . . It seems reasonable to presume that even two-week intervals are practical for infants, since general growth is rapid. For the infant, weight may be taken daily, or at shorter intervals. For the

school child the repetition of measurements should be less extensive and the time interval greater. At the Research Station, preschool children are measured once a month . . . older children are measured at six-month intervals . . ." (2, p. 6)

Recognition has thus been given to the importance of adjusting the interval between examinations to the rate of growth of the child. During infancy, when growth is rapid, the intervals between observations should be shorter than during periods when growth rates are declining. Age of subjects is, clearly, one factor to be considered in establishing programs of measurement. A second factor, not indicated by Baldwin, is the rate of growth of different dimensions. While, in general, growth is rapid during infancy in contrast to later years, it is not equally rapid for all dimensions. Frequency of measurement, then, depends on the age of the subject and on the specific dimensions to be studied. Both aspects of the problem are amenable to experimental investigation.

Table 8

Reliability of Measurements Relative to Mean Annual Increments: Reliability Criterion, Annual Gain, Ratio of Gain to Criterion, and Proposed Frequency of Measurement for Each of Twenty Dimensions

Dimension	90th Percentile of Differences, Millimeters	Mean Annual Increment, Millimeters	Estimated Frequency (times per year)	Proposed Calendar Frequency
Stature (6) (25)	6.5	65.2	10.0	Bimonthly
Hip width (6)	2.1	8.0	3.8	Quarterly (?)
Sitting height (29)	7.4	26.0	3.5	Every four months
Leg girth (27)	2.9	9.3	3.2	Every four months
Total arm length (41)	10.2	30.5	3.0	Every four months
Thigh girth (27)	7.8	14.6	1.9	Semiannually
Chest girth (6)	8.7	14.5	1.7	Semiannually (?)
Forearm girth (27)	3.1	4.9	1.6	Semiannually (?)
Knee width (6) (25)	1.7	2.4	1.4	Annually
Arm girth (27)	4.2	5.2	1.2	Annually
Shoulder width*	8.2	10.0	1.2	Annually
Arm length (41)	12.1	13.6	1.1	Annually
Forearm length (41)	10.3	10.8	1.0	Annually
Chest width (6) (28)	4.5	4.7	1.0	Annually
Hand length (41)	7.3	6.1	.8	Annually
Head length (6) (25)	3.1	2.1	.7	Biannually
Head breadth (6) (25)	2.0	1.4	.7	Biannually
Head girth (21)	6.5	3.8	.6	Biannually
Chest depth (6) (28)	4.8	2.5	.5	Biannually
Elbow width (6) (25)	1.5	.8	.5	Biannually

*Unpublished data, Iowa Child Welfare Research Station

Meredith (1936) presented an experimental approach to the problem of frequency of measurement from the standpoint of specific dimensions. In his study, "the assumption was made that seriatim observations are only profitably made at such intervals as the mean increment of growth equals or exceeds 90 per cent of the reliability differences." (25, p. 270) Upon this basis a table was constructed giving an estimated frequency of measurement of twelve dimensions for eight- and nine-year-old boys. The reliability criterion employed was based on repeated measurements by a single technician. Marshall (23) (1937), by the same technique, reported a similar table of proposed frequency of measurement. In his study, the findings were based upon differences between two technicians. Both estimates were for children of school age and both showed that whereas certain dimensions were profitably remeasured at quarterly intervals, other dimensions were only profitably remeasured at annual intervals.

The present study provides similar data for measurements on preschool children. For each of twenty dimensions, Table 8 gives the 90th percentile of the differences, the mean annual increment for the dimension, and the estimated frequency of measurement of children from three to six years of age. The estimated frequency in times per year was obtained by dividing the annual increment by the reliability criterion, the 90th percentile.

Mean annual increments were determined from various studies, largely on Iowa children. The reference number following each dimension gives the source for the increment. The mean annual increment for each dimension was derived by obtaining the difference between the mean at three years and the mean at six years and dividing this difference by three. In certain instances, this procedure gave more stable increment values than could have been obtained from a selected one-year interval.

The estimated frequency of seriatim measurement is one which implies that only one measurement value for each dimension is to be obtained at each examination. Since accurate individual curves are one of the primary needs in the study of physical growth and since some measurement differences are rather large, it seems advisable that, where resources will permit, two independent observations for a dimension be made at each examination.

By employing the mean of two independent observations rather than a single observation, the reliability of the recorded value is increased.

Table 9

An Anthropometric Schedule for Children of Preschool Age*

Dimension	September	October	November	December	January	February	March	April	May	June	July	August
Stature	X	X	X	X	X	X	X	X	X	X	X	X
Sitting height	X		X		X		X		X		X	
Hip width	X		X		X		X		X		X	
Leg girth	X		X		X		X		X		X	
Total arm length	X		X		X		X		X		X	
Thigh girth	X				X				X			
Chest girth	X				X				X			
Forearm girth	X				X				X			
Knee width	X				X				X			
Shoulder width			X						X			
Arm girth			X						X			
Arm length			X						X			
Forearm length			X						X			
Hand length			X						X			
Chest width			X						X			
Chest depth	X											
Head length	X											
Head breadth	X											
Head girth	X											
Elbow width	X											

*Each dimension to be measured independently by two trained technicians and the mean of their readings to be taken for the record

With reliability increased, seriatim examination can be profitably made at more frequent intervals.

Table 9 gives a suggested program of measurement of children of preschool age. The frequency of examination is greater than (not quite double) that shown in Table 8, the assumption being that when this schedule is used, two independent observations are made and their mean used as the value reported on the permanent record.

The significance of Table 8 and Table 9 is considered to lie not so much in the exact estimated frequency of measurement for each dimension, as in the fact that from these tables the relative frequency of measurement for specific dimensions can be determined. For example, stature may be profitably taken over three times as frequently as the length of the upper extremity, ten times as frequently as the width of the chest, and fourteen times as often as the three head measurements studied.

With the aid of these tables, it is possible to arrange other schedules adapted to particular administrative and research problems. If, for example, one were interested in studying individual curves of indices, the schedule could be adjusted in such a way that the component dimensions constituting the index would be taken at the same examination periods. The essential contribution of this section is the presentation of experimental data from which one can construct a schedule differentiated on the basis of reliability and rate of growth.

CHAPTER IX

SUMMARY AND CONCLUSIONS

Critical evaluation of the available experimental studies reveals that further research is especially needed with regard to the physical measurement of subjects of preschool age.

The present study was formulated to apply specifically to the child of preschool age and to avoid several of the shortcomings found characteristic of earlier studies. It was projected in the belief that it would yield a contribution useful for: (1) appraising current assumptions regarding anthropometric reliability, (2) evaluating research data on physical growth of the child, (3) establishing examination schedules, (4) assisting in the choice of anthropometric method, and (5) training technicians.

Thirty-five physical measurements were studied. These included ten length measurements, eleven transverse and antero-posterior diameters, nine circumferences, and five measures of thickness of skin and subcutaneous tissue. Of these thirty-five measurements, twenty-nine were measurements of different dimensions, while the remaining six were of dimensions included in the twenty-nine, but were taken with the subject in a second position. That is, stature, sitting height, length of lower extremities, chest width, chest depth, and chest girth were taken with the subject both in an erect and a recumbent position.

One hundred and thirty-one different children between three and six years of age served as subjects.

Reliability values were obtained by two procedures. The major procedure was that of obtaining the difference between observations made by two technicians. This was employed with all the measurements studied. A supplementary procedure was used for five of the measurements. Here the differences between two observations by the same technician were secured.

For a majority of the thirty-five measurements, there were 100 paired observation values. In all, the data consisted of 3,643 differences between independent observations of two technicians and 221 differences based on two observations by the same technician.

The differences obtained between the various pairs of observers, all values being considered positive, were distributed. The median and the 90th percentile were then computed for the distribution pertaining to each dimension. Certain general conclusions from the findings on absolute differences between successive, independent observations are:

1. For major length measurements of the body, stature yields smaller differences, in general, than does sitting height. Length of lower extremities, determined as stature minus sitting height, shows the largest differences.

2. The lengths measured on the upper extremity, ranked in order of size of differences from those showing least to those showing greatest differences, are: hand length, total arm length, forearm length, and arm length.

3. Head length gives larger differences, on the whole, than does head breadth.

4. Concerning diameters of the trunk, it is found that hip width is most reliable in terms of absolute differences, chest measurements rank next, while shoulder and abdomen diameters are least reliable.

5. When judged on the basis of absolute differences, girth of the head is more reliable than girth of thorax, which in turn is more reliable than abdomen girth.

6. The girths of forearm and leg, observed independently, show smaller differences than girth of the arm and of the thigh.

7. Five measurements of the thickness of skin and subcutaneous tissue varied in the agreement between observations. Differences at chest front and abdomen are smaller than those of the supra-iliac measurement.

8. No marked differences are found in reliability of measurements taken with the subject in two different positions. However, stature recumbent and sitting height erect are slightly more reliable than stature erect and sitting height recumbent. Chest depth supine yields a lower maximum difference than chest depth erect.

Differences derived from the observations for the two technicians with the greatest frequency of pairings were examined for bias in technique. A tendency was found for one examiner to obtain slightly higher (or lower) values in the measurement of seven of the dimensions.

Entirely separate analysis was made of the differences obtained between successive independent observations of a single technician. In general, these differences were smaller than those between two different technicians.

For the six dimensions taken with the subject in two different positions, comparison was made to determine the difference in observed values as well as the difference in reliabilities. The major absolute magnitude differences obtained were as follows:

1. Supine measurement of stature and sitting height, in general, was found to yield observations 7 mm. greater than stature and sitting height taken in the erect position.

2. On the assumption of the supine position, chest width was found to show a median increase of 5 mm.; chest depth a decrease of approximately one-half this amount. Chest girth measured approximately 11 mm. greater.

All the foregoing analysis was in terms of absolute differences. In addition, the median and maximum differences found for each dimension were related to the mean magnitude of that dimension, yielding a relative measure of reliability. From these ratios of the median difference for each dimension to the mean size of that dimension, selected findings are:

1. The ratio for stature is approximately one-fourth of 1 per cent, and fourteen of the twenty-eight dimensions studied revealed percentages equal to or less than 1 per cent.

2. Ratios for the various measures of skin and subcutaneous tissue were markedly above those for the other dimensions, ranging between 6 and 12 per cent.

Finally, reliability constants were related to mean annual increments of growth. The object here was to estimate the approximate frequency at which it would be profitable to repeat measurements on a given individual, assuming reliability comparable to that of the present study. Judged on the basis of the relation of reliability to the rate of growth, it was found that:

1. Stature could be profitably taken ten times a year during the preschool period.

2. Three to four times a year was the estimated frequency for hip width, sitting height, leg girth, and total arm length.

3. The three head measurements, chest depth, and elbow width should be taken only every two years, while the remaining measures could be profitably measured annually and in a few cases semiannually.

Based on this same relative frequency, a schedule was constructed allowing for more frequent measurement. This proposed schedule was presented for measurement of various dimensions of preschool children when dimensions are measured independently by two technicians and the mean of these observations used as the recorded value.

With regard to the direction of future research in this area, it appears that measurement of other dimensions at various age periods should be investigated. Also, other techniques of measuring the same dimensions should be explored with a view to obtaining more reliable techniques. For example, by placing the child against a board or in the supine position, it might be possible to reduce some of the large differences in the measurement of shoulder width which occur when measurement is made with the subject standing free.

Study of the reasons for small but rather consistent bias in technique, such as those found here in arm length measurements, may be expected to reveal possibilities of more objective and precise definitions.

REFERENCES

1. Bakwin, Harry, and Bakwin, Ruth Morris: Body build in infants. I. The technique of measuring the external dimensions of the body of infants. J. Clin. Investig., 1931, 10, 369-375.

2. Baldwin, Bird T.: Anthropometric measurements of children and adults. [Discussion] National Research Council, Division of Anthropology and Psychology, Committee on Child Development. Second Conference on Research in Child Development. Washington, D. C., 1927. Pp. 123. (p. 2-8) (p. 25) (Mimeographed)

3. Bayley, Nancy: Growth changes in the cephalic index during the first five years of life. Human Biol., 1936, 8, 1-18.

4. Bayley, Nancy, and Davis, Frank C.: Growth changes in bodily size and proportions during the first three years: A developmental study of sixty-one children by repeated measurements. Biometrika, 1935, 27, Parts 1 and 2, 26-87.

5. Boyd, Edith: The experimental error inherent in measuring the growing human body. Amer. J. Physical Anthropol., 1929, 13, 389-432.

6. Boynton, Bernice: The physical growth of girls: A study of the rhythm of physical growth from anthropometric measurements on girls between birth and eighteen years. Univ. Iowa Stud., Stud. in Child Welfare, 1936, 12, No. 4, Pp. 105.

7. Carey, Thomas F.: An anthropometric measuring board. Child Develop., 1936, 7, 57-59.

8. Carey, Thomas F.: The relation of physical growth to developmental age in boys. (Center for Research in Child Development, 1936, 1, No. 2) Washington, D. C.: Catholic University of America, 1936. Pp. 119.

9. Cates, H. A., and Goodwin, J. C.: The twelve-day old baby. Human Biol., 1936, 8, 433-450.

10. Davenport, C. B.: Guide to anthropometry and anthroposcopy. Baltimore, Md.: Waverly Press, 1927. Pp. 53.

11. Davenport, C. B.: Some principles of anthropometry. Amer. J. Physical Anthropol., 1937, 33, 91-99.

12. Davenport, C. B., Steggerda, Morris, and Drager, William: Critical examination of physical anthropometry on the living. Proc. Amer. Acad. Arts & Sci., 1934, 69, 265-284.

13. Dearborn, Walter F.: [Discussion] National Research Council, Division of Anthropology and Psychology, Committee on Child Development. Second Conference on Research in Child Development. Washington, D. C., 1927. Pp. 123. (p. 26) (Mimeographed)

14. Faber, Harold K.: A critique of the Pirquet feeding system. Amer. J. Dis. Child., 1923, 25, 339-349.

15. Gray, Horace: Against measuring the chest girth at rest. Amer. Med. Assoc. J., 1922, 79, 349-350.

16. Gray, Horace: Sitting-height and stem-length in private school boys. Amer. J. Dis. Child., 1922, 23, 406-418.

17. Gray, Horace, and Ayres, J. G.: Growth in private school children: With averages and variabilities based on 3,110 measurings on boys, and 1,473 on girls, from the ages of one to nineteen years. Chicago, Ill.: University of Chicago Press, 1931. Pp. xv, 282.

18. Hejinian, Lucea, and Hatt, Elise: The stem-length: Recumbent-length ratio as an index of body type in young children. Amer. J. Physical Anthropol., 1929, 13, 287-307.

19. Hrdlička, Aleš: Anthropometry. Philadelphia, Pa.: Wistar Institute of Anatomy and Biology, 1920. Pp. vi, 163.

20. Hrdlička, Aleš: Anthropometry. The Anthropometric Committee of the American Association of Physical Anthropologists. Note 2. (Jan. 29) Amer. J. Physical Anthropol., 1936, 21, 287-291.

21. Iowa Child Welfare Research Station: Physical traits of young children: Report of measurements of one hundred and fifty boys and one hundred and sixty-seven girls from three to six years of age. Amer. J. Dis. Child., 1929, 38, 541-546.

22. Lincoln, Edward A.: The reliability of anthropometric measurements. Ped. Sem. & J. Genet. Psychol., 1930, 38, 445-450.

23. Marshall, Everett L.: The objectivity of anthropometric reliabilities taken on eight- and nine-year-old white males. Child Develop., 1937, 8, 249-256.

24. Martin Rudolf: Lehrbuch der Anthropologie. In systematischer Darstellung: Mit besonderer Berücksichtigung der anthropologischen Methoden. Erster Band: Somatologie. Jena: Gustav Fischer, 1928. Pp. xvii, 578.

25. Meredith, Howard V.: The rhythm of physical growth: A study of eighteen anthropometric measurements on Iowa City white males ranging in age between birth and eighteen years. Univ. Iowa Stud., Stud. in Child Welfare, 1935, 11, No. 3, Pp. 128.

26. Meredith, Howard V.: The reliability of anthropometric measurements taken on eight- and nine-year-old white males. Child Develop., 1936, 7, 262-272.

27. Meredith, Howard V., and Boynton, Bernice: The transverse growth of the extremities: An analysis of girth measurements for arm, forearm, thigh, and leg taken on Iowa City white children. Human Biol., 1937, 9, 366-403.

28. Meredith, Howard V., and Knott, Virginia B.: Changes in body proportions during infancy and the preschool years: I. The thoracic index. Child Develop., 1937, 8, 173-190.

29. Meredith, Howard V., and Knott, Virginia B.: Changes in body proportions during infancy and the preschool years: III. The skelic index. Child Develop., 1938, 9, 49-62.

30. Montessori, Maria: Pedagogical anthropology. Trans. from the Italian by Frederic Taber Cooper. New York: Frederick A. Stokes, 1913. Pp. xi, 508. (p. 67-106)

31. Palmer, Carroll E.: The relationship of erect body length to supine body length. Human Biol., 1932, 4, 262-271.

32. Roberts, Charles: Manual of anthropometry: A guide to the physical examination and measurement of the human body. . . . London: J & A Churchill, 1878. Pp. xxiv, 118.

33. Scammon, Richard E.: [Discussion] National Research Council, Division of Anthropology and Psychology, Committee on Child Development. Second Conference on Research in Child Development. Washington, D. C., 1927. Pp. 123. (p. 23) (Mimeographed)

34. Scammon, Richard E., and Calkins, Leroy A.: The development and growth of the external dimensions of the human body in the fetal period. Minneapolis, Minn.: University of Minnesota Press, 1929. Pp. xxiii, 367.

35. Schultz, Adolph H.: The technique of measuring the outer body of human fetuses and of primates in general. Contributions of Embryology, No. 117. Reprinted from Cont. to Embryol., Carnegie Inst. Wash., 1929, Pub. No. 394, 213-257.

36. Todd, T. Wingate: The reliability of measurements based upon subcutaneous bony points. Amer. J. Physical Anthropol., 1925, 8, 275-279.

37. Todd, T. Wingate: [Discussion] National Research Council, Division of Anthropology and Psychology, Committee on Child Development. Second Conference on Research in Child Development. Washington, D. C., 1927. Pp. 123 (p. 20-22) (Mimeographed)

38. Todd, T. W.: Types and indices of growth in childhood. National Research Council, Division of Anthropology and Psychology, Committee on Child Development, Third Conference on Research in Child Development, Part I. Critical Reviews, 1929. Pp. 227. (p. 25-38)

39. Todd, T. Wingate, and Lindala, Anna: Dimensions of the body: Whites and American Negroes of both sexes. Amer. J. Physical Anthropol., 1928, 12, 35-119.

40. Todd, T. Wingate, and Pyle, S. Idell: Effects of maceration and drying upon the vertebral column. Amer. J. Physical Anthropol., 1928, 12, 303-319.

41. Wallis, Ruth Sawtell: How children grow: An anthropometric study of private school children from two to eight years of age. Univ. Iowa Stud., Stud. in Child Welfare, 1931, 5, No. 1, Pp. 137.

42. Whitacre, Jessie: Standing heights of school children as determined by two techniques. Amer. J. Physical Anthropol., 1934, 18, 457-465.

43. White House Conference on Child Health and Protection: Section I. Committee A. Growth and Development. Preliminary Report of Division V. Determination of Physical Status. [1932] Pp. 46.

44. Wilson, Charles A.: [Discussion] National Research Council, Division of Anthropology and Psychology, Committee on Child Development, Third Conference on Research in Child Development, 1929, Part II. Washington, 1929. Pp. 351. (p. 255)

Obtainable from the Department
of Publications

Paper bound, $1.00

Cloth bound, $1.35