INTRODUCTION

Radiographic observations of bone density are used as indications of mineral density in a variety of studies, such as the following:

1. Assessment of the status and progress of persons suffering from diseases that affect mineral deposition and removal.
2. Surveys of groups of human subjects in a search for factors that may affect mineral content, such as age and socio-economic conditions.
3. Nutritional experiments on human subjects.
4. Experiments on laboratory animals and livestock in which it is desirable to obtain serial observations on the bones of the same animal.

Visual assessment of the density of any X-ray shadow can be trusted only when differences are gross, because (a) the eye's interpretation is affected by the density of neighboring shadows, and because (b) correction for the overall density of the film is coarse and subjective. For objectivity and more precise quantification, therefore, a photoelectric densitometer is commonly used, and the reading on the bone shadow is corrected by reference to a standard object, such as an aluminum wedge, exposed

on the same film, the correction being often simply conversion of the bone reading into the equivalent thickness of the standard object.

The use of a metal wedge as a standard can be traced back to the beginning of this century, and since then many workers have employed metal wedges or other objects (such as bone slabs or ivory wedges). But bone densitometry research has been sporadic, with the principal exception of the studies initiated by Pauline Beery Mack[23] at Pennsylvania State College about thirty years ago, and since continued at that institution (now Pennsylvania State University) as well as by Dr. Mack and her collaborators at Texas Women's University.

The chief criticism of densitometry has come from radiologists, who know that gross demineralization by a tumor, in the vertebral column or the femur, can be masked by overlying soft tissue. This skepticism was supported by nitric acid demineralization of human cadaver bones and the use of a wax phantom to represent soft tissue—experiments which led Lachman and Whelan[20] to conclude that only under the most favorable circumstances could one expect to diagnose decalcification when it is less than 20 per cent. At the other extreme, a decade ago some nutritionists were so impressed by data produced at Pennsylvania State College that they suggested that X-ray densitometry might come to replace calcium-balance determinations in human subjects. This hope has not been gratified, but the use of bone densitometry for various purposes has increased greatly and so, therefore, has the number of applications for grants in aid of research projects which employ that technique.

Because of the conflicting opinions regarding the value of the method and discordant results obtained by different investigators, one of the grant-application reviewing committees of the U.S. Public Health Service, the Nutrition Study Section of the National Institutes of Health, sponsored in 1959 a two-day Workshop on Bone Densitometry, at which were assembled experts and others interested in the subject. Although the primary purpose of the workshop was enlightenment of the study

section members, the transcript[11] issued by the Chairman of the workshop, Stanley M. Garn of the Fels Research Institute, is a valuable source of information regarding the experiences of the participants, their conflicting opinions and unsolved problems. Of even more general value is the Annotated Bibliography on Bone Densitometry[5] compiled for the workshop by Arthur K. Clark at the same institute. Since that compilation, supplemented by contributions from other workers, has appeared in print,[10] there is no need for the present report to include a historical background or extensive bibliography.

PROBLEMS IN DENSITOMETRY

After contemplating a radiological physicist's description of the factors that determine the density of an X-ray shadow, we might despair of ever finding a biologically or medically useful expression of that density if we did not recall the fact that many useful quantities—readings on clinical thermometers and sphygmomanometers, for example—are the resultants of a multitude of factors, some known and many unknown. Those who practice bone densitometry are sustained by the knowledge that the density of a bone shadow is, in the main, due to the mineral content of the bone; but they are apt to be discouraged when they discover, in the transcript of the densitometry workshop, that those who have developed the most complex apparatus are not clear about the meaning of such terms as "bone density" and "mass coefficient."

It may be helpful, therefore, to return to a simple empirical point of view, to visualize a simple instrument (such as the Photovolt transmission densitometer used in this laboratory), to keep clearly in mind the gap between X-ray bone shadow density and mineral density, to enumerate the factors that cause this gap, and to examine methods of bridging it. These

methods must be scrutinized for hidden assumptions and other defects.

Mineral Density and X-ray Density. Mineral density of a bone, or of a certain region of a bone, can be expressed as the weight of a mineral constituent (of specified composition) per unit volume of the bone. The X-ray density of a bone, measured by densitometer, is expressed in the first instance as the densitometer scale reading produced by the passage of the beam of light through a chosen area of bone shadow. Other things being equal, the more radiopaque the bone itself, the lighter is its shadow, and the higher is the density;—that is, "dense" refers to the object X-rayed instead of to the blackness of the film, as is customary among workers concerned with the photographic aspects of X-ray technique.

Factors Affecting X-ray Density. If we think of a bone in a living limb filmed on two occasions, with its mineral density exactly the same on both occasions, we can group under four headings the numerous factors that can, or conceivably could, make the X-ray bone shadow differ in two films, and the same factors could hide a real difference in mineral content. These factors are mechanical, biological and even psychological, in contrast to the factors enumerated by a radiological physicist (absorption, scattering, secondary radiation, and so on); but it is through his factors that the factors in the following list influence shadow density.

1. Factors In or Directly Affecting the Bone. These factors include: thickness of bone; its inclination with reference to the film; the macroscopic, microscopic and submicroscopic arrangement of its mineral components; the density, structure and amount of its organic components, including the bone matrix, periosteum, marrow and blood vessels.

2. Factors In Adjacent Tissues and Other Objects. Overlying and underlying soft tissues affect the bone X-ray shadow density directly; but the underlying soft tissue can also affect it indirectly—for example, an increase in its thickness raises the

bone farther from the film and reduces the bone density reading. Soft tissues alongside the bone, and also other objects such as neighboring bones, aluminum wedges and a person holding an infant's limb in place, can influence bone shadows by the scattering of radiation.

3. X-ray Technique Factors. It is well known that every major controllable step in radiography, from positioning to film drying, can affect film density. But in addition there are factors, known or conceivable, which would be troublesome or impossible to control, such as: differences between films, even from the same box; differences between film holders; differences in a subject's muscle tension or posture, owing to greater familiarity with the X-ray procedure after his first visit; and, probably of greatest significance, some fluctuations of the timing mechanism and of line voltage that are entirely beyond our control, even in equipment of high quality.

4. Densitometer Factors. Some of these factors are common to many types of photometric instruments—for example: fluctuations of line voltage, instrument drift, and intra-observer variation, due to hurry or fatigue, or to an unperceived change in the manner of using the instrument (even, perhaps, a change in the manner of tapping the instrument to liberate the needle before taking a reading). Two factors of special relevance to bone densitometry are:

a) The photometer scale which, in the instrument used in this laboratory, resembles a logarithmic scale in that it becomes progressively coarser from lower to higher light intensities. Therefore, if an unchanged bone produces, owing to processing differences, a lighter shadow on one film than on another, the reading error of the two shadows will differ. Moreover, if the shadows differ so much that they have to be read on different density ranges (of which there are four) there may be a systematic difference between the ranges, in spite of great care in standardizing the instrument before use.

b) Film envelopes. In order to protect the film from contact with the densitometer search head and platform, it is enclosed in a transparent cellulose envelope. To use a new envelope for every few films is too expensive, and it is laborious to test the envelope frequently. Therefore, an arbitrary rule is adopted,

e.g., to start a new envelope after every fiftieth film; but this may still permit the increasing dullness of the envelope surface to affect the readings.

The above list of four groups of factors, specified or implied, is formidable; but in practice we can easily classify the factors more simply, according to whether we can, or cannot, hope to remove or reduce their effects by observations on the films or on the subjects themselves. Such observations will not enable us to differentiate clearly between differences in mineral content and other differences inside the bone; nor will they compensate for the "damping" effects of soft tissue. We might correct for intersubject *differences* in soft tissue thickness and density (and for differences in general film density) so completely that the (corrected) bone shadow densities in a series of films were essentially all the same, and yet there might be considerable differences in mineral content.

With these exceptions, however, something can be done in the way of removal or correction with regard to all the other factors, and it will be seen that one procedure will often correct for a large number of factors, including hidden factors, at the same time.

We may in the future be able to form densitometric estimates of absolute mineral content in living human bones, or of intersubject differences, or of intrasubject changes, by extrapolation from animal experiments or human cadaver bone experiments, but even then the uncertainties of extrapolation will remain. In the meantime, X-ray shadows will continue to be used as indicators of mineral density differences, and for this purpose it is desirable to improve densitometry by attention to the factors that are amenable to treatment. These factors combine with whatever true intersubject differences there may be in the bones themselves, to increase one or the other of the two risks inherent in all investigations:

1. Reduced sensitivity. Great variation in bone shadow density between subjects treated alike, in a nutrition experiment

for example, can cause a great overlap of the readings of the treatment groups and thereby mask a real intergroup difference in the bones themselves.

2. Bias. An excess of spuriously high (or low) readings in one of two treatment groups can produce a statistically "significant" difference when, in fact, no intergroup difference in the bones has been produced.

Reduction of Variation and Control of Bias

Previous bone densitometry in this laboratory has included a search for age and sex differences in the calcaneus,[29] five hand bones[30] and femur (unpublished) of healthy adults, and also studies of arthritic hands, including a search for treatment effects in a controlled drug trial. In all these, and in a study on cadaver bones of seven X-ray technique factors,[26] the primary question has been: How, and to what extent, can variation due to the method be reduced, and how can bias be controlled? An inexpensive densitometer has been used, because if a reliable technique could be devised it ought, if possible, to be within the reach of individual clinical and laboratory research workers.

The present study, while seeking for effects on infants' bones of dietary supplements prescribed to their mothers during pregnancy, has extended the investigation of densitometric methods in several directions which, owing to lack of suitable films, could not previously be explored in this laboratory.

Before presentation of this study, however, it appears desirable to look at the principles that ought to be applied in efforts to reduce variation and to control bias, because, although much statistical arithmetic appears in reports on bone densitometry, those principles, familiar to many research workers in industrial chemistry, applied biology and controlled therapeutic trials, are not conspicuous in densitometric literature or in the Densitometry Workshop transcript.

The three ways of reducing variation and controlling bias are: (a) Uniformity of technique. (b) Correction factors for general film density, soft tissue and bone size. (c) Randomization. For brevity in discussing these, and in the remainder of the report, "bone density" will imply "bone shadow density," and "intersubject variation" will, in general, include (a) intersubject differences in absolute bone shadow density, and also (b) intersubject differences in density changes of the same bone in the same subject.

Uniformity of Technique. Experiments on seven X-ray technique factors[26] have led to the insistence, in this laboratory, on greater uniformity at every stage in the X-raying and processing than is customary, even in the rules prescribed by the Bone Density Laboratory of Pennsylvania State College[3] which, it is reported,[34] were followed in preparing the Nutrition Study films used in the research presented here. In the experiments[26] on cadaver bones and on ivory wedges as bone substitutes, the tube position was fixed and the milliamperage and exposure time were set always at the values employed in hand filming; but the following seven factors were systematically varied: aluminum step-wedge position, kilovoltage, film position in processing, speed of developing, fixation time, washing time and method of drying (a warm film-dryer versus the open room).

Conversion of bone densities to equivalent thicknesses of aluminum removed the major part of the interfilm differences in density of the same bone shadow, but significant differences were still found to be associated with each of the seven factors (P often much less than 0.05). However, it was found that, if the differences purposely introduced were eliminated (e.g., if the kilovoltage were always set at the same level and the films were all processed with the same end up), the variation would be considerably reduced. One step of an ivory wedge was used to represent homogeneous bone of about the same shadow density as several bones in the living hand, and 128 films were processed in succession without changing developer or fixer. From

these films it was estimated that, if the seven factors were held constant (along with milliamperage, exposure time and tube position) and if random pairs of films were taken from the series, each film to be read once, the greater density estimate in the pair would, in about 95 per cent of pairs, exceed the smaller by less than 6.52 per cent of the smaller one. For the mean difference of 25 pairs, this value would, of course, be about 1.3 per cent, and for 100 pairs, 0.7 per cent. Therefore, in a short study, involving the comparison of means rather than individuals, attention to details of technique can make the radiophotographic component of the intersubject variation so small that it will not be likely to obscure treatment effects of the size we wish to detect.

Similar care in operating the Photovolt densitometer greatly reduces its contribution to the variation. The rules adopted in this laboratory exceed the instructions issued with the instrument in two respects: a long warm-up period (30 minutes) and restriction of the reading period on any one film to about 5 minutes, after which the instrument is re-set. It will be noted that the 6.52 percent interfilm variation, mentioned above, comprised not only X-ray technique variation but densitometer variation; and, since two independent rounds of readings were made on the same shadows, it was possible to estimate that, if reading variation were entirely removed (or made negligible by taking the average of a very large number of readings on the same area), the 6.52 per cent would have been reduced to 4.95 per cent—that is, the contribution of reading variation was small. In the reading of actual bone shadows, which are usually less homogeneous than an ivory wedge shadow, the contribution from reading variation is somewhat greater, but not much,[29, 30] and it can be reduced by taking the average of several points on the same shadow—without, however, using the elaborate equipment necessary for making a continuous series of readings across the whole shadow.

Correction Factors in General. Two assertions can be made

regarding all kinds of correction factors, at least in biological data:

1. A correction factor that is to be applied to a particular body of data is more likely to be reliable if it is derived from relationships found in those data than if it has been developed from theoretical considerations or from measurements made outside the data themselves.
2. No correction factor can be assumed perfect; and to prove that a correction factor, derived theoretically or from other data, is near enough to perfection for one's purpose would require extensive information, often far in excess of what is contained in the data under study.

Correction for General Film Density by Conversion to Wedge Thickness. In previous studies in this laboratory densities of bone and soft tissue have been corrected for interfilm differences in general density by conversion of tissue readings into equivalent thickness of a step-wedge, usually of aluminum, but sometimes of ivory. That is to say, the correction factors have been derived from thickness measurements made on the wedges, not solely from the density data under study. That method was impossible in the Nutrition Study films because, although the actual (smooth-sloped) aluminum wedge that had been exposed on the films was obtained, it was found to be warped and also irregularly bent at the narrow end. Therefore, wedge thickness measurements made now, or earlier measurements if obtainable, could not be assumed applicable to the wedge during the filming. This circumstance prompted the use of another correction method which had already been contemplated because of growing doubts about the conversion-to-thickness method. The conversion method appears to depend on at least four assumptions:

1. The assumption that the points chosen for density reading on the wedge shadow correspond closely enough to the points where thickness measurements have been made—perhaps a

reasonable assumption on an aluminum step-wedge, but doubtful on a sloping wedge or an ivory step-wedge, which is difficult to mill with precision.

2. The assumption that there is a negligible difference between the measured vertical thickness at a certain point and the path of the X-rays through the wedge at that point, or at least that this difference does not vary from one part of the wedge to another by more than a negligible amount.

3. The assumption that the errors in wedge thickness readings (both variable errors and systematic errors at different heights of the wedge) are negligible.

4. The assumption that linear interpolation between steps is safe enough in converting bone shadow readings to wedge thickness.

Correction for General Film Density by "Characteristic Curves." This method is illustrated by Facto, et al.[7] The curves are made by graphing the densitometer readings (Y) of the aluminum wedge against the corresponding steps (X) numbered 1, 2, 3, and so on. From a chosen set of observations, a standard curve is created, and then for each film in an investigation its own characteristic wedge curve is drawn and the standard curve is inserted on the same graph. A tissue density reading (Y) is located on the characteristic curve obtained from the same film and its position on the X-axis is noted, i.e., its location with reference to the aluminum wedge. This wedge value, located on the standard curve, leads to a different Y value from the one observed, and this new value is taken as the corrected tissue density. Like any correction method that requires measurements supplementary to those made on the material under investigation, this method introduces additional variation. For the proper spacing of the X values it would seem desirable to use actual step thicknesses instead of numbers, and this would introduce the problems and assumptions of thickness measurements mentioned above. In any case, the method involves the assumption that, if the characteristic curve of a

particular film were identical with the standard curve, the various tissue readings on that film would not differ systematically from the values obtained by the correction method.

A Direct Method of Correcting for General Film Density. The assumptions involved in the methods just described, especially the assumptions pertaining to wedge thickness, would be difficult to validate; but they can be avoided, and so can the labor of measuring the density of a number of steps on each film, by taking on each film a reading (X) at one particular point on the wedge shadow, along with the density reading of the tissue concerned (Y). Then, from the whole series of films there can be found the Y-on-X regression line, to represent (by the regression coefficient) the average upward slope of tissue density corresponding to greater wedge density. The relationship may be either straight or curved, but straight lines are easier to work with arithmetically, and to produce an approximately straight line it may be desirable to use a transformation of the readings, such as their logarithms, as in part of the present study.

In terms of a scatter diagram (wedge readings horizontal, soft tissue or bone readings vertical) with the regression line inserted, the deviation of any particular tissue reading from regression, i.e., its vertical distance above or below the line, represents the effect on the tissue density of factors other than general film density in so far as the regression line represents the relationship between the tissue density and the general film density.

If two or more treatment groups are being compared, the slope of the line is derived from the tissue-wedge density relationships among subjects treated alike, this information being pooled to form an intragroup regression; and the line is drawn through the general means (X and Y) of the subjects. It may then be found that some of the treatment groups tend to deviate more frequently, or more extensively, above the line and others below, and these contrasts can be tested for statistical "significance."

Precautions with the Direct (Regression) Method. Regression and other statistical techniques are becoming so familiar that they are likely to mislead us unless we try to discover exactly what they tell us, and fail to tell us, about the particular material on which we are using them. In the present instance, the key phrase in this search is "in so far as the regression line represents the relationship between the tissue density and the general film density." This leads to the following five remarks:

1. Although regression lines are found by the "least squares" method, which is said to give the line of "best fit," this does not imply that the regression method is a better density correction method than any other. Indeed, although regression methods are used for many purposes in this laboratory, the term "best" is not interpreted in any technical sense, such as "the most likely estimate of the regression coefficient that would be found in an infinite population of such readings." The method is used because it avoids psychological bias and is a common, easily-learned technique, adaptable to the simultaneous study of, and correction for, more than two variables (multiple regression).

2. The tissue-wedge relationship may appear linear (rectilinear) in a graph, and tests may show no evidence that a curve would represent the relationship more accurately, but this does not prove that the relationship is in fact linear. Indeed, we know that even the most familiar straight-line relationship in physics, between the length of a spring and the force applied to it, would be found only approximately true if we progressively increased the fineness of our measurements of any particular spring.

3. Measurement error in X—in this instance, the variation of instrument or observer in wedge density reading—makes the regression line more horizontal than it would be if no error existed.

4. Differences in scatter of Y values (tissue density) at different points on the X axis influence the estimated value of the regression coefficient; and, like unequal variation within samples in a t test, these differences can influence the results of significance tests in comparing treatment groups. Moreover, these risks may be present even if a test has shown no "statis-

tically significant" difference in the scatter of Y values (no heterogeneity of variance).

5. Differences in the tissue-wedge regression in the different treatment groups, even if not detected as stastistically "significant," can either mask a real difference in the comparison of treatment-group means, or create a spurious difference.

In view of all these risks it might be thought that conversion to wedge thickness should be used whenever possible, but the conversion method itself contains analogous risks, plus the previously discussed uncertainties incidental to measurement of the wedge. For these reasons, also, another argument in favor of the conversion method—that it would enable different observers in different places to compare their results directly—is obviously very questionable.

In reality, however, the problem of correcting tissue density for general film density is not nearly so complex as it appears when particular risks are enumerated. Regarding any correction method we have to ask two questions:

1. Does the method increase the precision of the densitometric shadow comparison sufficiently for our purpose? That is, does it reduce the general film density component of the intersubject variation sufficiently to reveal, in samples of reasonable size, intergroup differences in mean shadow densities that we wish to detect if they exist?
2. How can we control the risk of biased verdicts due to the imperfections of the correction method?

The second question will be answered below, under *Randomization;* and for material like that of the Nutrition Study, answers to the first question, dependent on the definition of "reasonable sample size," will be shown later.

Correction for Soft Tissue Density by Subtraction. The commonest method of correcting for soft tissue is to subtract from the bone shadow density reading a quantity derived from a density reading of adjacent soft tissue.[22, 6, 4] In 1950 P. B. Mack,[21] described briefly the method used in her laboratory.

Bone density and adjacent soft tissue density were determined on one film; and on another film, exposed at right angles to the density film, the thicknesses of the bone and soft tissue were measured. The density per unit volume of soft tissue was estimated, and an appropriate multiple of this (determined by the relative thickness of bone and soft tissue) was subtracted from the bone density measurement. The latest development of this method, described by Vose in the appendix to the Densitometry Workshop Transcript,[11] is the same in principle.

In the interval between these two descriptions, the densitometer technique initiated by Mack has been widely used, but there appears to have been no detailed publication to show that the subtraction method has solved the soft tissue problem—either the general damping effect or the effect of intersubject variation in thickness or composition. It may be noted, also, that at the Densitometry Workshop it was revealed that data which appeared to show a relationship between nutrition and calcaneus density had not been corrected for soft tissue. In 1959 Schraer and his collaborators[32] showed, by chemical analysis of rats' femora, that in X-ray densitometry during life the effect of soft tissue was apparently unimportant, but the report added the statement: "When large masses of soft tissue are traversed by the radiation, as in the roentgenography of the spine and femur in humans, the problem becomes formidable."

Even if an auxiliary X-ray tube is used in order to insure that a second film is taken at right angles to the density film without movement of the limb, there remain other difficulties in the subtraction method, including overlap of adjacent bone shadows, magnification problems, and irregularities of bone shape. More fundamentally, the method depends on the assumption that the densities of living bone and living soft tissue are simply additive over the whole range of densities (and thicknesses) that would be met in any particular investigation.

Correction for Soft Tissue by Water Immersion. Since water has approximately the same radiographic properties as soft tis-

sue, immersion of the part to be X-rayed in water of uniform depth for all subjects tends to obliterate the effects of intersubject differences in soft tissue,[15] and it is claimed that no numerical correction for soft tissue is required. The method has not been used in this laboratory, because the aim is to explore simple techniques which could be used widely, and without inconvenience to the subjects, for example in studies of arthritic limbs.

Direct Correction for Soft Tissue Density. In the subtraction method the correction for soft tissue is derived, initially, from a soft tissue density reading on the bone density film. Therefore, it would appear most appropriate to proceed, not by assumptions (e.g., of additivity) and by linear measurements on another film, but by finding out directly the relationship between the soft tissue readings and bone readings on the density films themselves, and to base the correction on that relationship. This has always been the method employed in this laboratory —regression of bone density on soft tissue density. Up to now, the wedge-equivalent thickness (in aluminum or ivory) have been used; but in the Nutrition Study films, since no wedge equivalents were obtainable, the actual densitometer readings on soft tissue (or their logarithms) were used. These were incorporated in the same multiple regression equation as wedge density readings, so that allowance for general film density and soft tissue could be made at the same time.

The regression method introduces, of course, the same kinds of risks as when it is used in correcting for general film density, but the subtraction method entails analogous risks, in addition to its own technical difficulties. As in correction for general film density, the problems can be summed up in the two questions regarding (a) sensitivity in the intergroup comparisons of bone densities, and (b) risks of bias. Again, the second question will be answered in the discussion of randomization, and, in answer to the first question, estimates of precision in the Nutrition Study films will be shown later.

Correction of Bone Density for Bone Size. It is customary to express the chemical composition of an organ in terms of the weight of particular constituents per unit weight (or volume) of the whole organ, or in an equivalent percentage form. In the comparison of groups of subjects such expressions are assumed to compensate for intersubject differences in total weight (or volume) of the organ; but this implies an assumption of uniformity of percentage composition regardless of total size and that is a questionable assumption. Therefore, there is an increasing tendency among biologists to seek, in the material under study, for the actual relationship of constituents to total weight or size, and to make the correction accordingly, by using the regression of weight of the constituent on the total weight (or volume).

In bone densitometry there is the further difficulty of ascertaining total volume. Assumption of spherical or elliposidal shape, or of "average" shapes derived from cadaver bones, are anatomically unrealistic, and the more accurate determination of the actual shape and volume of individuals' bones by body-section radiography is an elaborate procedure suitable only for very limited and special studies.

A bone densitometer reading represents the opacity of a cylinder of bone traversed by the X-rays, and the most direct method of correcting for the amount of bone traversed (the height of the cylinder) is to find the bone thickness in a film at right angles to the density film. Apparently this can be done with considerable precision by the method described by Vose,[11] which includes body-section radiography, but again the procedure is elaborate. In only a few bones, such as the middle phalanx of the little finger, is it possible to obtain a shadow that is free from overlap of other bones in two films at right angles to each other. Therefore, in most bones, filmed by a simple technique, the only available substitute for a thickness measurement is a linear measurement (width or length) obtainable on the density film itself. Correction of bone density for bone size can then be made by regression of density on this

linear dimension, but its effectiveness will depend on the (unknown) correlation between this dimension and the thickness.

Even in a search for density changes in adult bones, where the sizes are commonly unchanged, the sizes should nevertheless be studied because, although the actual size may be unaltered, a change in the size of the shadow may indicate a change in the tilt of the bone, which is a source of bias in density studies.

Randomization. The foregoing discussion of all kinds of correction methods, and of methods of standardizing techniques, has left open the question: How can we avoid bias due to the imperfections of these methods? For example, if a dietary supplement was tested in one group of children (A) against a control group (B), more A's than B's might, in their post-treatment visits, be X-rayed at a time when the conditions of processing resulted in lighter films. If the correction method for general density were less successful with lighter than with darker films, a "significant" intergroup difference in the increase of bone density might be wrongly attributed to the dietary supplement. The same thing might happen if the densitometer or observer differed in reading levels at different times in the period of study; and, from similar causes, either X-raying or densitometry might mask a real treatment effect. Imperfections in soft tissue correction or in bone size correction can likewise mislead.

The way to control the bias that may result from such factors, as well as bias from intersubject differences and intercurrent events during the trial, is: (a) strictly random assignment of the subjects to the treatments and (b) strict adherence to the schedule of pre- and post-treatment X-raying. The densitometer readings can be made in the same (random) order as the order of admission to the trial.

The inference after such an experiment is simple and takes a form such as the following: "The intergroup difference in bone density change is due either to randomization or to randomization plus the dietary supplement; but such large differences

occur in less than 5 per cent (or less than 1 per cent) of randomization experiments (e.g., card shuffling); therefore we shall attribute it to the dietary supplement." This would not necessarily imply that the effect was on the bone itself. The supplement might have promoted the growth (or possibly increased the density) of soft tissue, and, through an undetectable imperfection in the soft tissue correction method, this could have produced the bone shadow difference. We could not rule out this possibility by finding a "nonsignificant" intergroup difference in soft tissue density change. As in most research, the explanation of a result cannot be found by a single experiment or by a repetition of the same kind of experiment. In this instance, other types of evidence, perhaps from animal experiments, would be necessary.

Experiments and Surveys. The reference to randomization recalls an important distinction, which is often overlooked when "significance" tests are applied. In the Nutrition Study, the assignment of the treatments or factors under test (dietary supplements) was under the control of the investigators—that is, the investigation was an experiment in the strict sense, and the treatments could be assigned by a strictly random method, either throughout the whole series or within racial groups, sex groups, or other groups if desired. But the study was intended also to examine differences in bone density associated with race and sex, and since these factors could not be assigned at random, a "significance" test could not lead to a simple inference: Either randomization or the factors under test. A "significant" difference might be due to undetected factors hidden by the race or sex labels—selection biases of various kinds, including "competition between selection rates," which will be discussed later (p. 90). To distinguish such observational studies from experiments in the strict sense, they are best called "surveys."

THE NUTRITION STUDY

The films that provided the data for this report were prepared in a Nutrition Study conducted at Pennsylvania Hospital, the purpose of which was to seek for the effects on infants of dietary supplements (a polyvitamin concentrate and a protein concentrate) prescribed for women during pregnancy. More than 1400 infants, born in the period 1947–1952, were in the study, and of these more than 300, born in 1951 or 1952, were X-rayed. Reports on the total group of infants[16, 17, 18, 19] revealed no detectable effect of the treatment upon the size at birth or on postnatal growth. The films from the X-rayed children were investigated for skeletal maturation by study of the frequencies and sizes of ossification centers in the calcaneus, proximal end of tibia and distal end of femur,[34] and again no treatment effect was found. Bone densitometry, although originally intended, had not been performed, and in 1957 the films were offered to this laboratory by Miss Dorothy G. Wiehl of the Milbank Fund.

The first step was a pilot densitometric study of 26 films available in New York (each film from a different infant), in order (a) to ascertain whether the technique previously used

on adult bones could be applied to these films, and (b) to obtain an estimate of observational and intersubject variation in X-ray bone density estimates. Although the films were not a strictly random sample of the total, there appeared no reason to believe them exceptional in bone density, and the intersubject variation (including observational variation) was found to be sufficiently small to justify the hope that intergroup differences, if large enough to be of interest, could be demonstrated in the complete set of films.

It was decided to study the complete set of films very thoroughly, not only because of the outlay of effort and expense in producing them, but because the current hesitation to expose infants to radiation, except for diagnostic purposes, would probably make it difficult to obtain such a large and potentially informative series again.

The following information has been obtained partly from the publications already indicated and partly from Miss Wiehl.

The Primary Sample. All prenatal women who registered at the Philadelphia Lying-In Hospital were referred to the Nutrition Research Clinic if the estimated duration of gestation was not more than 16 weeks, if the woman was married, if she showed no indication of chronic disease or syphilis, and if she did not refuse to attend clinic in the afternoon. With these exceptions, the women were believed to be an unselected series from the ward service of the hospital stratum of the population.

The films studied were from infants believed to have been born at term, i.e., they weighed 5.5 pounds or more at birth, except for three of 39 weeks' gestation that weighed 5 to 5.5 pounds. If evidence of fetal age from the distal femoral epiphyseal center alone is considered, it may be noted that in the previous survey of these films,[34] that center was reported absent in the first postnatal week in 4 infants out of 314, but it was present in all of the infants used for densitometry; therefore it can be accepted that all, or very nearly all, had passed the seventh fetal (lunar) month.[14]

Treatment Groups. The women were assigned to one or other of the following groups on a random basis controlled for race (Negro or white), age and gravidity:

A. Control group, no nutritional supplement.
B. Vitamin supplemented group.
C. Protein supplemented group.
D. Vitamin and protein supplemented group.

The supplements prescribed were: Polyvitamin concentrate (Upjohn's Zymacaps), 3 capsules per day; protein concentrate (Mead Johnson and Company's Protenum) to furnish 50 gm. protein and 1.5 gm. calcium daily. The same basic marginal diet of protein and calcium was prescribed for all subjects. However, dietary histories, taken twice during the second half of pregnancy, showed that even without the supplements the majority of the women were consuming more protein than was recommended in the basic diet. Very few could be considered materially deficient in calories, protein or calcium.

With regard to protein the study was designed to evaluate benefits that might result from a high protein diet, as compared with an average good diet for this socio-economic group. On each visit to the clinic during pregnancy the women were given a new supply of the protein supplement unless they had sufficient to carry them to the next visit. On the basis of this information the investigators could decide with fair confidence whether a woman was consuming on the average at least half of the prescribed amount of protein supplement. Only the infants of mothers in that category have been included in the treatment comparisons in the bone density analysis. (The excluded cases are indicated in Table 1 by the symbols C(?) and D(?).)

All groups of women received the same general prenatal care and management.

X-ray Films. Films of the infants' right feet and right knees were taken by the technique prescribed by the Growth Study Center at Pennsylvania State College. On each film were two

separate exposures: foot (dorsiplantar view) and lower limb centered at the knee (frontal view). Alongside each of these images was the shadow of the same aluminum-alloy wedge, enclosed in a rectangular metal frame. The wedge, obtained for examination in this laboratory, was 15 cm. long, had a square 1.5 cm. base, and tapered smoothly in width as well as in height from its base.

Table 1 shows the distribution by age at first filming, race, sex and maternal treatment, of the 286 infants from whom one or more films were used in the present densitometric study. It

TABLE 1. FILMS USED FOR DENSITY STUDY—DISTRIBUTION BY AGE, RACE, SEX AND PRENATAL DIETARY TREATMENT.

TOTAL INFANTS FROM WHOM ONE OR MORE FILMS WERE ASSESSED: 308
EXCLUDED BECAUSE NO FILM FROM THE FIRST WEEK WAS AVAILABLE: 22
REMAINDER: 286

Age (Days)	0	1	2	3	4	5	6	7	8	Total
No. of Infants	2	3	32	114	74	55	3	2	1	286

	Treatment of Mother						
Race and Sex	A	B	C	D	C(?)*	D(?)*	Total
White Males	34	26	17	22	1	3	103
White Females	32	20	15	15	3	1	86
Negro Males	13	14	9	7	2	5	50
Negro Females	18	9	8	7	4	1	47
Total	97	69	49	51	10	10	286

* C(?) and D(?) indicate that the mothers had probably taken less than half of the prescribed amount of the protein supplement. These infants were excluded from the treatment comparisons.

had been planned to X-ray each infant three times—immediately after birth, at one month and at six months. The degree of departure from that scheme will be shown in Table 14 and discussed in the text.

Purposes of the Densitometric Study

The purposes of the present study can be summarized thus:

1. A search for effects of the prenatal dietary supplements on X-ray density of bone (and soft tissue).
2. A comparison of bone (and soft tissue) density by race and sex.
3. A search for relationships between density changes and bone size changes, in the hope of obtaining a clue to relationships between density changes and changes in the amount of mineral matter.
4. Further exploration of methods developed in this laboratory for soft-tissue and general-density correction.
5. A review of outstanding problems in bone densitometry in the light of this experience.

Densitometric Technique

Equipment. The transmission densitometer was Photovolt Model 501A, light-box Model 52 with aperture for light-beam 1/16 in. in diameter. To promote stability, before each reading session the instrument was allowed to warm up for more than half an hour, i.e., it was kept at the "Warm up" setting for 5 minutes, and then at the "On" setting for 30 minutes before it was adjusted to the light-beam and tested on each density range.

Preparation and Reading of Films. Each film was inserted in an envelope composed of two 0.01-inch transparent cellulose acetate sheets, and tracing paper was clamped to the envelope by paper clips. Then the film was placed on an X-ray viewing box, outlines of the bones were drawn by pencil on the tracing paper and the selected reading points, enumerated below, were

also marked. The reading points were then pricked through the tracing paper with a needle, without touching the cellulose envelope. When the film, with tracing paper, was placed on the light-box, the beam of light, traversing the aperture, was centered on the needle prick, the tracing paper was turned aside and, with the film held firmly in place, the search-head was lowered on to the surface of the envelope and the density reading was taken.

The author's wife, Ruth M. Mainland, made all the densitometric and bone size readings, and at this stage she was not aware of the differences in treatment indicated by the letters A, B, C and D.

Reading Points. In the bones, the regions selected were those in which the cross-section is roughly circular or elliptical, in order to minimize the variation due to differences in positioning of the limbs in different children and in the same child X-rayed on different occasions. For this reason, the triangular-sectioned tibial shaft was excluded. The bones studied were: shaft of first (great toe) metatarsal, distal epiphysis of femur, and shaft of femur. In the pilot study the calcaneus had been read also; but in the main study it was found difficult in a considerable number of films to distinguish its shadow from that of the talus. (This confusion was probably the reason why, in a previous study of these films,[34] the center for the calcaneus was reported absent from a number of heels.) With each bone, an area of adjacent soft tissue was read, and also a black area of the film, for background density. In detail, the density reading points were as follows:

First Metatarsal: Center of bone, determined by millimeter measurement, longitudinally and transversely. Soft tissue at mid-point between first and second metatarsal, located by eye. Black area on medial side of foot in the same transverse line as the readings on bone and soft tissue.

Distal Epiphysis of Femur: Center of bone, located by eye. Soft tissue 3 mm. from the bone on the lateral side in the same

transverse line as the bone reading. Black area at medial side of knee 3 mm. from soft tissue (but somewhat variable, because the area was sometimes obscured).

Shaft of Femur: Center of bone, found by measurement longitudinally and transversely, except when the proximal end was not visible, and then the reading was taken at the narrowest part of the middle region of the bone. Soft tissue 3 mm. from the bone on the lateral side, in the same transverse line as the bone reading. On many films no area near the middle of the femur was found free from shadows; therefore in all films the black-area reading near the epiphysis was used instead.

Aluminum Wedge Shadow: The outline of the wedge shadow was marked on a piece of transparent cellulose and a small hole was made at the point where the wedge shadow density was to be read on all films—one wedge shadow on the foot film and one on the femur film. In the reading of each film the outline on the cellulose was made to coincide with the wedge, with the tracing paper between, and a fine pencil was inserted into the hole to mark the reading point on the tracing paper.

Order of Reading. Before the reading of any film was started, the densitometer baseline (on Range No. 1) was recorded, and also the reading of a particular area on a standard bone-shadow film. By repeating this standard reading after all the readings on the particular film under examination, it was possible to detect instability of the instrument, but throughout the whole survey re-reading of a film was very seldom found necessary.

The readings were always taken in exactly the same order in each film, first on the foot exposure and then on the knee exposure. The time spent on the reading of any one film (both exposures) was approximately five minutes.

Previous experience had shown that reading variation was a minor component of interfilm variation, and duplicate readings in the pilot study of these films had confirmed this; therefore only one round of readings was conducted.

Size Measurements. With the films on a viewing box and

covered by the transparent cellulose, the following measurements were made by fine calipers applied to a steel scale which, with vernier, read to 0.01 cm.:

First Metatarsal: length; width at mid-point.

Calcaneus: length (antero-posterior axis); greatest measurement at right angles to length.

Distal Epiphysis of Femur: height (vertical axis); width (transverse axis).

Shaft of Femur: width at middle. Femur length was not measured because the proximal end was often obscured.

Randomization. In order to control the bias that reading variation might create in the *intergroup* comparisons, the films from each child were treated as a unit, and these units were read in a sequence determined by random numbers. In order to minimize the effect of reading variation on *intrasubject* comparisons, i.e., Visit (1) versus Visit (2) and Visit (2) versus Visit (3), all the films of the same child were read on the same occasion. To control any residual bias due to the order of reading on any one occasion, the sequence of $V(1)$, $V(2)$ and $V(3)$ films was randomized for each child separately. A more complex, systematic design of reading orders—1, 2, 3; 3, 2, 1, etc., in equal numbers of infants—could fortunately be avoided, because the stability of the instrument was known to be high.

The various points were read in the same order in all films for two reasons:

1. The non-osseous readings were to be used as correction terms for bone readings, and, if there were any drift of the instrument (or observer) during the reading of the same film, randomization of order of reading of bone and wedge for example, would tend to vary the relationship between the correction factor and the reading on which it was to be used.

2. No safe comparison could, in any case, be made between the densities of, say, the femur and metatarsal, because of the bias that might have been introduced by the sequence in which

they were filmed. Therefore, there was no need to avoid a bias (systematic error) due to constant order of reading.

Analysis of Data

The design of an experiment—the particular stratification scheme, followed by random assignment of treatments within strata—dictates the form of analysis. Therefore, the appropriate analysis of bone density differences, in the first postnatal week for example, ought to have taken the form: between races, between sexes, between maternal age groups, between gravidity classes, between treatments, interactions among these various factors or attributes, between infants who were treated alike and were alike in the foregoing attributes. The numbers of infants in the various subclasses would, however, have been so unequal (probably with zero in some subclasses) that this comprehensive analysis, even if it could have been carried through (with the aid of numerous assumptions about the absence of interactions between factors) would have led to rather obscure inferences. In such a situation, the best substitute is to make the treatment comparisons within each subclass separately. Then the verdict may be obvious, or some further analysis, such as combination of probabilities, may be desirable.

Such piecemeal analysis was used in this study, but the stratifications by maternal age and gravidity were ignored, because even the division into four race-sex groups, with four treatment groups in each, left very few infants in some of the subgroups.

In view of the abundance of statistical arithmetic that appears in the subsequent pages, the following remarks seem desirable:

1. The standard procedures (such as linear regression, product-moment correlation and the t test), familiarized by Fisher[8] were used, and the random sampling probability (P) values were determined from the tables of Fisher and Yates.[9]

2. Where the term "significant" is used without mention of a P value, the conventional 5 per cent level of significance is implied.

3. When many significance tests are applied to the same body of data, even if chance alone is operating we must expect a certain proportion of the tests to give verdicts of "significance"—on the average, about 5 per cent of the tests would produce P values less than 0.05. Our evaluation of verdicts must, therefore, depend on confirmation in other parts of the same data, or on *a priori* knowledge, or on an investigation of new material if that is possible.

4. The number of decimal figures that are shown in means, coefficients of correlation and other estimates may suggest an unwarranted claim to precision. They are shown to assist readers who may wish to use the estimates in trying to answer questions that were not answered in the analysis that is presented here.

DENSITIES IN THE FIRST WEEK OF BIRTH

General Mean Densities. Table 2 shows the observed general mean densities, without distinction by race, sex or treatment. Two sets of means are shown, because in the treatment comparisons the doubtful cases, C(?) and D(?) in Table 1, were excluded. When it was discovered, as will be shown, that there was no suggestion of a treatment effect, the doubtful cases were reintroduced. The femur shaft sample is smaller than the metatarsal and femur epiphysis samples, because in about 20 films a shadow, apparently of cloth, lay over the thigh.

Preliminary Comparison of Treatments. In the first stage of the analysis each of the four race-sex-bone subgroups was examined separately. Numerous scatter diagrams were made, in order to discover the general nature of the relationships be-

tween the three densities (wedge, soft tissue and bone) taken in pairs. On the basis of these diagrams, choice was made between actual density units and their logarithms, for use in regression analysis, the preferred unit being one that appeared to give a more nearly straight-line relationship and a more nearly equal scatter of dots at different parts of the line (homogeneity of residual variation). For each of the twelve race-sex-bone groups, regression equations were then produced, to show the relationships of soft tissue density (log units) to wedge density (actual units) and the relationship of bone density (log units) to wedge and soft tissue density—intra-treatment-group regressions. From these equations were estimated the adjusted mean densities shown in Table 3. For example, in the metatarsal of white males, if the mean wedge density and mean soft tissue density had been the same in all four treatment groups as in the white male group as a whole, the mean bone density would have been lowest (806 units) in Group *A*, and highest (840 units) in Group *B*.

TABLE 2. GENERAL MEAN DENSITIES DURING FIRST WEEK AFTER BIRTH.

Treat. Comp. = Infants Used in Comparison of Treatments.
Log = Mean of Logarithms of Densities.
Geom. = Geometric Mean.
Arith. = Arithmetic Mean.
ST = Soft Tissue.
Numbers of Infants are in Parentheses.

	Mean Densities						
Bone	Wedge Arith.	ST Log	ST Geom.	ST Arith.	Bone Log	Bone Geom.	Bone Arith.
Metatarsal							
Total (281)	4104	2.6337	430		2.8927	780	
Treat. Comp. (262)	4154	2.6534	450	700	2.9130	818	1190
Femur Epiphysis							
Total (280)	3558	3.0983	1254		3.3083	2034	
Treat. Comp. (260)	3623	3.1158	1305	1821	3.3216	2097	2562
Femur Shaft							
Total (259)	3588	3.3950	2483		3.6145	4116	
Treat. Comp. (241)	3649	3.4075	2556	3007	3.6210	4179	4350

TABLE 3. GEOMETRIC MEAN DENSITIES IN TREATMENT GROUPS DURING THE FIRST WEEK AFTER BIRTH, ADJUSTED FOR DIFFERENCES IN WEDGE AND SOFT TISSUE.

B = Bone. ST = Soft Tissue. WM = White Male.
WF = White Female. NM = Negro Male. NF = Negro Female.

Adjusted means are derived from linear regressions (ST on W; B on W and ST) within the four treatment groups, for each race-sex group separately. Treatments (A, B, C, D) and numbers of infants are in parentheses.

Bone	Group	Tissue	Observed General Means	Adjusted Means of Treatment Groups Lowest	Highest
Metatarsal	WM (97)	ST	432	397 (B, 26)	505 (D, 17)
		B	822	806 (A, 33)	840 (B, 26)
	WF (82)	ST	439	397 (C, 15)	478 (D, 15)
		B	800	762 (C, 15)	823 (D, 15)
	NM (42)	ST	432	389 (D, 7)	448 (B, 13)
		B	755	743 (C, 9)	784 (D, 7)
	NF (41)	ST	545	496 (B, 9)	599 (D, 7)
		B	920	907 (A, 17)	959 (C, 8)
Femur Epiphysis	WM (96)	ST	1320	1265 (A, 34)	1371 (C, 16)
		B	2081	2013 (A, 34)	2160 (D, 21)
	WF (80)	ST	1204	1181 (D, 14)	1229 (B, 20)
		B	2051	2002 (B, 20)	2081 (A, 32)
	NM (42)	ST	1266	1105 (D, 6)	1326 (B, 14)
		B	2012	1939 (A, 13)	2072 (D, 6)
	NF (42)	ST	1532	1515 (A, 18)	1586 (D, 7)
		B	2322	2074 (C, 8)	2460 (D, 7)
Femur Shaft	WM (89)	ST	2446	2313 (A, 31)	2598 (B, 25)
		B	4122	4081 (A, 31)	4157 (C, 15)
	WF (73)	ST	2515	2411 (C, 15)	2647 (B, 17)
		B	4095	3986 (B, 17)	4191 (C, 15)
	NM (41)	ST	2593	2352 (D, 6)	2938 (B, 13)
		B	4298	4036 (B, 13)	4577 (D, 6)
	NF (38)	ST	2873	2627 (C, 8)	3264 (D,6)
		B	4349	4142 (C, 8)	4518 (B, 8)

The figures throughout the table suggest, both for soft tissue and bone, that the *A* groups tended to contribute the lowest values and the *D* groups the highest values; but comparison of the rank orders of the treatment-group means in the twelve race-sex groups (four in each bone) did not reveal any consistent pattern, and this was confirmed by a more comprehensive analysis, described below.

Because the race-sex groups were analyzed independently, Table 3 does not permit a vertical comparison of adjusted means, such as the lowest values for metatarsal density in *WM, WF, NM,* and *NF;* but the piecemeal analysis permitted comparison of the regression coefficients among the race-sex groups, and this revealed no race or sex distinction. Therefore, a more comprehensive analysis appeared safe.

Correlations of Densities. The data that had been used to produce the individual regressions in the preliminary analysis were now pooled to produce estimates of correlation and regression common to all the sixteen race-sex-treatment groups (Tables 4 and 5). The correlations between bone, wedge and soft tissue were of the same order (+0.75 and upward) as has been found in adult limbs, and such high values confirm the impression, created by the scatter diagrams, that within the observed range of values the relationships were to a very large extent linear. The partial correlation of bone and soft tissue density ($B \times ST.W$) shows that, after differences in general film density had been allowed for, much of the remaining intersubject variation in bone density was associated with soft tissue density. Doubtless this represents largely the effect of thickness (and possibly actual tissue density) of the soft tissue covering the bones, but it may in part represent an effect of general film density that is not fully corrected by eliminating the interfilm differences in wedge density.

The partial correlation of bone and wedge density ($B \times W.ST$) reveals that, although correction for soft tissue removed most of the bone density variation that was associated with

TABLE 4. CORRELATIONS OF DENSITIES DURING THE FIRST WEEK AFTER BIRTH.

The coefficients represent pooled correlations within the 16 race-sex-treatment groups.
B = BONE (LOG UNITS). ST = SOFT TISSUE (LOG UNITS).
W = WEDGE (ACTUAL UNITS). B × ST.W. = CORRELATION OF BONE AND SOFT TISSUE AFTER ELIMINATION OF DIFFERENCES IN WEDGE DENSITY.
NUMBERS OF INFANTS ARE IN PARENTHESES.

CORRELATION	METATARSAL (262)	FEMUR EPIPHYSIS (260)	FEMUR SHAFT (241)
B × W	+0.891	+0.937	+0.743
B × ST	+0.989	+0.974	+0.819
W × ST	+0.881	+0.950	+0.807
B × ST.W	+0.952	+0.764	+0.554
B × W.ST	+0.286	+0.171*	+0.244

* For B × W.ST in femur epiphysis, the random sampling probability (P) is between 0.01 and 0.001; for all the other coefficients, P is much less than 0.001.

TABLE 5. LINEAR REGRESSION RELATIONSHIPS OF DENSITIES DURING THE FIRST WEEK AFTER BIRTH.

The coefficients represent pooled regressions within the 16 race-sex-treatment groups.

B = BONE (LOG UNITS). ST = SOFT TISSUE (LOG UNITS). W = WEDGE (ACTUAL UNITS).

Cases marked "C(?)" and "D(?)" in Table 1 are omitted.

Metatarsal (281 infants):

$$ST = 0.0003680(W) + 1.1232$$
$$B = 0.00003312(W) + 0.8486(ST) + 0.5218$$

Femur Epiphysis (280 infants):

$$ST = 0.0002801(W) + 2.1016$$
$$B = 0.00002833(W) + 0.6592(ST) + 1.1650$$

Femur Shaft (259 infants):

$$ST = 0.0001800(W) + 2.7492$$
$$B = 0.00002657(W) + 0.3159(ST) + 2.4467$$

interfilm differences in general density, some of this general density effect, measured by wedge density, is still demonstrable. A combined correction, by regression of bone density on wedge and soft tissue, was therefore desirable (Table 5).

Estimates of Absolute Bone Density. An attempt was made to answer the question: What would be the average bone densities if no soft tissue were present? It was assumed that, in the absence of soft tissue, the reading alongside the bone would be that of the adjacent black area; mean at metatarsal = 1.8676 log units (74 units); mean at femur shaft and epiphysis = 1.8618 log units (73 units). These values were inserted for *ST* in the equations of Table 5, while the general mean wedge densities were retained for *W*. The bone density estimates were then as follows:

Metatarsal = 2.2425 log units (175 units)
Femur Epiphysis = 2.4931 log units (311 units)
Femur Shaft = 3.1299 log units (1349 units)

Such figures are, of course, little more than speculative, because we do not know quantitatively how the complete absence of soft tissue would affect the adjacent shadow-free area of the film, either in X-raying (absence of scattered rays) or in processing. However, it may be pointed out that the estimate did not require extreme extrapolation (to zero on the soft tissue abscissa), and the lowest observed values in the bone-soft-tissue scatter diagrams gave no suggestion of departure from linearity. Indeed, in the metatarsal series, the mean black-area value (74 units) used in the calculation was actually within the region of observed soft tissue values.

Density Relationship Estimated in Densitometer Scale Units. Although the scatter diagrams suggested that log units for soft tissue and bone density would provide more nearly linear regressions than the densitometer scale units themselves, it was of interest to examine the relationships with untransformed scale units, especially since these units were found pref-

erable to log units in analyzing the density changes after the first week. The following correlation coefficients and regression equations were obtained from white male infants:

Metatarsal (97 infants): $r_{BW} = +0.8444$; $r_{BST} = +0.9808$

$$B = 0.1160(W) + 1.3963(ST) - 211.22$$

Femur Epiphysis (100 infants): $r_{BW} = +0.9322$; $r_{BST} = +0.9701$

$$B = 0.2780(W) + 0.8047(ST) + 98.66$$

Femur Shaft (89 infants): $r_{BW} = +0.8879$; $r_{BST} = +0.9157$

$$B = 0.2759(W) + 0.4604(ST) + 1970.86$$

The high correlations again indicate that the relationships were not far from linear; and the defect of this method, as compared with the use of log units, probably lies chiefly in the inequality of bone density variation at different points along the regression lines.

To form rough estimates of what the bone densities (arithmetic means in scale units) would be if no soft tissue were present, *ST* in the above equations was replaced by the corresponding mean black area densities (94.85 units at the metatarsal and 104.01 units at the femur). Then the estimated mean bone densities were: *Metatarsal* = 400 units; *Femur Epiphysis* = 1158 units; *Femur Shaft* = 3003 units. As would be expected, these are larger than the corresponding geometric means.

Intersubject Variation in Density. Table 4 shows that much of the variation in soft tissue density within the 16 race-sex-treatment groups was associated with differences in general film density, represented by wedge density; and much of the intragroup bone density variation was associated with differences in wedge density and in soft tissue density. For example, the correlation coefficient for $W \times ST$ in the metatarsals, +0.881, shows that $0.881^2 = 77.6$ per cent of the soft tissue variation (expressed as sums of squares of deviations from the group mean) was associated with differences in wedge density. The remaining 22.4 per cent of the variation represents deviations from regression of soft tissue density on wedge density, and this appears as the standard deviation, ±0.2095 log units in Table 6.

Similarly, the residual intersubject variation in metatarsal bone density appears as the standard deviation ±0.05773 log units. All such figures represent pooled estimates from the 16 groups, because there was no notable intergroup difference in intersubject variation.

The percentage equivalents, obtained from antilogs, are shown because they convey a clearer picture of magnitudes of variation than do the logarithmic forms; but in most calculations involving the standard deviations we must return to the logarithmic forms. For example, the ±2*SD* range for femur shaft density is ±0.16564 log units, i.e., –31.7 per cent and +46.4 per cent, which are not the same as the values (–34.8 and +42.0 per cent) that are obtained by doubling the percentage *SD* in Table 6.

The percentages can, however, be misleading, and it is desirable to apply them to the geometric mean densities in Table 2, in order to obtain rough estimates of the standard deviations

TABLE 6. RESIDUAL INTERSUBJECT VARIATION IN DENSITIES DURING THE FIRST WEEK AFTER BIRTH—DEVIATIONS FROM REGRESSIONS SHOWN IN TABLE 5.

The variations are expressed as intersubject standard deviations (log units) within the 16 race-sex-treatment groups. Equivalent percentages are in the form 100(O − E)/E, where O is the observed value and E is the value estimated from the regression equation.

Metatarsal:

Soft Tissue: ±0.2095 (−38.3% and +62.0%)
Bone: ±0.05773 (−12.5% and +14.2%)

Femur Epiphysis:

Soft Tissue: ±0.1222 (−24.5% and +32.5%)
Bone: ±0.07033 (−15.0% and +17.6%)

Femur Shaft:

Soft Tissue: ±0.1741 (−33.0% and +49.3%)
Bone: ±0.08282 (−17.4% and +21.0%)

in absolute units. For example, in percentage form the soft tissue variation is much greater than the bone variation; but in actual units the difference is not so great. Thus, the femur shaft soft tissue variation, 33.0 per cent, applied to the mean, 2556 units, gives 852 units, and the corresponding bone value, 17.4 per cent, applied to its mean, 4179 units, gives 727 units.

The chief use of the intersubject standard deviations of Table 6 was comparison of bone densities in race, sex and treatment groups by the *t* test, and since this test is derived from Gaussian distributions it was desirable to see whether the residual intersubject variation differed greatly from the Gaussian form, particularly in the tail regions. The distributions (log units) were bell-shaped, and the following figures were obtained for percentage frequencies in relation to certain multiples of the intersubject standard deviation:

	Gaussian [Per Cent]	*Metatarsal* (260 infants) [Per Cent]	*Femur Epiphysis* (280 infants) [Per Cent]
Below – 1.96 *SD*	2.5	2.5	2.9
Below – 1.28 *SD*	10.0	7.5	8.2
Above + 1.28 *SD*	10.0	10.0	8.2
Above + 1.96 *SD*	2.5	2.8	1.8

The distribution for femur shaft was similar to these, and therefore, since distributions of means of random samples are even more nearly Gaussian than their parent distributions of individual measurements, there need be no hesitation in accepting the verdict of *t* tests regarding statistical "significance."

Comparison of Treatment Groups. For each bone in each infant, the corresponding wedge density and soft tissue density were inserted into the appropriate equation in Table 5 and the "expected" bone density was estimated. The difference between this value and the observed bone density, i.e., the deviation from regression, provided an index of the density status of that

particular bone with reference to the total group of children after correction for wedge and soft tissue density. Some of the deviations were positive, others negative, and the (algebraic) mean of the deviations in any particular race-sex-treatment subgroup represented the status of that subgroup with reference to the mean density of the whole group of infants. These mean deviations were found for each of the 16 race-sex-treatment groups, and within each race-sex group the mean deviations of the four treatment groups were ranked (1, 2, 3, 4) from the largest negative to the largest positive mean deviation. Table 7 shows the mean deviations converted from logarithmic differences to percentage differences, along with the rank orders.

Because of the unequal numbers of infants in the subgroups, step-by-step analysis was used instead of an all-inclusive analysis of variance. Within each race-sex group the mean deviations of the treatment groups were compared with each other. When the verdict was not clear to the eye, the t test was used, and where it appeared possible that the t-test results from all the race-sex groups in any one bone, although not individually "significant," might in combination show evidence of a treatment effect, the probability (P) values from the individual tests were combined by conversion to chi-square and summation (Fisher,[8] Sect. 21.1). In none of the comparisons of dietary supplement groups, either singly or in combination, was there found a bone density difference that could not easily be attributed to the random assignment of the supplements, and the same was true in the contrasts of supplement groups with control groups. That is, in all the comparisons P was greater than 0.05.

The soft tissue density differences were not analyzed in the same detailed manner, but no suggestion of treatment effects was detected.

The foregoing analyses confirmed the verdict previously obtained from comparison of treatments within each race-sex group separately (Table 3); but the advantages of using re-

gression relationships derived from all infants were:

1. A more accurate estimate of intersubject variation.
2. An opportunity to compare race groups and sex groups after pooling the treatment groups. These contrasts are presented in Table 8.

TABLE 7. INTERGROUP DIFFERENCES IN BONE DENSITIES DURING THE FIRST WEEK AFTER BIRTH—MEAN DEVIATIONS FROM REGRESSION ON WEDGE AND SOFT TISSUE DENSITY (TABLE 5).

Conversion of mean logarithmic deviations produced percentage deviations in the form 100(O − E)/E, where O is the observed value and E is the value estimated from the regression equation.

Rank orders within each race-sex group run from the largest negative to the largest positive value.

WM = White Male. WF = White Female. NM = Negro Male. NF = Negro Female
A, B, C, D = Treatment Groups. N = Number of Infants.

	Metatarsal			Femur Epiphysis			Femur Shaft		
Group	N	Mean Dev. (%)	Rank	N	Mean Dev. (%)	Rank	N	Mean Dev. (%)	Rank
WM A	33	+2.1	1	34	−4.7	1	31	−1.0	2
B	26	+6.7	4	25	−0.1	2	25	+1.7	4
C	17	+5.0	3	16	+0.1	3	15	+0.9	3
D	21	+4.0	2	21	+3.0	4	19	−3.3	1
Total	97			96			90		
WF A	32	+1.0	3	32	+5.0	4	29	−1.1	2
B	20	+0.9	2	20	+0.9	1	17	−3.4	1
C	15	−4.9	1	14	+4.5	3	15	+0.6	4
D	15	+3.3	4	14	+2.7	2	12	−1.1	3
Total	82			80			73		
NM A	13	−5.7	2	13	−4.8	1	13	+6.7	3
B	13	−4.0	3	14	+0.4	4	13	−3.4	1
C	9	−5.9	1	9	−1.0	2	9	+2.5	2
D	7	−1.1	4	6	−0.3	3	6	+9.2	4
Total	42			42			41		
NF A	17	−8.7	1	18	−1.9	2	16	−1.0	3
B	9	−5.7	3	9	+2.9	4	8	+1.5	4
C	8	−1.7	4	8	−2.6	1	8	−5.5	1
D	7	−6.4	2	7	+2.8	3	6	−2.2	2
Total	41			42			38		
Grand Total	262			260			241		

TABLE 8. RACE AND SEX DIFFERENCES IN MEAN DENSITIES DURING THE FIRST WEEK AFTER BIRTH, ADJUSTED BY EQUATIONS IN TABLE 5.

Percentage Difference = 100 (larger mean minus smaller mean)/(smaller mean).

W = WHITE. N = NEGRO. M = MALE. F = FEMALE.
P = RANDOM SAMPLING PROBABILITY. NS = NOT SIGNIFICANT AT THE 10 PER CENT LEVEL.

NUMBERS OF INFANTS CORRESPOND TO THOSE IN TABLE 7.

BONE AND COMPARISON		SOFT TISSUE DENSITY			BONE DENSITY		
		Sign	Per Cent	P	Sign	Per Cent	P
Metatarsal							
W – N	M	+	2.1	ns	+	9.6	<0.001
	F	–	2.7	ns	+	6.3	0.01
M – F	W	–	3.2	ns	+	3.7	0.1
	N	–	8.2	ns	+	0.5	ns
Femur Epiphysis							
W – N	M	+	3.6	ns	+	2.2	ns
	F	–	2.5	ns	+	4.9	0.1
M – F	W	+	11.6	0.01	–	4.0	ns
	N	+	5.2	ns	–	1.4	ns
Femur Shaft							
W – N	M	–	5.5	ns	–	2.2	ns
	F	+	1.8	ns	+	0.1	ns
M – F	W	–	0.6	ns	+	0.6	ns
	N	+	6.7	ns	+	3.0	ns

Comparison of Race and Sex Groups. The only soft tissue density difference in Table 8 that exceeds the 5 per cent level of significance appears under the femur epiphysis—the male-female difference in white infants ($P = 0.01$ approximately) and even this may well be fortuitous, because out of six *M-F* contrasts of soft tissue density three are positive and three are negative.

Among the bone density differences, only two have P values less than 0.05, but both are white-Negro differences, both have the same sign, both are in the metatarsal and in both the P values are rather low. Even without estimating how often

such a concatenation would occur by chance alone, we can assume that it is much more rare than the single occurrence in the soft tissue densities. Moreover, Table 7 shows that in the metatarsal 7 of the 8 treatment groups in white infants had positive signs, while all 8 in the Negroes had negative signs. Therefore it will be appropriate to ask, later, whether this apparently greater bone density in white infants than in Negro infants of the same sex could be explained by a difference in bone size.

Confidence Limits of Treatment Differences. Although there was no evidence that the prenatal treatments had any effect on bone density, the question can still be asked: If, by increasing the numbers of infants in the investigation, a difference (say, between treated and control groups) were discovered, how large might it be? On the assumption of approximately Gaussian distributions, the 95 per cent confidence limits of the difference were estimated from the formula: Observed difference between sample means ± twice the standard deviation of the difference. The mean bone densities, adjusted for wedge and soft tissue densities, were used, and the results were expressed as percentages in the form:

100 × (mean treated minus mean control)/mean control.

For the metatarsal in white infants, male and female (65 controls and 114 treated) the confidence limits were: –2.8 per cent and +5.6 per cent. That is, if the control and treated samples, studied here, were random samples of their respective populations, and if very large numbers were examined, we would be unlikely to find the mean differences outside the specified range. For the femur shaft, the corresponding limits were: –4.8 per cent and +5.1 per cent.

Comparison of Tissue Density and Body Measurements. The survey[18] of approximately 1,400 infants, of whom the present (X-rayed) series was a subgroup, included measurements at birth of body weight, crown-sole length, crown-rump

length, rump-sole length, head circumference, chest circumference, hip breadth and calf circumference. With minor exceptions it was found that in group-mean values (a) males exceeded females of the same race, and (b) white infants exceeded Negro infants of the same sex. Although most of the differences in Table 8 are "nonsignificant" and some are negative, they are not incompatible with these race and sex differences in body measurements.

For the study of intragroup relationships, in the first postnatal week, between body measurements and bone density (adjusted for wedge and soft tissue densities) the following correlation coefficients were obtained from the femur shafts of white male infants (numbers of infants are in parentheses):

Bone Density × *Weight* (91): +0.0831 ($P = 0.4$–0.5)
Bone Density × *Crown-Sole Length* (77): –0.1493 ($P = 0.2$ approx.)
Bone Width × *Weight* (80): +0.4304 (P less than 0.001)
Bone Width × *Crown-Sole Length* (69): +0.4263 (P less than 0.001)

The last two coefficients show the degree of relationship that one expects to find between total body measurements and linear components of the body. By contrast, the first two coefficients, showing no evidence of a relationship between bone density and body size, lead one to suspect either (a) that there was no close relationship between mineral content and bone size, or (b) that the relationship was masked by soft tissue or by uncorrected variation in film density.

Bone Sizes. Before bone sizes were studied in relationship to bone density, they were examined directly. No suggestion of a relationship to the dietary treatments was found; therefore, the treatment groups were pooled for race and sex comparisons (Table 9). The tibia was included as an indication of leg length, and the calcaneus was measured because it was

TABLE 9. BONE SIZES (CM.) DURING THE FIRST WEEK AFTER BIRTH, AS MEASURED ON X-RAY FILMS.

The sizes were not corrected for magnification. Intersubject standard deviations are pooled estimates from within the four race-sex groups.

Percentage difference = 100(larger mean minus smaller mean)/(smaller mean).

W = WHITE. N = NEGRO. M = MALE. F = FEMALE.
P = RANDOM SAMPLING PROBABILITY. NS = NOT SIGNIFICANT AT THE 10 PER CENT LEVEL.
CV = COEFFICIENT OF VARIATION. CALC. = CALCANEUS.

BONE AND NO. OF INFANTS	GENERAL MEAN	STAND. DEV'N.	CV PER CENT	DIFFERENCE BETWEEN GROUP MEANS: GROUP		SIGN	PER CENT	P
Metatarsal Width	0.4907	0.0411	8.4	W – N	M	–	2.3	<0.1
(WM, 101; WF, 85					F	–	1.2	ns
NM, 49; NF, 46				M – F	W	+	5.1	<0.001
Total, 281)					N	+	6.2	<0.001
Metatarsal Length	1.4448	0.0989	6.8	W – N	M	–	3.4	<0.01
(Nos. as for					F	–	0.7	ns
width)				M – F	W	+	0.7	ns
					N	+	3.3	<0.02
Femur Epiphysis, Transv.	0.7339	0.1854	25.3	W – N	M	+	3.6	ns
(WM, 99; WF, 85					F	+	8.1	ns
NM, 49; NF, 47				M – F	W	–	4.8	ns
Total, 280)					N	–	0.5	ns
Femur Epiphysis Vert.	0.5389	0.1333	24.7	W – N	M	+	4.0	ns
(Nos. as for					F	+	5.6	ns
transv.)				M – F	W	–	6.3	<0.1
					N	–	4.8	ns
Femur Shaft Width	0.6940	0.0591	8.5	W – N	M	+	1.7	ns
(WM, 98; WF, 80					F	+	8.0	v. small
NM, 48; NF, 46				M – F	W	+	1.4	ns.
Total, 272)					N	+	7.7	v. small
Calcaneus Ant.-post.	1.3189	0.1636	12.4	W – N	M	–	5.9	<0.02
(WM, 83; WF, 59					F	–	1.6	ns
NM, 39; NF, 39				M – F	W	–	1.4	ns
Total, 220)					N	+	3.6	ns
Calcaneous Transv.	0.8860	0.0919	10.4	W – N	M	+	4.8	ns
(WM, 83; WF, 59					F	+	5.7	ns
NM, 40; NF, 38				M – F	W	+	0.5	ns
Total, 220)					N	+	1.4	ns
Tibia Length	6.1579	0.4952	8.0	W – N	M	+	2.2	ns
(WM, 85; WF, 77					F	+	4.4	<0.01
NM, 42; NF, 43				M – F	W	+	0.4	ns
Total, 247)					N	–	1.8	ns

at first hoped that density readings could be obtained on it, but even its linear measurements could often not be determined with confidence. Therefore the discrepancies between the results obtained from the two axes of those calcanei that could be measured (Table 9) probably reflect merely the lack of precision of the measurements.

Among the other bones, the femur and tibia agreed with the body measurements, mentioned above, in showing a greater mean size in whites than in Negroes, but in the metatarsal the reverse was true, and this is curious because, as already noted, the metatarsal density was greater in whites than in Negroes. Sex differences also, were not clear cut, except in the metatarsal width, which was greater in males than in females of both races.

In a previous study of these films[34] measurements of the calcaneus, proximal tibial epiphysis and distal femur epiphysis were made, and it was possible to compare the transverse femur epiphysis measurements in Table 9 with those of the previous survey:

		Previous Survey (314 infants)	*Present Survey* (280 infants)
General mean		6.2 mm.	7.3 mm.
W–N	*M*	+3.4 per cent	+3.6 per cent
	F	+6.5 "	+8.1 "
M–F	*W*	−10.0 "	−4.8 "
	N	−6.9 "	−0.5 "

In the present survey the samples were smaller than in the previous survey because some of the films previously measured were no longer available. The difference between the general means (1.1 mm.) appears too large to be attributed entirely to interobserver differences, and presumably it was due partly to the disproportion in subgroup numbers—for example, the previous survey contained 18 more white males than the present survey, but only 2 more Negro females. In spite of the discrepancies, the direction of the race and sex differences was the same in both surveys.

The standard deviations and coefficients of variation in Table 9, although representing chiefly intersubject variation, were doubtless somewhat inflated by race and sex differences and by the fetal age differences which are present even in a group of "mature" infants. However, the figures suffice to show the striking contrast between the variation in the femur epiphysis, a center which had recently appeared, and the long-established primary centers in metatarsal and femoral shafts.

Correction of Bone Sizes for Magnification. The measurements in Table 9 do not allow for differences in magnification of the bones due to differences in limb thickness. There are considerable technical difficulties in correcting for magnification by taking a film at right angles to the density film without moving the limb from its original position; but an indirect measure was available in this study—the differences in soft tissue density readings, which are undoubtedly due largely to differences in thickness of the tissue. Comparison of Tables 8 and 9 shows that some of the bone size differences might have been partly caused by differences in thickness; for example, the male-female difference in the Negro infants' femur shaft was positive both for bone size and soft tissue density. On the other hand, in some instances a negative size difference was accompanied by a positive soft tissue difference and vice versa. In particular, the sex difference in metatarsal width could not be accounted for by the soft tissue difference, which was in the opposite direction.

Special attention was paid to the size relationships of the femur shaft because its cross-section would be more nearly circular than the cross-sections of metatarsal or femur epiphysis, and therefore its width, if adjusted for magnification, would approximate its antero-posterior axis, through which the X-rays passed in producing the shadow, although the exact relationship of the two axes would depend largely on the angle between the femur and the film. Table 10 shows how an adjustment for differences in magnification can be made by using

soft tissue density. It was, of course, necessary to use wedge density as a correction for soft tissue density, and when that was done, two relationships were demonstrated:

1. A positive correlation between femur width and soft tissue density, after elimination of wedge density differences.

2. A negative correlation between femur width and wedge density after elimination of soft tissue density differences.

The first correlation is what would be expected, as a manifestation of the influence of soft tissue lying between the femur and the film; but the second (negative) correlation is puzzling. In so far as differences in general film density are produced by X-ray exposure factors (kilovoltage, milliamperage and timing) one would expect that the lower wedge readings, indicating darker films, would be associated, if anything, with reduced size of bone shadows—a "burning out" effect on the periphery of the bones. Whatever the explanation of the relationship between bone size and wedge density may be, it was necessary to make allowance for it in correcting bone size for soft tissue, and this is done by the equation at the foot of Table 10.

TABLE 10. RELATIONSHIP OF FEMUR SHAFT WIDTH AND SOFT TISSUE DENSITY DURING THE FIRST WEEK AFTER BIRTH (89 MALE WHITE INFANTS).

S = Bone Size (width in cm.).
W = Wedge Density (units).
ST = Soft Tissue Density (log units).
P = Random Sampling Probability.

Relationship	Correlation Coefficient	P
S × ST	+0.1325	greater than 0.1
W × ST	+0.8601	very small
S × W	−0.0653	very large
S × ST. W	+0.3707	less than 0.001
S × W. ST	−0.3546	less than 0.001

$$S = -0.00003182(W) + 0.1472(ST) + 0.3223$$

Relationship of Bone Density to Bone Size. If thicknesses of the bones were known, their densities, after correction for wedge and soft tissue density, could be expressed as so many units of density per unit thickness, or as density units per unit volume of the bone cylinder that corresponds to the aperture for the light-beam. It would, however, be preferable to ascertain the actual relationship between bone thickness and density; and where, as in this study, bone thickness is not known, we must use another available dimension in seeking for the size-density relationship. In the metatarsal and femur shaft the width was appropriate; in the femur epiphysis the vertical axis was used because it was somewhat less variable than the transverse axis (Table 9).

In all three bones (Table 11) the size-density correlation was positive, but only in the femur epiphysis was it even moderately high. In general, however, a positive correlation would be expected; therefore, even in the metartarsal, where the correlation was very low, it appeared desirable to examine the effect of the apparent relationship; and Table 12 shows the effect of using the regression coefficients of Table 11. Two points are noteworthy:

TABLE 11. RELATIONSHIPS OF BONE DENSITY AND BONE SIZE DURING THE FIRST WEEK AFTER BIRTH.

B = BONE DENSITY (LOG UNITS) ADJUSTED FOR WEDGE AND SOFT TISSUE DENSITY.
S = BONE SIZE (CM.), I.E., WIDTH IN METATARSAL AND FEMUR SHAFT, VERTICAL AXIS IN FEMUR EPIPHYSIS. EXCEPT IN THE SECOND SET OF FEMUR SHAFT COEFFICIENTS, THE OBSERVED BONE SIZES WERE USED.
P = RANDOM SAMPLING PROBABILITY.

BONE	GROUP	NO. OF INFANTS	CORRELATION COEFFICIENT (B × S)	P	REGRESSION COEFFICIENT (B/S)
Metatarsal	White males	97	+0.049	v. large	+0.06627
Femur Epiphysis	All groups	280	+0.626	v. small	+0.3200
Femur Shaft	White males	89	+0.314	<0.01	+0.2316
*Femur Shaft**	White males	89	+0.194	0.1–0.05	+0.1549

* Femur shaft width adjusted for differences in soft tissue density.

TABLE 12. ESTIMATED DIFFERENCES IN MEAN BONE DENSITY RELATED TO BONE SIZE DIFFERENCES IN TWO SAMPLES OF *n* INFANTS DURING THE FIRST WEEK AFTER BIRTH.

Sizes are widths in metatarsal and femur shafts, vertical axis in femur epiphysis. Femur shaft width differences were adjusted for differences in soft tissue density. In estimation of percentage density difference, the denominator was the sample with smaller mean bone size.

Bone	General Mean Size (cm.)	Diff. in Size (Per Cent of General Mean Size)	Difference in Density (Per Cent) Mean	*n*	Range (±2SD of Diff.) Lower	Upper
Metatarsal	0.4907	0	0.0	25	− 7.2	+ 7.8
				50	− 5.2	+ 5.5
				100	− 3.7	+ 3.8
		+10	+0.8	25	− 6.5	+ 8.6
				50	− 4.5	+ 6.3
				100	− 3.0	+ 4.6
		+20	+1.6	25	− 5.8	+ 9.5
				50	− 3.7	+ 7.1
				100	− 2.2	+ 5.4
Femur Epiphysis	0.5389	0	0.0	25	− 6.9	+ 7.4
				50	− 4.9	+ 5.2
				100	− 3.5	+ 3.7
		+10	+4.1	25	− 3.1	+11.8
				50	− 1.1	+ 9.4
				100	+ 0.4	+ 7.8
		+20	+8.2	25	+ 0.8	+16.3
				50	+ 2.9	+13.9
				100	+ 4.5	+12.2
Femur Shaft	0.6940	0	0.0	25	−10.2	+11.4
				50	− 7.3	+ 7.9
				100	− 5.3	+ 5.5
		+10	+2.6	25	− 7.9	+14.2
				50	− 5.0	+10.7
				100	− 2.6	+ 8.2
		+20	+5.2	25	− 5.6	+17.2
				50	− 2.6	+13.5
				100	− 0.3	+11.0

1. Percentage differences in density are much smaller than the corresponding percentage differences in size.

2. If the distribution of differences in sample-pairs can be accepted as approximately Gaussian, with a $2SD$ range including about 95 per cent of the differences, even with $n = 100$ and 20 per cent differences in bone size, a number of negative values would be met within the $2SD$ range in metatarsal and femur shaft, indicating that the sample with the larger mean size had the smaller mean density.

Required Sample Sizes. After bone density differences have been adjusted for differences in (a) wedge density, (b) soft tissue density, and (c) bone size, intersubject variation still remains in each of the groups whose mean bone densities are to be compared, and the question arises: How large must the difference in adjusted group mean bone densities be before it can be safely attributed to something more than a random assignment of subjects to the respective groups? Table 13 gives answers to that question, based on three assumptions:

1. It is assumed that the frequency distributions of intersubject differences in bone density are sufficiently like the Gaussian distribution to justify comparison of means by the t test. Evidence in support of this assumption has already been given, and it is in agreement with observations in other bone density studies conducted in this laboratory.

2. It is assumed that the intersubject variation found in comparing other samples would be the same as was found in this series. For safety in planning another study on the same kind of material, it would be desirable to assume that the intersubject standard deviations in Table 13 were underestimates found in random samples of the material to be studied, and then to estimate how large the new standard deviation might be. A simple method, using Fisher and Yates'[9] table of variance ratios, i.e., Snedecor's F,[33] can be employed.

3. It is assumed that the 5 per cent level of significance would be acceptable; but it would be easy to make sample size esti-

mates for other levels of significance.

For sample sizes from 50 upward, the differences required for significance at the 5 per cent level are not excessive in work with biological material; but the last column of Table 13 illustrates a point that is often overlooked. For instance, with two samples of 50 infants, it is estimated that a difference in mean metatarsal density of 5.5 per cent or greater would be significant at the 5 per cent level; but let us suppose that 5.5 per cent is the "true" (population) difference in mean metatarsal density between two race, sex or treatment groups, and

TABLE 13. ESTIMATES OF INTERGROUP DIFFERENCES IN MEAN BONE DENSITY REQUIRED FOR SIGNIFICANCE (P LESS THAN 0.05) DURING THE FIRST WEEK AFTER BIRTH.

Bone densities were adjusted by regression on wedge density, soft tissue density and bone size.

Percentage difference between means = 100(larger minus smaller)/(smaller).

The estimates represent twice the standard deviation of the difference, derived from the residual SD's (reproduced from Table 6). The population estimates represent values necessary to ensure 95 per cent probability of meeting the required sample differences.

Bone	Intersubject SD (log units)	No. in Each Sample	Required Sample Diff. (Per Cent)	Required Population Diff. (Per Cent)
Metatarsal	0.05773	25	7.8	14.6
		50	5.5	10.2
		100	3.8	7.1
*Femur Epiphysis**	0.07033	25	9.6	18.2
	(0.05487)		(7.4)	(13.9)
		50	6.7	12.5
			(5.2)	(9.7)
		100	4.7	8.7
			(3.7)	(6.7)
Femur Shaft	0.08282	25	11.4	21.7
		50	7.9	14.9
		100	5.5	10.3

* The parenthesized figures for femur epiphysis represent residual variation after regression of bone density on size as well as on wedge and soft tissue density. In the other two bones the adjustment for size-density relationship had a negligible effect on residual variation.

that we could take a series of random samples of 50 infants, one sample from each group. In about half the pairs the difference would be less than 5.5 per cent. Therefore, we must ask: How great must the population difference be in order to insure that in the vast majority of pairs the difference would be at least 5.5 per cent? The answer in the last column of Table 13, i.e., 10.2 per cent, was obtained by translating "vast majority" into 95 per cent, and utilizing the fact that in a Gaussian distribution 95 per cent of the items lie above the value, mean minus 1.6449*SD*. (In this instance the items would be differences between samples of 50 infants.)

Relationship between Metatarsal and Femur Densities. Although the shadows of metatarsal and femur shaft were obtained by two separate exposures, they were on the same film. Therefore, it could not be assumed that a significant correlation between the densities of the two bones represented anything more than shared experiences—the similar behavior of the generator during successive exposures, and also the film processing. On the other hand, the absence of a correlation might be informative; therefore, the following coefficients were calculated for metatarsal and femur shaft densities in log units from $V(1)$ films of 90 male white infants:

Before correction for wedge and soft tissue: +0.7629; after correction: –0.0421.

The corrections removed all evidence of correlation, including the actual positive relationship that one would expect to exist, just as there existed a correlation of + 0.4847 between the widths of metatarsal and femur shaft in the same group of white males.

CHANGES AFTER THE FIRST POSTNATAL WEEK

Drop-outs and Irregular Attendances. Although it had been planned to re-examine the infants at one month and at six months after birth, the ages at Visit (2) and at Visit (3) varied greatly, and so did the intervals between the visits (Table 14). In the early part of the study, although the infants were brought to the clinic, films were not taken because the X-ray schedule was not fully operating. Moreover, many infants were not brought back at all. Of the 286 infants who provided films from the first postnatal week (Table 1), 61 provided only that film, while 146 provided also two subsequent films.

TABLE 14. FILMS USED FOR STUDY OF CHANGES IN DENSITY AFTER THE FIRST POSTNATAL WEEK—AGES, INTERVALS BETWEEN VISITS, RACE AND SEX DISTRIBUTIONS.

AGES AND INTERVALS ARE IN DAYS

FREQUENCY DISTRIBUTION	VISIT (1) — VISIT (2)			VISIT (2) — VISIT (3)		
	Age at V(1)	Age at V(2)	Interval	Age at V(2)	Age at V(3)	Interval
Minimum	0	21	18	21	155	77
10th Percentile	2	28	25	29	173	124
Median	3	32	28	33	184	149
Mean	3.5	32.7	29.3	38.8	184.6	145.9
90th Percentile	5	40	36	65	199	161
Maximum	8	50	42	104	218	182

Exclusions: At V(1), infants more than 8 days old.
V(1) — V(2) intervals greater than 42 days (6 weeks).
V(2) — V(3) intervals greater than 182 days (26 weeks).

RACE & SEX	V(1) — V(2)	V(2) — V(3)
White males	52	45
White females	39	35
Negro males	29	35
Negro females	31	28
Total	151	143

Race and sex comparisons were as follows:

	One Film Only	*Three Films*	*Total*
White Males	23 (32 per cent)	48 (68 per cent)	71
White Females	28 (45 ")	34 (55 ")	62
Negro Males	6 (15 ")	35 (85 ")	41
Negro Females	4 (12 ")	29 (88 ")	33
Total	61	146	207

The sex differences were not significant, but the Negro groups contained a much greater proportion of 3-film infants than did the corresponding white groups. This was familiar to the investigators in the Nutrition Study and occurred because many white mothers started taking their infants to private physicians shortly after birth.

In a search for bias due to drop-outs and irregular attendances, two comparisons were made on the Visit(1) films of the 1-film and 3-film groups in each of the four race-sex groups:

1. Mean metatarsal density after adjustment for wedge and soft tissue density. In two of the groups the 1-film versus 3-film difference was positive, in the other two groups it was negative.
2. Mean length of tibia. Three of the differences were positive, and one was negative, but the differences were far from significant at the 5 per cent level.

Lack of obvious bias is, however, by no means proof that bias did not exist; and even if there were no bias-causing differences at Visit(1) this would not tell anything about the factors associated with nutrition or disease that affected attendance at subsequent visits. Therefore, the main purpose of this part of the study was not to seek for relationships between intervisit density change and prenatal treatment, or the relationships between density and absolute age, but to determine:

1. The effect of wedge and soft tissue corrections on changes

of density in the same bone.

2. The relationship of bone size change to density change.

Consequently, variation in ages at the same visit and variation in the intervisit intervals were not serious impediments. No attempt was made to estimate, for each infant, a daily rate of density change or the amount of change that would be expected if, say, all infants had been filmed immediately after birth and on the 28th day. Such estimates would assume, for each infant, a linear change in density with time, whereas it seemed preferable to ascertain empirically the general relationship between density change and the length of the interval.

Intervisit Changes in Density. In the text and tables the word "change" has been used to indicate an intervisit difference (later visit minus earlier visit) between films from the same infant, in order to avoid repetition of the more correct phrase "apparent change," and to avoid confusion between this difference and other differences, such as the intersubject and intergroup differences in the amount of apparent change. Scatter diagrams of the density changes in bone, soft tissue and wedge, plotted against each other, showed no evidence that a logarithmic or other transformation would improve the linear fit of the variables; therefore densitometric cunits were used throughout the analysis of intervisit changes.

The negative signs attached to the wedge densities in Table 15 show that, on the average, the $V(2)$ films were darker than the corresponding $V(1)$ films, and that the $V(3)$ films were darker than the $V(2)$ films. Therefore the question arose: When this difference is allowed for, what will be the effect on the changes in soft tissue and bone density?

Correlations of Density Changes. Table 16 shows the close relationships of the changes in wedge, soft tissue and bone. As in the first week (Table 4), the high values of the zero order correlations (the first three lines of Table 16) confirmed the impression from the scatter diagrams that the relationships

TABLE 15. MEAN DENSITY CHANGES AFTER THE FIRST POSTNATAL WEEK—OBSERVED (UNADJUSTED) VALUES IN DENSITOMETER UNITS.

Mean intervals: Visit (1) — Visit (2) = 29.3 days.
Visit (2) — Visit (3) = 145.9 days.

N = Number of Infants. W = Wedge. ST = Soft Tissue. B = Bone.

	Visit (2) minus Visit (1)				Visit (3) minus Visit (2)			
Bone	N	W	ST	B	N	W	ST	B
Metatarsal	148	−596	−163	−321	139	−285	+330	+280
Femur Epiphysis	149	−595	−326	−438	137	−778	+163	+180
Femur Shaft	131	−612	−169	−276	95	−1022	−0.2	−304

TABLE 16. CORRELATIONS OF DENSITY CHANGES (IN DENSITOMETER UNITS) AFTER THE FIRST POSTNATAL WEEK.

Coefficients for V(1) — V(2) were pooled estimates from within the 16 race-sex-treatment groups; for V(2) — V(3) they were pooled from the 4 race-sex groups, disregarding treatments.

B = Bone W = Wedge. ST = Soft Tissue.
B × W.ST = Correlation of Bone Density Change with Wedge Density Change after Elimination of Soft Tissue Density Change.
Numbers of Infants are in Parentheses.

The random sampling probability (P) was very small for all coefficients except B × W.ST in femur shaft V(2) — V(3), where P was approximately 0.6.

	Visit (2) minus Visit (1)			Visit (3) minus Visit (2)		
Correlation	Met. (148)	Fem. Ep. (149)	Fem. Shaft (131)	Met. (139)	Fem. Ep. (137)	Fem. Shaft (95)
B × W	+0.780	+0.949	+0.888	+0.840	+0.917	+0.849
B × ST	+0.918	+0.976	+0.911	+0.978	+0.972	+0.954
W × ST	+0.756	+0.902	+0.863	+0.789	+0.898	+0.883
B × W.ST	+0.331	+0.731	+0.413	+0.542	+0.426	+0.048
B × ST.W	+0.802	+0.882	+0.560	+0.946	+0.844	+0.823

were essentially linear. The table shows also that elimination of soft tissue density changes ($B \times W.ST$) did not suffice, except in the femur shaft in $V(2)$-$V(3)$, to eliminate the effects of general film density, represented by the wedge. Consequently, W and ST are in the bone regression equations (Table 17). In that table it will be noted that for any particular bone (or soft tissue) the regression coefficients for $V(1)$-$V(2)$ are rather similar to the corresponding coefficients for $V(2)$-$V(3)$, except in the femur shaft in the second interval, where, as was revealed by the correlation coefficient ($B \times W.ST$), the wedge contributed little to the estimation of bone density change.

Intergroup Differences in Density Change. Before the data from the race, sex and treatment groups were pooled to produce Table 18, intergroup differences were examined, especially in

TABLE 17. LINEAR REGRESSIONS OF DENSITY CHANGES (IN DENSITOMETER UNITS) AFTER THE FIRST POSTNATAL WEEK.

The coefficients were estimated from within the same groups as the correlation coefficients in Table 16, and the symbols (B, W, ST) have the same meaning as in that table.

SD = INTERSUBJECT STANDARD DEVIATION, REPRESENTING DEVIATIONS FROM REGRESSION. NUMBERS OF INFANTS ARE IN PARENTHESES.

INTERVAL & BONE	REGRESSION EQUATION	SD
V(2) MINUS V(1):		
Metatarsal (148)	ST = 0.3493(W) + 44.82	487
	B = 0.1437(W) + 1.1915(ST) − 40.74	210
Femur Epiphysis (149)	ST = 0.8548(W) + 182.47	720
	B = 0.3870(W) + 0.7162(ST) + 25.39	279
Femur Shaft (131)	ST = 0.8960(W) + 379.93	931
	B = 0.3185(W) + 0.4388(ST) − 7.68	517
V(3) MINUS V(2):		
Metatarsal (139)	ST = 0.4421(W) + 456.11	704
	B = 0.1270(W) + 1.0225(ST) − 20.93	249
Femur Epiphysis (137)	ST = 0.7873(W) + 776.07	836
	B = 0.2264(W) + 0.8539(ST) + 216.47	455
Femur Shaft (95)	ST = 0.8982(W) + 971.43	1028
	B = 0.0237(W) + 0.7590(ST) − 279.78	542

the bone density data—race and sex differences, and, in the $V(1)$-$V(2)$ interval, antenatal treatment differences. No suggestion of treatment effects was detected; nor was any consistent difference associated with race or sex. For example, the adjusted mean $V(1)$-$V(2)$ increase in femur epiphysis density was greater in Negro females than in white females by 164 units ($P = 0.01$—0.02), but this was not confirmed in males or in the other bones, and can be most plausibly interpreted as one of those "significant" differences that must be expected when many tests are applied to the same data.

Mean Soft Tissue Density Changes. Regarding the soft tissue changes in Table 18, three points are noteworthy:

1. All three areas of soft tissue show an increase in density, no doubt attributable chiefly to increase in thickness.
2. Although the average $V(2)$-$V(3)$ interval was five times the average $V(1)$-$V(2)$ interval, the soft tissue increases dif-

TABLE 18. ADJUSTED MEAN CHANGES OF DENSITY (IN DENSITOMETER UNITS) AFTER THE FIRST POSTNATAL WEEK.

Estimates of intervisit changes from equations in Table 17, i.e., mean ST change for zero change in W, and mean B change for zero change in W and ST.

Densities at V(1) for soft tissue are the observed arithmetic means from Table 2; for bone, the estimated arithmetic means with ST equated to the density of the mean black area of the films. P = random sampling probability.

Mean intervals: V(1) − V(2) = 29.3 days; V(2) − V(3) = 145.9 days.

		Visit(2) minus Visit(1)			Visit(3) minus Visit(2)		
Tissue	Mean at V(1)	Adj. Mean Change	Per Cent of V(1)	P	Adj. Mean Change	Per Cent of V(1)	P
Soft Tissue:							
Metatarsal	700	+ 45	+ 6.4	0.2–0.3	+456	+65	v. small
Femur Epiphysis	1821	+182	+10.0	<0.01	+776	+43	v. small
Femur Shaft	3007	+380	+12.6	v. small	+971	+32	v. small
Bone:							
Metatarsal	400	− 41	−10.2	0.02	− 21	− 5.2	0.3–0.4
Femur Epiphysis	1158	+ 25	+ 2.2	0.2–0.3	+216	+18.7	v. small
Femur Shaft	3003	− 8	− 0.3	0.8–0.9	−280	− 9.3	v. small

fered from the 5 to 1 ratio—at metatarsal, 10 to 1; at femur epiphysis, 4 to 1; at femur shaft, 2.6 to 1. External measurements might have elucidated these differences; but the only limb-girth measurement that was made in the Nutrition Study, the calf circumference, had low precision (Miss Wiehl—personal communication). Lateral radiographs, taken under highly standardized conditions, would be necessary for a more precise determination of changes in thickness of muscle and fat.

3. When the intersubject standard deviations that accompany the soft tissue equations in Table 17 are considered in relation to the mean changes in Table 18, and when the frequency distributions are visualized as roughly Gaussian, it will be seen that, although the means were positive, the changes in many infants were negative—decreases rather than increases. The differences far exceeded densitometer reading variation; and several explanations might be suggested, for example:

a) Incompleteness of the correction for general film density. In view of the rather high correlations between wedge and soft tissue density (Table 16), this hardly appears likely to be the complete explanation.

b) Difference in the positioning of the limbs on the two visits, so that in some infants, although the limbs had grown in bulk, less soft tissue was in the path of the X-rays than on the preceding visit.

c) Actual loss of volume of soft tissue. Further study of this suggestion would require antero-posterior and lateral radiographs taken without the tissue-flattening effect of pressure against the film holder or table.

Mean Bone Density Changes. The adjusted bone density changes in Table 18 were unexpected. The femur epiphysis, the most rapidly growing of the three bones, was the only one that showed positive mean intervisit differences, whereas in the other two bones the differences were either approximately zero or significantly negative—metatarsal in $V(1)$-$V(2)$ and femur shaft in $V(2)$-$V(3)$. These observations will be discussed later in relation to bone size changes.

In order to obtain an estimate of the percentage change in

bone density, it would be rather misleading to use the observed bone densities at $V(1)$ because they included soft tissue. Therefore, the somewhat speculative estimates, obtained by equating the soft tissue density at $V(1)$ to the density of the adjacent black area of the film, were used as denominators; and, since the intervisit changes were expressed in densitometer units (not log units), it was necessary to use the $V(1)$ estimates obtained without logarithmic transformation.

Changes in Bone Size. In Table 19 no account is taken of the fact that, when a limb has increased in thickness between two visits, the contained bone will usually be farther from the film than on the previous visit, and therefore will appear larger, even if it has not increased in size. The geometrical relationships are simple; but even if a density radiograph is supplemented by a lateral film taken exactly at right angles to it, the application of the geometry is complicated, because the shadows on the lateral film are also magnified images. Change in soft tissue density, however, is largely due to change in soft tissue thickness; therefore, it can be used to correct bone size change, and bone density changes can therefore be corrected simultaneously for changes in wedge density, soft tissue density and bone size.

TABLE 19. BONE SIZE CHANGES (CM.) AFTER THE FIRST POSTNATAL WEEK—OBSERVED VALUES, NOT ADJUSTED FOR SOFT TISSUE CHANGES.

Denominators of percentage differences are the general mean sizes at Visit(1) in Table 9: Metatarsal width = 0.491 cm. Femur epiphysis, vertical = 0.539 cm. Femur shaft width = 0.694 cm. NUMBERS OF INFANTS ARE IN PARENTHESES.

	VISIT(2) MINUS VISIT(1) Mean Difference		VISIT(3) MINUS VISIT(2) Mean Difference	
DIMENSION	cm.	Per Cent	cm.	Per Cent
Metatarsal Width	+0.035 (148)	+ 7	+0.098 (128)	+20
Femur Epiphysis Vert.	+0.141 (149)	+26	+0.369 (125)	+69
Femur Shaft Width	+0.059 (131)	+ 9	+0.269 (95)	+39

Tables 20 and 21 show the relationships among these variables; but the femur shaft was explored more fully than were the other bones, because, in order to maintain an approximately circular cross-section, it would increase by approximately the same amount in thickness as in width.

After elimination of differences in wedge density, the correlations between the changes in femur shaft width and in soft tissue density, although significant at the 5 per cent level, were low (+0.2 and +0.3), and the regression coefficients, converted to mm. per 100 units of increase in soft tissue density, appeared small: +0.012 mm. for $V(1)$-$V(2)$, and +0.036 mm. for $V(2)$-$V(3)$. However, this relationship, which must be partly a direct (magnification) effect and partly concomitant size increase, cannot be ignored, because if the wedge and soft tissue changes are equated to zero, the changes in femur width, in comparison with the observed (unadjusted) changes in Table 19, are as follows:

$V(1)$-$V(2)$: +0.0567 cm. instead of +0.0592 cm.
$V(2)$-$V(3)$: +0.2235 cm. instead of +0.2687 cm.

The unadjusted value is, therefore, about 4 per cent greater than the adjusted value in $V(1)$-$V(2)$ and about 20 per cent greater in $V(2)$-$V(3)$. It will be observed that the adjusted increase in size in $V(2)$-$V(3)$, a five-month interval, is about four times the adjusted increase in $V(1)$-$V(2)$, a one-month interval. When the femur width changes were adjusted for soft tissue changes, none of the race or sex differences in the amount of width change were significant at the 5 per cent level.

Relationships of Bone Density Change and Size Change. In Table 20, three of the four $B \times S$ correlation coefficients are positive, but small; for the femur shaft the coefficient in $V(2)$-$V(3)$ is very small and negative. When wedge and soft tissue changes were both eliminated in the femur shaft, the correlation between bone density change and size change became approximately +0.1 in $V(1)$-$V(2)$ and −0.1 in $V(2)$-$V(3)$—values that would often occur in random sampling from a popu-

TABLE 20. CORRELATIONS OF DENSITY CHANGES (IN DENSITOMETER UNITS) AND BONE SIZES (IN CM.) AFTER THE FIRST POSTNATAL WEEK.

B = BONE. ST = SOFT TISSUE. W = WEDGE.

B = BONE SIZE, I.E., WIDTH IN METATARSAL AND FEMUR SHAFT, VERTICAL AXIS IN FEMUR EPIPHYSIS.

In metatarsal and femur epiphysis, bone density changes were adjusted for zero change in wedge and soft tissue, but bone sizes were not corrected for soft tissue change. Correlations in femur shaft begin with observed (unadjusted) intervisit differences and show stages in elimination of W and ST.

P = RANDOM SAMPLING PROBABILITY.

Numbers of infants: femur shaft V(1) − V(2), 131; femur shaft V(2) − V(3), 95; metatarsal, 148; femur epiphysis, 149.

		VISIT(2) MINUS VISIT(1)		VISIT(3) MINUS VISIT(2)	
BONE	CORRELATION	Coefficient	P	Coefficient	P
Femur Shaft	B × S	+0.206	0.02–0.05	−0.084	0.4–0.5
	ST × S	+0.213	0.02–0.05	−0.053	0.6–0.7
	W × S	+0.119	0.2–0.3	−0.223	0.02–0.05
	B × S.W	+0.222	0.01–0.02	+0.204	0.05 approx.
	B × S.ST	+0.031	0.7–0.8	−0.111	0.3–0.4
	ST × S.W	+0.216	0.02–0.05	+0.314	<0.01
	B × S.W + ST	+0.124	0.1–0.2	−0.101	0.3–0.4
Metatarsal	B × S	+0.119	0.1–0.2		
Femur Epiphysis	B × S	+0.266	<0.1		

TABLE 21. LINEAR REGRESSION RELATIONS BETWEEN DENSITY CHANGES (IN DENSITOMETER UNITS) AND CHANGES IN BONE SIZE (IN CM.) AFTER THE FIRST POSTNATAL WEEK.

The equations were derived from the same data as were the correlation coefficients in Table 20.

SD = INTERSUBJECT STANDARD DEVIATION, REPRESENTING DEVIATIONS FROM REGRESSION.

BONE AND INTERVAL	REGRESSION EQUATION	SD
Metatarsal:		
V(2) − V(1)	B = 673.0(S) − 64.6	208 units
Femur Epiphysis:		
V(2) − V(1)	B = 1046.0(S) − 121.8	268 units
Femur Shaft:		
V(2) − V(1)	S = −0.000007535(W) + 0.00001234(ST) + 0.0567	0.0495 cm.
	B = 0.3269(W) + 0.4252(ST) + 1104.56(S) − 70.3	526 units
V(3) − V(2)	S = −0.00004424(W) + 0.00003551(ST) + 0.2235	0.106 cm.
	B = 0.003545(W) + 0.7751(ST) − 490.31(S) − 168.6	542 units

Met. V(3) − V(2), from unweighted *means* of 16 race-sex-treatment groups:
B = 0.0869(W) + 1.0022(ST) + 283.40(S) − 41.9

lation in which there was no correlation at all.

Such low correlations, even if real, account for very little of the total intersubject variation. Thus, comparison of the intersubject standard deviations in Table 21 with the corresponding values in Table 17 shows that the correction of density changes by size changes adds essentially nothing to the precision of the estimates of bone density changes.

The equations in Table 21 represent the regression relationships corresponding to the correlation coefficients in Table 20. For example, the equation for the femur shaft density change in the $V(1)$-$V(2)$ interval would lead one to expect that, for any particular amount of change in wedge or soft tissue density, a width increase of 1 mm. would, on the average, be associated with a density increase of approximately 110 units, but in the $V(2)$-$V(3)$ interval one would expect an average *decrease* of 49 units, in agreement with the negative correlation ($B \times S.W + ST$) in Table 20.

Again, the constant at the end of each bone density equation is the estimate of the average change that would occur if there were no change in wedge density, soft tissue density or bone size—provided, of course, that the regression was linear throughout. Comparison of these figures with the corresponding values in Table 18, which were not adjusted for zero change in bone size, suggests that, if bone size had not increased, the densities would have decreased between the visits more than they actually did, and in the femur epiphysis the density would have decreased instead of increasing.

An Interpretation of Size-Density Relationship. If we accept the suggestion from Table 20 that the apparent size-density correlations were merely fortuitous, all bone density estimates from Table 21 become meaningless; but the combined evidence from the first week and the intervisit intervals merits more than such a casual dismissal, and the following considerations seem to elucidate the problem.

One would expect that, on the average, a larger bone would

contain more mineral matter than a smaller bone of the same kind. Therefore, at any one age (for example, the first postnatal week) one would expect a positive correlation between size and X-ray density, and when bones increased in size, as they did in the two intervisit intervals, one would expect an average increase in density and a positive correlation between size increase and density increase. When these expectations are not fulfilled, two explanations (which are not mutually exclusive) may be considered:

1. Masking of the size-density relationship by the damping effect of soft tissue or by the defects of the densitometric method, such as reading error or imperfect correction for differences in general film density and in soft tissue density.
2. A tearing down (remodelling) of the interior of bones while the girth increases.

These suggestions can be considered in relation to the findings in this study:

1. At Visit(1): A definite positive correlation between density and size in the femur epiphysis, but no convincing evidence of any correlation in the metatarsal, or in the femur shaft after correction of size by soft tissue differences (Table 11).
2. In the intervisit interval: A density increase in the femur epiphysis; decreases of density in the metatarsal and femur shaft, significant at the 5 per cent level in two instances (Table 18). No convincing intragroup correlation between density change and bone size change after correction for wedge and soft tissue (Table 20).

It is difficult to conceive how the mere masking of a positive size-density relationship could account for these diminutions in density, and it appears more likely that the densitometer has detected the well-known phenomenon in normal bone development—destruction accompanying construction. In the middle of a long bone, such as the femur, the process results in the

enlargement of the medullary cavity, but even in a region that remains cancellous throughout life, as do the ends of long bones, the spaces become larger as the bone grows. In any particular bone the amount of internal destruction per unit increase in bone girth doubtless varies from child to child, even within the same narrow age range, but these individual differences cannot be distinguished by studying the mean density change of a group, or the correlation between density change and size change.

Among the three bones under examination, the femur epiphysis, just before and just after birth, undergoes rapid growth, which is primarily a spread of calcification into the cartilaginous epiphysis. The shafts of metatarsal and femur have passed through this phase by the end of the third fetal month. At the time of birth their growth in circumference is due to subperiosteal deposition, accompanied, it would appear, by internal destruction which neutralizes the size-density relationship.

If we adopt this interpretation, and accept the relationships in Table 21 as more than fortuitous, Table 22 will provide a

TABLE 22. CHANGES IN DENSITY CORRESPONDING TO CHANGES IN BONE SIZE BETWEEN VISITS (1) AND (2)—ESTIMATES FROM EQUATIONS IN TABLE 21.

Denominators for percentage changes are general mean sizes at V(1) in Table 9, and estimated arithmetic mean bone densities in Table 18.

Bone Size Change	Estimated Average Bone Density Changes					
	Metatarsal		Femur Epiphysis		Femur Shaft	
Per Cent of V(1)	Units	Per Cent of V(1)	Units	Per Cent of V(1)	Units	Per Cent of V(1)
0	− 65	−16.3	−122	−10.5	− 70	− 2.3
+10	− 31	− 7.8	− 65	− 5.6	+ 3	+ 0.1
+20	+ 3	+ 0.8	− 9	− 0.8	+ 84	+ 2.8
+30	+ 36	+ 9.0	+ 48	+ 4.1	+162	+ 5.4
+40	+ 70	+17.5	+104	+ 9.0	+239	+ 8.0
+50	+104	+26.0	+161	+13.9	+316	+10.5

picture of the size-density relationships in the $V(1)$-$V(2)$ interval. The figures suggest the presence of two opposing forces acting on the bones—one force (size increase or something associated with it) tending to increase the X-ray density, and the other force tending to reduce it. Unless the size increase exceeds a certain amount, differing for each of the bones (and presumably for the same bone in different infants), it cannot prevent a decrease in density. The femur epiphysis was the only bone in either interval that showed a dominance of the density-increasing factor (Table 18), whereas in the femur shaft there appeared to be such a balance between the two tendencies that the size-density relationship was questionable in both intervals.

Required Sample Sizes. Table 23 is analogous to Table 13,

TABLE 23. ESTIMATES OF MEAN BONE DENSITY CHANGES REQUIRED FOR SIGNIFICANCE (P LESS THAN 0.05) AFTER ADJUSTMENT FOR CHANGES IN WEDGE DENSITY, SOFT TISSUE DENSITY AND BONE SIZE.

For single samples the estimates (in densitometer units) represent twice the SD of the mean change, estimated from the intersubject SD's in Table 21. For comparing mean changes in two samples the required difference is twice the SD of the difference between means.

The denominators of the percentage differences are the estimated arithmetic mean densities at V(1) in Table 18.

N = NUMBER OF INFANTS IN EACH SAMPLE.

	METATARSAL, V(2) − V(1)		FEMUR EPIPHYSIS, V(2) − V(1)	
N	Mean Change	Sample Diff.	Mean Change	Sample Diff.
25	84(21.0 Per Cent)	119(29.8 Per Cent)	108(9.3 Per Cent)	153(13.2 Per Cent)
50	60(15.0 Per Cent)	85(21.3 Per Cent)	76(6.6 Per Cent)	108(9.3 Per Cent)
100	42(10.5 Per Cent)	60(15.0 Per Cent)	54(4.7 Per Cent)	77(6.6 Per Cent)
	FEMUR SHAFT, V(2) − V(1)		FEMUR SHAFT, V(3) − V(2)	
N	Mean Change	Sample Diff.	Mean Change	Sample Diff.
25	212(7.1 Per Cent)	300(10.0 Per Cent)	217(7.2 Per Cent)	307 (10.2 Per Cent)
50	150(5.0 Per Cent)	213(7.1 Per Cent)	154(5.1 Per Cent)	217(7.2 Per Cent)
100	105(3.5 Per Cent)	149(5.0 Per Cent)	109 (3.6 Per Cent)	154(5.1 Per Cent)

and is based on similar assumptions (p. 56). Both tables show the relationship between sample size and the precision or sensitivity of the method, but Table 13 is concerned only with the comparison of two groups of infants of approximately the same age, whereas Table 23 is concerned with two types of comparison:

1. The change in the same group of infants between one visit and another.
2. The difference of this change in two groups of infants, e.g., race groups, sex groups or treatment groups.

Relationship of Bone Density to Length of Intervisit Interval. In Table 18 a rough estimate of average daily change in bone density can be found by dividing the mean changes by the mean intervals, approximately 30 days and 146 days; but this does not reveal the relationship of time and density change within a group. For that purpose, the following correlation coefficients were estimated for the femur shaft in the $V(1)$-$V(2)$ interval, the density changes being adjusted for zero change in wedge and soft tissue:

Bone Density Change × *Interval* in 42 white males: +0.181 (P = 0.2–0.3)
Bone Density Change × *Interval* in 31 white females: –0.015 (P large)

For comparison, the coefficients involving width change in the white males were:

Bone Width Change × *Interval:* +0.408 (P less than 0.01)
Bone Width Change × *Bone Density Change:* +0.282 (P = 0.05–0.1)

Although the changes in density and size were positively correlated, and the size change and interval were positively correlated, there was no clear-cut correlation of density change and length of interval, and this can perhaps be taken as a reflection of the conflict, already discussed, between destructive and constructive phenomena.

For further exploration, the longer interval, $V(2)$-$V(3)$, was used. In the four race-sex groups the correlation between length of interval and femur shaft density change (adjusted for zero changes in wedge, soft tissue and bone width) ranged from –0.203 to –0.393, and for the pooled data, from 87 infants, the coefficient was –0.333 (P less than 0.01). That is, if the bone size change were zero, or any other fixed value, the longer the interval the less the increase (or the greater the decrease) in bone density. This, again, suggests a destructive force, manifesting itself when it is not masked by the addition of material to increase the size of the bone.

Relationship between Bone Density Change and Initial Density. A question that is of interest in all studies of infant development can be applied to bone density in this form: Do infants who are below the average in density at birth tend to make up the deficit in the first few months, and do those who are advanced at birth tend to gain more slowly than the less advanced?

It was realized that an attempt to answer that question by correlation of density change with initial density would introduce the risk of an artifact through errors in reading. For example, let it be supposed that two infants have, in actuality, the same $V(1)$ density and the same $V(2)$ density, greater than $V(1)$. In one infant there are no errors in density reading, but in the other the $V(1)$ density is underestimated and the $V(2)$ reading is accurate. That will automatically increase the apparent intervisit gain. Such errors would tend to produce negative correlations between the gain and the initial density. However, the magnitude of these effects of reading error is commonly small in studies of gross development (e.g., weight changes); therefore this method was applied to data from femur shafts in 42 white males and 31 white females, i.e., $V(1)$ bone densities (log units) and $V(1)$-$V(2)$ changes (scale units), both variates having been adjusted for differences in wedge and soft tissue densities. The correlation coefficients

were –0.553 in white males and –0.740 in white females (P in each group less than 0.001).

These values were so large as to create suspicion of an artifact much greater than the reading error would be expected to create. No clues were provided by the bone size or length of the intervisit interval, the correlation coefficients in the 42 white males being as follows:

$V(1)$ *density* × $V(1)$-$V(2)$ width change: +0.027
$V(1)$ *density* × $V(1)$-$V(2)$ interval: –0.061

An explanation was then sought by making two rather plausible assumptions: (a) that much of the intersubject variation in bone density at $V(1)$, and again at $V(2)$, was due to the failure of the average adjustments for wedge, soft tissue and bone size to correct for these factors fully in individual films, and (b) that the amount of this discrepancy in any infant at $V(1)$ had no relationship to the amount of the discrepancy in the same infant at $V(2)$. Then, although the overall intersubject variation might be the same on both visits, the position of the infants in the density series at $V(2)$ would to a large extent be independent of their position at $V(1)$, as if their positions had been assigned almost at random; and this would be true whether or not there was a significant mean change in density. If there were random assignment, the infants who had very low readings at $V(1)$ would be likely to have higher readings at $V(2)$, rather than equally low or lower readings, and those with very high readings at $V(1)$ would tend to have lower readings at $V(2)$. This speculation led to randomization experiments which can be summarized thus:

For each of the 31 white females the bone density at $V(1)$, adjusted for wedge and soft tissue, was estimated in actual units. Then the differences among these (the deviations from regression) were redistributed among the infants by random numbers, to produce a fictitious "$V(2)$" reading and "intervisit change" in density. The correlation between this "change" and the deviations at $V(1)$ was –0.6746.

Although this result seemed to confirm the speculation, it was somewhat artificial, because the actual deviations at $V(1)$ had been used in the randomization. A more realistic conception was that the factors, whatever they might be, that produced intersubject variation between adjusted bone densities at $V(1)$ were still operating at $V(2)$—that is, the standard deviation of the $V(2)$ population would be the same as that of the $V(1)$ population; but the actual multiples (or fractions) of the standard deviation that the individuals in the sample received at $V(2)$ would be a random sample of possible multiples (or fractions). Such random samples, for normal (Gaussian) distributions are obtainable from the Rand Tables,[31] and so a random sample obtained from those tables was distributed to the V(1) films to create again a fictitious "change" in density. The correlation between this "change" and the actual deviations from regression at $V(1)$ was –0.6550, again in agreement with the expected relationship.

To imitate the $V(3)$ minus $V(2)$ change in density, the $V(1)$ intersubject variation was reassigned by use of the Rand table to the $V(1)$ densities of the 31 infants in the WF group. The readings so created were labeled "$V(3)$" and from them were subtracted the "$V(2)$" readings created by the first randomization. The correlation between these "changes" and the actual $V(1)$ deviations from regression was +0.0366. As might be expected, the correlation that had been found after the first randomization had disappeared. In the real data there was still a trace of it. For the femur shaft, the correlations in the four race-sex groups were as follows:

WM, –0.2924; *WF*, –0.1969; *NM*, –0.2194; *NF*, –0.2737; all groups pooled, –0.2449 ($P = 0.02$–0.05).

These negative correlations were not explored further, because the evidence was sufficient to show the risk of an artifact in the correlation of change with initial value. The risk may be very general—for example, the correlation between change in size (or weight) and initial size (or weight). Thus, the artifact may have been responsible for the negative correlation between

the $V(1)$-$V(2)$ femur width increase and the femur width at $V(1)$ in 42 white males: –0.2606 ($P = 0.05$–0.1).

Direct Correlation between Densities at $V(1)$ and $V(2)$. If the original purpose had been to study the relationship between $V(1)$ and $V(2)$ densities, the appropriate method would have been to find those densities (corrected for wedge, soft tissue and bone size) separately for each visit, and then estimate the regression of the $V(2)$ density on the $V(1)$ density. In a conceptually perfect experiment (that is, with no residual variation due to technique or observation), if each bone changed its density by exactly the same amount, the regression would be linear and, in a graph with equal scales for $V(1)$ and $V(2)$, the line would slope upward at an angle of 45 degrees to the horizontal. If the bones that were initially below average density increased more than the initially dense bones, the angle would be less than 45 degrees, or the line might even slope downward.

As a substitute for the independent estimation of adjusted $V(2)$ densities, the adjusted changes in femur shaft densities from $V(1)$ to $V(2)$ in the WF group (31 infants) were added (algebraically) to the adjusted $V(1)$ densities. When these $V(2)$ values (varying from 3508 to 5164 units) were plotted in a scatter diagram against the $V(1)$ values (3403 to 5082 units) there was no suggestion of an upward or downward trend, and the correlation coefficient was very low (+0.0423). However, the disposition of the dots was quite compatible with a regression line sloping upward at 45 degrees, or curving upward (indicating a greater increase for initially high values), or with a line that sloped (or curved) downward. Two explanations of this inconclusive evidence appear possible:

1. The actual changes in bone density varied so much between subjects that they destroyed any general relationship between $V(1)$ and $V(2)$ densities.

2. The corrections for general film density and for soft tissue density were imperfect; that is, the intersubject variation in

bone density at $V(1)$ and again at $V(2)$ was largely due to one or both of these factors.

The second explanation appears likely to be the chief one, because it is difficult to suppose that in a month the actual changes in bone density varied so much as to leave at $V(2)$ no trace of a relationship to the density at $V(1)$.

SUMMARY OF RESULTS

In the following statements, "bone density" means X-ray bone shadow density corrected, by linear regression, for differences in soft tissue density and differences in general film density, as measured by aluminum wedge density.

First Postnatal Week. The most noteworthy results can be summarized as follows:

1. There was found no evidence of a relationship between the infants' bone densities and the dietary supplements prescribed to their mothers during pregnancy.
2. In both sexes the mean density of the metatarsal shaft in white infants exceeded the mean density in Negro infants of the same sex by amounts that could not be easily attributed to random sampling variation. Nor could they be accounted for by the racial difference in metatarsal size (as measured by width), because it was in the opposite direction to the bone density difference. No other significant racial or sex difference in density was found in the three bones examined (metatarsal, distal femur epiphysis and femur shaft).
3. After correction for differences in bone size, the residual intersubject variation in density in all three bones was low enough to justify a verdict of "significant at the 5 per cent level" if, in two samples of 100 infants, the greater mean density exceeded the smaller by 5 per cent.
4. Although the femur shaft width showed correlations of

about +0.4 with weight and with crown-sole length, there was no evidence of a correlation between femur shaft density and these body measurements.

5. The correlations between bone size and bone density were positive in all three bones, but only in the femur epiphysis was the evidence of a relationship conclusive. In the femur shaft, after the size differences had been corrected for differences in magnification (by using the relationship of size and soft tissue density), the size-density correlation was low (+0.194) and did not quite reach the 5 per cent level of significance. In the metatarsal the coefficient was very low.

Intervisit Changes. The visits were: Visit(1), in the first postnatal week; Visit(2), at median age approximately one month; Visit(3), at median age approximately six months. A change in bone density, real or apparent, was expressed as: later visit minus earlier visit, with positive or negative sign. Bone density changes were corrected, by linear regression, for changes in densities of wedge and soft tissue. Femur shaft data were analyzed more fully than data from the other two bones because, owing to the roughly circular cross section of the bone, the width, measurable on the film and corrected for magnification, would approximate the thickness (antero-posterior axis).

1. No treatment effects on density change were detected in the Visit(1)-Visit(2) interval, and there were no consistent race or sex differences, either in that interval or in the Visit(2)-Visit(3) interval. Because of drop-outs and irregular attendances, these results could not be accepted as bias-free, but they permitted the pooling of data for the study of certain relationships between density change, size change and length of intervisit interval.

2. The femur epiphysis was the only bone that clearly showed increase of density between the visits. In the metatarsal and femur shaft the differences were either effectively zero or, in two instances, negative (P less than 0.05).

3. On the unprovable assumption that the observed regres-

sions of bone density on wedge and soft tissue densities at Visit(1) would be applicable even if soft tissue were absent, estimates of bone density, free of soft tissue, were obtained by substituting the mean densities of the black (shadow-free) areas for soft tissue in the regression equations. These rather speculative bone densities were then used as denominators for expressing intervisit density changes as percentages of the first-week densities in the estimation of sample sizes necessary to show statistically "significant" mean changes in bone density. With samples of 100 infants, the following mean changes in the $V(1)$-$V(2)$ interval would be significant at the 5 per cent level: metatarsal, 11 per cent; femur epiphysis, 5 per cent; femur shaft (in both intervals), 4 per cent.

4. In the femur shaft, in the $V(1)$-$V(2)$ interval, there was no detectable correlation between the change in density and the length of the interval; in the $V(2)$-$V(3)$ interval, after elimination of femur width changes, the correlation was negative and significant at the 1 per cent level.

5. In the femur shaft there were negative correlations (–0.55 in one race-sex group, –0.74 in another; P less than 0.001) between the $V(1)$ densities and the $V(1)$-$V(2)$ changes in density; but negative coefficients of similar magnitude were obtained by taking the intersubject variation in $V(1)$ densities and distributing it by random numbers to the $V(1)$ values, to create an artificial $V(2)$ density and an artificial $V(1)$-$V(2)$ change, and then finding the correlation between the actual $V(1)$ densities and these "changes." When this kind of artifact was avoided by correlating $V(1)$ and $V(2)$ density values for the femur shaft, there was no suggestion of a relationship.

DISCUSSION OF BONE DENSITOMETRY IN GENERAL

In his editorial introduction to the Transcript of the Workshop on Bone Densitometry[11] Garn provides a very pertinent and comprehensive list of eight problems that should be explored in an effort to evaluate densitometry. In the following discussion, Garn's presentation of each problem will be quoted, along with comments arising from the present study or previous studies in this laboratory, without implying either disagreement with the quoted statements or full treatment of the topics. Then a number of other problems will be discussed.

Underlying Problems. The complexities of the individual problems are, of course, simplified if we recognize the two underlying problems—lack of sensitivity and risk of bias.

Two kinds of sensitivity-reducing factors should be distinguished:

1. Factors that increase intersubject variation, i.e., reduce the precision of an intergroup comparison of bone shadows.
2. Factors that mask differences between amounts of mineral matter. These factors include the quality of the radiation employed, the film emulsion and, probably of greatest im-

portance, the soft tissue on the bones.

In other words, the distinction is between (a) sensitivity to intergroup differences in bone shadow density, and (b) sensitivity to intergroup (or interindividual) differences in mineral content. (a) is a necessary, but not sufficient, condition of (b). In many discussions of densitometry this distinction does not seem to be very clear. Except where specific reference is made to mineral content, the following discussion will refer to bone shadow density.

The second underlying problem, the risk of bias, makes it necessary to bear constantly in mind the distinction between a survey and an experiment in the strict sense (p. 26). Because in a properly conducted experiment randomization controls bias—that is, it enables us to attach a numerical statement of error to an inference regarding an intergroup difference, e.g., in the comparison of treatments.

Observational Reliability. "The absolute reliability of current roentgenogrammetric methods of densitometry needs to be determined. Appropriate short-term and long-term reliability coefficients involving inter-observer and intra-observer determinations have yet to be published".[11]

In its technical sense, a "reliability coefficient," as used by psychologists, some students of nutrition and other workers, is often a correlation coefficient, such as can be estimated when an observer examines the same film on two occasions or when two observers examine the same series of films independently. Like all correlation coefficients, it can be deceptive. For example, a coefficient of + 0.95 is impressive, but it does not give us the information that we need, i.e., answers to two questions:

1. Has the systematic intra- or interobserver difference produced a spurious difference (or masked a real difference), for example, between groups in different regions or between groups in the same region examined at different times?
2. Did the intra- or interobserver variability seriously reduce

the sensitivity of comparisons?

To answer those questions, we need estimates of intra- and interobserver error (systematic and variable), but the publication of such information would do little more than show that under such and such conditions it was, or was not, possible to achieve comparability that would allow the detection of intergroup density differences of a certain magnitude. To use such estimates in the actual analysis of data from other investigations, even by the same observers, would be dangerous.

In an experiment, the chief reason for determining intraobserver variation in densitometry is to ascertain how much reduction of intersubject variation can be achieved by examining the films more than once, as was done in density studies of adult bones.[29, 30] In surveys, however intra- and interobserver error imply much more than differences due to an observer and his densitometer, and they are best considered as a part of the overall observational variation along with variations due to radiographic technique.

Variations Due to Technique. "The extent of errors introduced by 'uncontrollable' variations in kilovoltage, exposure, type of radiation, films, holders, processing and aging needs to be known. Certainly some of the interstate differences in the Western Regional Survey must stem from technical sources of error".[11]

For the present discussion we include with these errors the observer (and densitometer) errors, and we consider all this variation associated with (a) time, i.e., differences at different times in the same clinic or laboratory, and (b) place, i.e., interregional differences. We assume that correction for differences in general film density is made, but we cannot assume that it is equally applicable to all films throughout an investigation.

As with observer reliability, publication of the magnitude of the effects of radiographic variation would merely show what precautions should be taken, and what could be achieved under certain conditions. For the techniques employed in this labora-

tory, this has been demonstrated on films from living adults who paid two visits to the X-ray unit[29, 30] and also on films of cadavar bone;[26] but it would not be safe to use the actual estimates of technique effects in analyzing data from another experiment or survey.

The "error term" to be used in comparing group mean densities is not observer variation or technique variation or both together, but intersubject variation, which contains them along with true intersubject differences. In an experiment, the random assignment of treatments controls the bias due to any or all of these elements of variation. If a survey (e.g., for the study of differences associated with race, sex and age) is conducted in the same location by the same observers, bias due to technical and observational factors can likewise be controlled, provided that the subjects are investigated in strictly random order. Such randomization is not always easy, especially if the investigator must depend on volunteer subjects; but in its absence, doubt regarding conclusions must always remain as, for instance, regarding the sex difference in the densities of certain adult hand bones,[30] because most of the women were X-rayed at one period in the survey.

If two geographical regions are to be compared for bone density, with radiography at different centers, it seems essential that (a) the standard object be the same on all films (e.g., a precision-milled aluminum sheet or wedge, split in two halves), (b) all densitometry be done in one laboratory, and (c) all films from both regions be randomly arranged before reading.

Even in spite of these precautions, bias may enter. If, for example, the cardboard film holders at one center tilted the aluminum wedge slightly, whereas at the other center the wedge was more horizontal, the correction for general film density (whatever form it took) would not be strictly applicable to either set of films. It would not be feasible for a sufficient number of persons to be X-rayed at both centers. Therefore, the best substitute would appear to be cadaver bones mounted in wax—a number of different specimens of the same type as

the living bones under study, and exchanged frequently between the centers throughout the survey. By careful organization, it would thus be possible to obtain a kind of interregional correction term.

Shape and Size of Bones. "The shape factor needs more extensive work so that the actual shape of individual bones measured in two-dimensional and even one-dimensional densitometry can be ascertained".[11]

Bone shape concerns us because of its relationship to bone size, particularly the thickness, i.e., the axis traversed by the X-rays. In an experiment, the bones of different shapes and sizes are randomly assigned to treatments, and in adults the inference connecting treatment difference with difference in density change is direct, if the effect of treatment on soft tissue can be eliminated. Correction for bone size, therefore, merely increases the sensitivity of the experiment, and the correction need not be complete. In children, however, an apparent increase in density may be due to an increase in thickness, and it is desirable to distinguish between the two effects.

If we wish to estimate, from numerous density readings on the same bone shadow, the total mineral content or the mineral content per unit volume, the bone shape, in relation to bone size, presents a serious problem, and the only solution would seem to be serial body-section radiography of each bone studied. If, however, we are content to take as it were, a core-like sample out of the middle of a bone shadow (or out of the middle of a particular region, such as a metacarpal head), and correct the intersubject differences in density by regression on bone size, the problem is not so difficult. It may well be that the dimension measurable on the density film is not highly correlated with the "thickness," and then it would be desirable to obtain a film at right angles to the density film, probably by body-section radiography because of overlap of bone shadows; but it would hardly seem necessary to find "exactly" the same plane as the one traversed by the X-rays in producing the

"sample core."

Soft Tissue. "The interfering effects of the various soft tissues need further study. Water-immersion and wax-immersion techniques introduced by European investigators merit evaluation here. The limits set by soft tissue need to be given in table form."[11]

As has been previously emphasized, it is important to recognize the distinction between (a) the damping effect of soft tissue, which cannot be removed by correction factors or, presumably, by immersion in water or wax, and (b) the effect of differences in soft tissue density which can be removed to a large extent, e.g., by regression of bone density on soft tissue density. In other words, correction for soft tissue is not equivalent to removal of it; and the substitution of the density of the black (shadow-free) area for soft tissue in the regression equations rests on an assumption not provable on the films themselves. The same is true of the "subtraction" method of correcting for soft tissue.

The question, then, is: What can be told about "true" differences in bone density or mineral content after correction for soft tissue differences? That question will be discussed below; but it appears unlikely that tabular presentation of the limits set by soft tissue, as found by one investigator, would apply, except in a general way, to the work of others, or even to his own later work, without re-investigation.

One way of trying to remove the effect of soft tissue is to place the standard object (a step-wedge or bone) between the limb and the film, and at the thigh this can be done without much risk of overlap of the shadows of femur and wedge. This method has not been often used, but recently Heuck and Schmidt[13] have applied it, using an apatite step-wedge of known composition. For the calcaneus, they immerse the foot and the wedge in water; and for both femur and calcaneus they find bone thickness from films taken at right angles to the density films. The bone mineral content is estimated as mg.

of hydroxyapatite per ml. of bone.

The report on this method, applied to the femoral neck for example, seems to imply much greater faith in the sensitivity of densitometry in muscle-laden regions than is held by many radiologists, who doubt whether in the hip region a change in mineral content of less than 30 per cent can be detected. The danger of bias also seems to receive too little attention. For example, to estimate the mineral content of the femoral neck, the wedge is placed behind the medial part of the upper third of the thigh, whereas the soft tissue that overlies the femoral neck includes the massive gluteus maximus.

Scintillation Counters. "The possibility of scintillation counters has barely been explored. They offer the advantage of direct measurement eliminating the troublesome film and processing . . . the scintillation counter can be combined with a film".[11]

Without experience of this method, few critical comments are possible. The soft tissue problem would remain; but the taking of a film would permit allowance for differences in bone size.

The Meaning of Bone Density . "Though there is excellent agreement on *how* body sections are measured in radiographic densitometry, there is very little concern with exactly *what* is being measured. In the Workshop, participants were unable to come up with a satisfactory definition of what the trace curve was, what the 'mass coefficient' really included and exactly what 'bone density' measured".[11]

In this laboratory, "bone density" is simply bone shadow density, i.e., a densitometer scale reading made under certain conditions and, in previous work, translated into other units, e.g., mm. of thickness of a certain aluminum wedge. Differences between such readings are sometimes partly due to differences in the amount of mineral matter traversed by the X-rays in creating the shadows. The question, how large a difference

must be before it can be so explained, must be decided empirically; but the same is true of the more elusive concepts of bone density which are associated with more elaborate instruments and techniques.

Equivalents of Bone Density. "Though it was agreed that the aluminum alloy utilized by several laboratories represented the most practical wedge material, it was considered that the present wedge-equivalent values might be translated into calcium or calcium-salt equivalent values".[11]

For a time, ivory step-wedges were used instead of aluminum wedges at Pennsylvania State College, because ivory is akin to bone in its chemical, physical and radiographic properties. The advantage seems to be somewhat illusory, however, because one of the most important things measured in bone densitometry is soft tissue density, and that tissue does not resemble bone in its radiographic properties. There appears to be no proof that a correction for soft tissue, by whatever method, is more reliable if derived from ivory wedge readings than if derived from readings of aluminum.

If the more recent suggestion, that aluminum-alloy equivalent values be translated into calcium-salt values, were carried out, it would probably create the impression that the investigator, using a densitometer, learned more about the composition of a particular bone than, in fact, he does. In the present state of ignorance regarding the causes of discrepancies between the results obtained at different laboratories, expression of density as calcium-salt (e.g., apatite) values would not appear to be helpful. A different attitude might be adopted to the use of apatite wedges themselves, if the technique of Heuck and Schmidt,[13] mentioned above under *Soft Tissue,* substantiated its claim to sensitvity and lack of bias.

Scope of Densitometry. "There was evidence [in the workshop] that roentgenogrammetric densitometry, as presently carried out, is most applicable to short-term studies on adult

individuals where relatively large changes in bone mineral can be expected. . . . In contrast, the purely survey utilization of the method, involving large numbers of individuals of different ages (and with different degrees of mineralization) would seem to be beyond the present technical limits of the method despite the fact that such surveys are unquestionably of interest to nutritional research".[11]

Short-term studies of human beings in any field of research have great advantages over long-term studies, because of the risk of drop-outs and of secular changes in circumstances, including techniques and equipment; but the crucial points in the quoted statement seem to be:

1. The observation of *changes*—that is, longitudinal studies rather than the cross-sectional contrasts of groups.
2. The magnitude of the change to be expected.

The emphasis on the individual should not mislead us because, unless the change in an individual is so great as to render densitometry unnecessary, a group of individuals is needed to show what allowance must be made for observational error.

The preference, expressed in the quoted statement, for adults as subjects in whom large changes can be expected, has hardly been supported by the present study of films from infants.

It is true that a cross-sectional survey of any variable is less likely to detect the effects of factors, such as age and diet, than is a longitudinal study of the same variable when the same group can be followed for the necessary length of time; but it should be remembered that in a cross-sectional survey (relating, for example, to diet, race or socio-economic status) correction for age can be made.

Densitometric Surveys. Before any strict limits are prescribed for the application of densitometry, even in the difficult area of large surveys, an effort should be made to obtain the most satisfactory results of which current densitometric methods are capable; and this effort should include the applica-

tion of the principles and techniques of sample surveys, sometimes called the "epidemiological method." Some of these techniques have been already mentioned, such as the filming of subjects in random order, and devices for detecting and correcting technique differences in different geographical areas.

There are, however, other essentials in sample surveys that are often less obvious to those investigators who are most likely to be concerned with bone density studies, such as laboratory workers, radiologists, clinicians and students of nutrition. In any survey that attempts to find a causal factor, we ought continually to ask: What was it that brought these subjects into the survey, and what kept the others out? In what ways may these selection factors have introduced bias, leading to a fallacious conclusion?

Competition between Selection Rates. One such risk of bias, which is often ignored, arises from differences in selection rates (or admission rates). For example, in the Nutrition Study, a higher proportion of the Negro infants than of the white infants returned after the first visit—that is, in the $V(1)$-$V(2)$ survey the selection rate for Negroes was higher than for whites. Let us now imagine that infants who, in one way or other, did not thrive well after birth did not increase their bone density as rapidly as those who were in good condition, and that there was a greater tendency to bring an infant back to the clinic if he was not thriving than if he was doing well. This would mean that the selection rate for those with smaller increase in bone density was greater than for those whose density increased more.

For simplicity, we visualize two classes: those whose bone density increased (*I's*) and those whose density did not increase (*NI's*). Let us assume that the proportion of *NI's* in the total white infants was the same as in the total Negro infants; and in order to see how bias could arise let us exaggerate by imagining that all Negro infants were returned for Visit(2), whereas white infants were returned only if they were

not thriving, i.e., if they were *NI's*. Obviously, 100 per cent of the white infants observed in the survey would be *NI's* but less than 100 per cent of the Negroes would be in that class; or, in terms of measured density changes, the mean density increase would be greater in the Negro infants than in the whites, even although in the total infants (observed and not observed) the mean density changes were identical in the two race groups.

In less extreme, more realistic, cases it can be shown by simple arithmetic[27, 28] that the same kind of bias occurs. In general terms, the conditions are as follows:

1. In two populations, *A* and *B*, the proportion of *X's* is the same.
2. The selection rate of *A's* for admission to the survey is higher than the selection rate of *B's*.
3. The selection rate of *X's* is higher than the selection rate of not-*X's*.

Then, in the survey the *B's* will have a higher proportion of *X's* than do the *A's*. If a subject is an *A*, he has a relatively good chance of getting into the survey, whether he is an *X* or a not-*X*, whereas if he is a *B*, he has a relatively poor chance of admission unless he is also an *X*.

This phenomenon is a kind of competition or interplay between selection rates which was, apparently, first described in print by Berkson[1] in 1946, when he showed how it could, in the study of hospital patients, create entirely false conclusions regarding the association between two diseases, or between a disease and some supposed causal factor; but he had previously recognized the same danger in the study of autopsy data.[2] In fact, the danger can arise in any sample survey in which strictly random samples of a population, animate or inanimate, are not taken, including surveys in human anatomy and physiology;[24] and it can mask a real association as well as create a spurious one. Its possible influence on a study of age differences in human bone density has been discussed elsewhere.[29] (A somewhat detailed discussion of this problem in relation to various kinds of medical research has appeared recently.[25])

The crux of the difficulty is that we do not know the selection rates; if we did, we would know the composition of the populations, and would not need to make a sample survey. All that we can do to combat the danger is to consider in advance the possible ways in which unequal selection rates could fallaciously cause, or could cover up, associations that we intend to seek in our survey, and then search during the survey for factors that could cause such inequalities of selection rates.

For example, let us make the following suppositions regarding the Nutrition Study: the selection rate for Visit(2) was higher in Negroes (*A's*) than in whites (*B's*); the infants who did not thrive after birth had no increase in bone density; the mothers of those infants, whether Negro or white, tended to take them to private physicians rather than back to the clinic. Then those with increase of bone density would be *X*'s in our general scheme, and at *V*(2) they would be more frequent among the whites than among the Negroes, even if there were, in fact, no racial difference in amount of intervisit change in bone density. With such mechanisms of bias in mind, although an investigator could not force all mothers to bring their children back to the clinic, he might, by special inquiries, reassure himself to some extent regarding the absence of some of the selection factors.

Preparation by Pilot Studies. The foregoing remarks imply no criticism of the Nutrition Study, which began over fifteen years ago. In those fifteen years much has been learned about human sample surveys, and about clinical trials, which are analogous to the dietary experiment. Nowadays, those who are acquainted with what has been learned would start with a pilot study, in order to explore such dangers as drop-outs, failures to attend at scheduled times, failures to measure specified body dimensions, failures to adhere to the prescribed treatment; and in order, also, to estimate intersubject variation in all features to be studied, and hence the sizes of sample necessary if a large study were to be undertaken.

Norms of Bone Density. It is becoming rather widely recognized that many of the anatomical, physiological and biochemical "normal" values applied in clinical medicine rest upon very shaky foundations; but this is hardly an excuse for proposing that any of the estimates of intersubject variation in bone density, so far published, be used as standards in diagnosis or nutritional assessment, even in the centers where they have originated. The reasons for this caution include:

1. Uncertainty of criteria to be applied in the selection of subjects as a "standard" population.
2. Difficulty in obtaining, especially in adults, a sufficient number—let alone random samples—in each of the populations (race-sex-age groups) which would subsequently furnish the subjects to whom the norms would be applied in clinical assessment.
3. Uncertainty regarding the effects of incompletely corrected soft tissue differences.
4. Ignorance of the effect of changes, from time to time, in radiographic and densitometric techniques.

It is, however, desirable to make some estimates of the orders of magnitude of intersubject variation that would be found if conditions remained the same as in a particular survey. Table 6 shows such estimates as standard deviations in logarithmic form and also as percentages of the group geometric mean densities which are analogous to coefficients of variation. In adults[29, 30] the coefficients of variation for densities of the calcaneus, wrist and hand bones ranged from 5.3 per cent to 7.5 per cent, except for the middle phalanx of the little finger (11.7 per cent). These are considerably lower values than were found in the infants in the study reported here.

Standard deviations obtained from apparently healthy (asymptomatic) subjects can be used to give rough estimates of the limits of "normality." For example, the mean ± 1.28SD cuts off the upper and lower 10 per cent of measurements in a Gaussian frequency distribution. In actually establishing a

standard, however, the Gaussian assumption should be questioned. Although the frequency distributions representing residual intersubject variation in infants' bone densities (p. 44) were sufficiently like the Gaussian distribution to justify the comparison of group means by the *t* test, it must not be lightly assumed that this, or any other frequency distribution of human measurements, is sufficiently Gaussian to render multiples of the standard deviation safe in setting up limits of "normality." The percentile method[25,12] is preferable because it avoids the Gaussian assumption.

If we have decided to accept a group of subjects, chosen by certain criteria, as randomly representative of a standard population, and if we have chosen the 10th and 90th percentiles (the lower and upper 10th percentiles) as the limits of "normal" bone density, we may, by continually using these density values, exclude more (or fewer) than 10 per cent of healthy subjects at each end of the distribution. Our sample of standard subjects must, therefore, be large enough to reduce this risk of misclassification to an amount that we are willing to tolerate; and Herrera,[12] discussing this question, presents tables from which the following information is extracted. If we use the 10th percentile, estimated from a random sample of 100 subjects, we can have considerable confidence (about 95 per cent probabiltiy) that we shall not cut off fewer than 5 per cent or more than 16.5 per cent of the standard population. By taking a sample of 200, we can reduce these figures to 6.3 per cent and 14.5 per cent. A sample of 50 is comparatively unreliable, because the 10th percentile, estimated from it, may cut off anything from 3.4 per cent to 19.6 per cent of its parent population.

In an effort to combat variation of technique in establishing norms, something might be done toward detecting this variation by exposing, throughout the survey of the selected group, a set of bones, imbedded in wax, that were comparable to the living bones under examination. The same bones could then be exposed in testing subjects against the norms.

Bone Density and Mineral Content. The only way to discover exactly what X-ray bone density tells us about mineral content is by experiment—densitometry of living limbs followed by chemical analysis of the bones—as did Schraer and his collaborators[32] with rats' thighs. Even then, we must avoid being deceived by low *average* differences between estimates obtained by the two methods, and by high correlation coefficients, i.e., by close proximity of dots to the line of regression of mineral content on X-ray density. For example, if the correlation in a series of bones were +0.99 and we estimated the mineral content from the X-ray density, the deviations from regression would amount to only 2 per cent of the total variation in mineral content (expressed as the sum of squares of deviation from the mean mineral content of the series). To show how deceptive this apparently trivial 2 per cent can be, it was applied to some published figures of bone ash weight[21] and it was found that the usual ±2SD allowance for error in estimation would, in some of the bones, amount to ± 10 per cent of their true ash weight.

In the study of bone density of living healthy human subjects the evidence for its relationship to mineral content must be indirect, except for an occasional opportunity to study a limb densitometrically before amputation and by bone analysis afterward, or an opportunity to make a similar study before and after death of a subject. (In the United States the likelihood of ante-and postmortem study of a large series of subjects is not great, owing to restrictions on the extent of autopsy mutilation.)

Indirect evidence, in this connection, may be of various kinds. The least indirect information could emerge from a study of animals of similar limb size to man, such as certain domestic animals before and after slaughtering. Sample sizes should be adequate, and the problem of soft tissue should receive more attention than it apparently has received in such studies[7].

In human beings themselves, indirect evidence comprises

a relationship between an X-ray bone density finding and some phenomenon that is known independently of the densitometry —a relationship that is plausible from our knowledge of physiological or pathological processes. For example, in the middle phalanx of the little finger in adult men aged 20 to 84, there was a significant negative correlation between age and mid-shaft bone density, even after elimination of soft tissue differences.[30] In the twelve oldest men (aged 66 to 84), without the observer knowing the age or density score, an "index of articular degeneration" at the two interphalangeal joints of the little finger was derived by scoring the flattening of the articular surfaces, diminution of joint spaces and angularity (or lipping) at the articular margins. Even after elimination of age, there was a correlation of –0.554 between bone density and the index of articular degeneration ($P = 0.05$–0.1), suggesting that the bone density difference reflected an osteo-articular degeneration.

Another possible piece of indirect evidence appeared in the same study[30] of five bones of wrist and hand. In each bone the mean difference in density (right minus left) was positive (P less than 0.001 in four of the bones), even after correction for soft tissue density. The differences could not be accounted for by the bone size differences as measured on the films, but even if they could have been so explained, the finding would have suggested that densitometry was sensitive enough to detect a difference in the amount of mineral matter present. (In the calcaneus,[29] the mean right-left density difference was also significant at the 5 per cent level, but negative.)

In such examples, of course, there still remains the question: Were the corrections, especially for soft tissue, quite adequate? Thus, it might be thought that the negative correlation between age and bone density that has been found in several studies[13] would be a good piece of indirect evidence because it agrees with our experience with the bones of older persons and with some actual determinations of bone mineral content. However, apart from the dangers in translating an age *differ-*

ence, found in a cross-sectional study, into an age *change,* the question arises: Have age differences in soft tissue been sufficiently eliminated?

In this connection it is perhaps pertinent to recall that in the study of the calcaneus[29] and of five wrist and hand bones[30] —a phalanx, a metacarpal, two carpals and the radius—the densities of all six bones, before correction for soft tissue, were negatively correlated with age, but after the soft tissue correction (by the regression method) three of the age-density coefficients became positive, and all were low in value and far from significant at the 5 per cent level, except in the middle phalanx of the little finger (–0.431; $P = 0.001$–0.01). There are various possible explanations of the difference between the phalanx and the other bones, such as an actually greater loss of mineral matter in the phalanx with age; but perhaps the most important difference is the relatively small bulk of overlying soft tissue.

In the present study of infants' films the peculiar features of the intervisit data may be very plausibly interpreted as evidence that shadow changes had a real relationship to changes in mineral content. They cannot, however, reveal how close the relationship was, i.e., the sensitivity of the method, even in the epiphysis of the femur, where the relationship appeared most clearly. The reason is that, in the attempt to relate change in bone shadow density to change in mineral content, the latter could be represented only by change in bone size, and if we could measure the mineral matter more directly we might well find a closer relationship to the X-ray density determinations.

This lack of knowledge of the degree of sensitivity is, however, no barrier to the useful application of densitometry. If no intergroup difference (e.g., in a nutrition trial) is found, it may be due to lack of sensitivity of the method; but if a significant difference is discovered, that in itself is important. The interpretation, and an estimate of the degree of the effect, must depend on other evidence, often including animal experiments.

Complex versus Simple Equipment. The complex equipment developed at Pennsylvania State College sprang from a belief that a continuous density reading, through the whole length or width of a bone would give more accurate information about mineral content than does spot photometry—the reading of density at one point or at a few separate points. That is, of course, not necessarily so; and even if it were desirable to make many readings on each shadow, their number and location should be determined by methods analogous to those used for sampling areas of land, communities and other heterogeneous aggregates.

When readings are made along a tracing path extending from one end of a bone to another, the intersubject variation in shape at the ends of long bones and at the whole periphery of short bones (e.g., the calcaneus) introduces intersubject variation in bone density estimates, and it is noteworthy that, in some of the most recent work at Pennsylvania State University on the relationship between shadow density and mineral content[32], the ends of long bones were excluded. Indeed, even if all points on bone shadows could be read for density, the information would not be very useful unless the shape of each bone were equally well determined.

The claim that the continuous-reading method is greatly superior to spot photometry does not appear to have been established by systematic and comprehensive comparison of the two methods. In such a comparison, techniques appropriate to the study of variation should be used. For example, the "reproducibility" of the continuous-reading method may appear high when the two ends of the tracing path are marked on a film and the same marks are used as guides in each replicate reading. In order to estimate the contribution of reading variation to the variation between two films of the same bone in the same subject, or between films of the same bone in different subjects, the tracing path must, of course, be chosen afresh each time that any film is read. That is the reason, in this laboratory, for the marking of reading points on tracing paper

attached to, and detachable from, the films.

If the exploration of variation is properly designed, with appropriate randomizations, the data can be analyzed by the "components of variance" method, which was used in this laboratory to separate three components: between subjects, between films from the same subject, between readings on the same film.[29, 30] Applied to the comparison of mineral content and X-ray density, it could separate such components as: between chemical analyses of different regions of the same bone, between analyses of the same region performed on different occasions.

Such explorations might prove that spot photometry gives much less reliable information about mineral content than does the continuous-reading method; but it is desirable at this point to ask: What information do we desire?

In densitometry, a particular bone is taken, more or less tacitly, as a sample of the osseous tissue of the body, or at least of bones of its own type; but any subjects' bones, even of similar type and structure, doubtless differ in their mineral content per unit volume or unit weight. Therefore, if it were desirable to obtain the full information about any one bone, it would surely be desirable to do the same for another bone, and another, and another—at least until it was shown how closely two or three bones represented the whole skeleton or the total bones of their class. This is, however, not the way in which histological and physiological sampling is done. We take a few drops of blood from a finger tip—not the whole blood of the finger, or even all its capillary blood. In clinical measurement of blood pressure we do not find the average of readings with the cuff at different positions on the arm or on different limbs. There is room for improvement in such sampling methods, but not by attempting to collect all possible information on the thing to be measured. Improvement comes by adopting, and adhering to, certain rules of procedure which will reduce the observational variation, and risk of bias, as far as is necessary in the application of the method.

Previous studies in this laboratory (p. 16) have shown that when spot photometry is carefully applied its reading variation makes only a small contribution to the variation between films of the same bone that are processed at the same time. Previous studies, as well as the present study, have shown that, even when intersubject variation is added, the sample sizes necessary to demonstrate intergroup differences in bone shadow density (or changes within a group) are not exorbitant. The question still remains: How large must a difference in mineral content be before it will produce a demonstrable difference in bone shadow density? If soft tissue masks differences in mineral content, it is immaterial whether photometer readings are made at one point or at all points on the bone shadows. Until the soft tissue question is settled, therefore, we may well doubt the wisdom of further propagating complex and costly equipment.

SUMMARY

Bone shadow densities were determined by spot photometry on the shaft of metatarsal I, femur shaft and femur distal epiphysis in the right limbs of infants, born in 1951 or 1952, whose mothers, from the ward-service hospital population of Philadelphia, had been the subject of a prenatal dietary supplement experiment, in which the women, during the first 16 weeks of pregnancy, had been assigned by a random procedure to one or other of four groups: polyvitamin concentrate; protein concentrate (with calcium); polyvitamin concentrate plus protein concentrate; control (with no supplement supplied). The films examined were obtained from 286 infants (189 white and 97 Negro) in the first postnatal week, 151 of the same infants at a median age of approximately one month, and 143 of the same infants at a median age of approximately six months.

Corrections for interfilm differences in general density and in soft tissue density were made by a "direct" method, i.e., readings were made at a particular point on the shadow of an aluminum-alloy wedge that had been exposed on all films, and on areas of soft tissue adjacent to the respective bones, and these two readings formed the independent variables in a multiple

linear regression equation in which the dependent variable was bone density. The equation provided an estimate of what the shadow density in a particular bone would have been if the intersubject differences in bone density were entirely due to differences in general film density and in soft tissue density, as represented by the equation. The difference between this estimated density and the observed density (the deviation from regression) was an index of the density status of that bone in relation to the mean density for the same bone in the whole group of infants. The means of these deviations for each of the three bones in each of the 16 race-sex-treatment subgroups were used for comparison of treatments.

There was found no evidence of a relationship between the mother's prenatal dietary supplement and the infant's bone density in the first postnatal week or its change in density between the first and fourth week.

The only demonstrable racial or sex difference was found in the metatarsal. In both sexes the mean density in white infants during the first week exceeded the mean density in Negro infants of the same sex by amounts that could not be readily attributed to random sampling variation, nor could they be accounted for by differences in metatarsal size (as measured by width).

The investigation was therefore concerned chiefly with:

1. The precision of the method for the detection of differences in bone shadow density. For example, it was shown that, after correction for differences in bone size, wedge density and soft tissue density, the residual intersubject variation was low enough to justify a verdict of "significant at the 5 per cent level" if, in two samples of 100 infants, the greater mean bone shadow density exceeded the smaller by 5 per cent.

2. Evidence of the sensitivity of the method in reflecting differences in actual mineral content. Because it is rarely possible in living human subjects to obtain direct evidence of this sensitivity, indirect evidence must be sought—for example, the demonstration of density differences that cannot be easily accounted for by differences in general film density, in soft tissue

density, in bone size, or in shadow magnification, or by random sampling variation. For instance, in the intervisit intervals, although the femur epiphysis clearly showed increase of density, in the metatarsal and femur shaft the differences were either effectively zero or negative, and the most plausible explanation was that the densitometer had registered the effect of the destruction (remodelling) inside those bones, which at this stage of development competes with the subperiosteal deposition of bone.

In the light of this study and of previous studies of adult bones and cadaver bones by the same investigators, a detailed discussion of current problems in bone densitometry is presented.

REFERENCES

[1] Berkson, J.: Limitations of the Application of Fourfold Table Analysis to Hospital Data. *Biometrics Bulletin,* (1946), 2: 47–53.

[2] ————: The Statistical Study of Association Between Smoking and Lung Cancer. *Proceedings of the Staff Meetings of the Mayo Clinic* (1955), 30: 319–348.

[3] Brown, W. N.: INSTRUCTIONS FOR EXPOSING AND DEVELOPING X-RAY FILMS FOR BONE DENSITY EVALUATION. Pennsylvania State College, Bone Density Research and Evaluation Center, 1953.

[4] Bywaters, E. G. L.: The Measurement of Bone Opacity. *Clinical Science,* (1948), 6: 281–287.

[5] Clark, A. K.: AN ANNOTATED BIBLIOGRAPHY ON BONE DENSITOMETRY PREPARED FOR THE WORKSHOP ON BONE DENSITOMETRY. Yellow Springs, Ohio, The Fels Research Institute for the Study of Human Development, 1959.

[6] Engstrom, A., and Welin, S.: A Method for the Quantitative Roentgenological Determination of the Amount of Calcium Salts in Bone Tissue. Its Use in the Study of Calcium Retention in the Phalanges of the Fingers in Chronic Polyarthritis. *Acta radiologica,* (1949), 31: 483–502.

[7] Facto, L. A., Kastelic, J., and Catron, D. V.: Estimating Ash Content of Bone, Using Standardized Radiographs. *Journal of the Biological Photographic Association,* (1956), 24: 13–21.

[8] Fisher, R. A.: STATISTICAL METHODS FOR RESEARCH WORKERS. New York, Hafner Publishing Company, 1954.

[9] Fisher, R. A., and Yates, F.: Statistical Tables for Biological, Agricultural and Medical Research. London and Edinburgh, Oliver and Boyd Ltd., 1943.

[10] Garn, S. M.: An Annotated Bibliography on Bone Densitometry. *American Journal of Clinical Nutrition,* (1962), 10: 59–67.

[11] Garn, S. M., Harrison, H. E., and Mason, K. E. (Eds.): Transcript of the Workshop on Bone Densitometry Held at The Stone House, National Institutes of Health, December 4, 5, 1959. Yellow Springs, Ohio, The Fels Research Institute for the Study of Human Development, 1960.

[12] Herrera, L.: The Precision of Percentiles in Establishing Normal Limits in Medicine. *Journal of Laboratory and Clinical Medicine,* (1958), 52: 34–42.

[13] Heuck, F., and Schmidt, E.: Die praktische Anwendung einer Methode zur Quantitativen Bestimmung des Kalksalzgehaltes gesunder und kranker Knochen. *Fortschritte auf dem Gebiete der Roentgenstrahlen und der Nuklearmedizin,* (1960), 93: 761–783.

[14] Hill, A. H.: Fetal Age Assessment by Centers of Ossification. *American Journal of Physical Anthropology,* (1939), 24: 251–272.

[15] Jackson, H.: Problems in the Measurement of Bone Density. *British Journal of Radiology,* (1951), 24: 613–616.

[16] Kasius, R. V., Randall, A., Tompkins, W. T., and Wiehl, D. G.: Maternal and Newborn Nutrition Studies at Philadelphia Lying-in Hospital. Newborn Studies. I. Size and Growth of Babies of Mothers Receiving Nutrient Supplements. *In* The Promotion of Maternal and Newborn Health. [Proceedings of a Round Table at the 1954 Annual Conference of the Milbank Memorial Fund.] New York, 1955, pp. 153–168.

[17] ———————————————: Maternal and Newborn Nutrition Studies at Philadelphia Lying-in Hospital. Newborn Studies. III. Size and Growth of Babies During the First Three Months of Life. *The Milbank Memorial Fund Quarterly,* (1955), 33: 341–367.

[18] ———————————————: Maternal and Newborn Nutrition Studies at Philadelphia Lying-in Hospital, Newborn Studies. V. Size and Growth of Babies During the First Year of Life. *The Milbank Memorial Fund Quarterly,* (1957), 35: 323–372.

[19] ———————————————: Maternal and Newborn Nutrition Studies at Philadelphia Lying-in Hospital. Newborn Studies. VI. Infant Size at Birth and Parity, Length of Gestation, Maternal Age, Height, and Weight Status. *The Milbank Memorial Fund Quarterly,* (1958), 36: 335–362.

[20] Lachman, E., and Whelan, M.: The Roentgen Diagnosis of Osteoporosis and Its Limitations. *Radiology,* (1936), 26 : 165–177.

[21] Mack, P. B.: Results from the Study of Bone Density in the Appraisal of Calcium Status. *In* NUTRITION IN RELATION TO HEALTH AND DISEASE. Papers Presented at a Round Table at the 1949 Annual Conference of the Milbank Memorial Fund, New York, 1950, pp. 30–63.

[22] Mack, P. B., Brown, W. N., Jr., and Trapp, H. D.: The Quantitative Evaluation of Bone Density. *American Journal of Roentgenology,* (1949), 61 : 808–825.

[23] Mack, P. B., O'Brien, A. T., Smith, J. M., and Bauman, A. W.: A Method for Estimating the Degree of Mineralization of Bones from Tracings of Roentgenograms. *Science,* (1939), 89: 467.

[24] Mainland, D.: An Experimental Statistician Looks at Anthropometry. *Annals of the New York Academy of Science,* (1955), 63: 474–483.

[25] ———: ELEMENTARY MEDICAL STATISTICS, Philadelphia, W. B. Saunders Company, 2nd edition, 1963.

[26] ———: Measurement of Bone Density. *Annals of the Rheumatic Diseases,* (1956), 15: 115–118.

[27] ———: The Risk of Fallacious Conclusions from Autopsy Data on the Incidence of Disease, with Applications to Heart Disease. *American Heart Journal,* (1953), 45: 644–654.

[28] ———: Safety in Numbers. *Circulation,* (1957), 16: 784–790.

[29] ———: A Study of Age Differences in the X-ray Density of the Adult Human Calcaneus—Variation and Sources of Bias. *Journal of Gerontology,* (1957), 12: 53–61.

[30] ———: A Study of Age Differences in the X-ray Density of Five Bones in the Adult Human Wrist and Hand. *Journal of Gerontology,* (1957), 12: 284–291.

[31] The Rand Corporation: A MILLION RANDOM DIGITS WITH 100,000 NORMAL DEVIATES. Glencoe, Illinois, The Free Press, 1955.

[32] Schraer, H., Schraer, R., Trostle, H. G., and D'Alfonso, A.: The Validity of Measuring Bone Density from Roentgenograms by Means of a Bone Density Computing Apparatus. *Archives of Biochemistry and Biophysics,* (1959), 83: 486–500.

[33] Snedecor, G. W.: STATISTICAL METHODS APPLIED TO EXPERIMENTS IN AGRICULTURE AND BIOLOGY. Ames, Iowa, The Iowa State College Press, 1956.

[34] Tompkins, W. T., and Wiehl, D. G.: Epiphyseal Maturation in the Newborn as Related to Maternal Nutritional Status. *American Journal of Obstetrics and Gynecology,* (1954), 68: 1366–1377.