Selection of Variables, Timing of Examinations, and Retention

Alex F. Roche and Shumei Guo

Department of Community Health, Wright State University, Yellow Springs, Ohio, USA

In any clinical trial, the selection of variables to be measured, the timing of examinations, and the retention of subjects form a triad of critical elements that in large part determine whether suitable data will be available for hypothesis testing at the end of the trial. These elements, which are particularly important in clinical trials of infants, must receive careful consideration during the development of the protocol. Because clinical trials involving infant nutrition are brief, there are only limited opportunities to modify these trials once they have begun.

The examples used in the present chapter relate to common measures of infant growth, but they are applicable to other measurements that might be made. Consideration of these examples can lead to the elaboration of some general principles and attitudes that are pertinent to the biochemical measures and to data relating to immunocompetence and health that are likely to be recorded and analyzed in clinical trials of infant nutrition.

It is assumed that a control group will be included in the trial. Usually, the control group and the experimental group will be constructed by random assignment, and the control group will receive a well-established infant formula. A comparison group of breast-fed infants can assist the interpretation of the findings, although it is difficult or impossible to standardize the intake and duration of breast-feeding, and there will not be random assignment involving the breast-fed group. In addition, comparisons are usually made with commonly used reference data to detect whether the growth status of some of the infants in the experimental group is outside the normal range. These latter comparisons should take covariates into account, but there are some limitations to the implementation of such adjustments. The current U.S. National Center for Health Statistics (NCHS) reference data for infancy were obtained from the Fels Longitudinal Study (1–3), and the data for matching covariates are incomplete and not readily available. It is expected that revised NCHS reference data will be published in 1997 and that data for covariates will be made available at that time for the infants in national surveys. These revised data will be based on national
samples at birth and at ages older than 3 months. From birth to 3 months, however, they will be based on other data sets for white infants born at term (3–5). Adjustments for covariates will be difficult or impossible during these young ages.

SELECTION OF VARIABLES

In accordance with common practice, the measurements of infant growth are likely to be weight, length, and head circumference. These are good choices. Each of these measurements increases rapidly during infancy, particularly soon after birth. The more rapid the rate of growth, the more likely it will be modified by the infant formula being tested.

During the period from 1 to 12 months, the relative increases in the median values are about 145% for weight, 40% for length, and 26% for head circumference. Part of the much more rapid increase in weight is related to its three-dimensional nature. In the Fels Longitudinal Study, the coefficients of variation of increments, expressed as percentages of the means, are larger for weight (40%, 25%, and 16% for 1- to 2-, 2- to 3-, and 3- to 4-month increments, respectively) than for length and head circumference (about 20%, 15%, and 10% ) for the same age intervals (6,7). For the increments from 10 to 11 and from 11 to 12 months, the coefficients of variation are 20% for weight, 8% for length, and 16% for head circumference. These differences between measurements reflect the fact that changes in weight tend to be much more rapid than those in length or head circumference. Therefore, significant effects of a formula being tested are more likely to be detected in weight than in length or head circumference.

Weight, which is the sum of the weights of all body components, is a less specific measure than length or head circumference. In the presence of disease, there may be extremely rapid changes in weight during a period of a few days, which usually reflect alterations in body fluids and are quickly reversed if the disease is treated effectively. Changes in adipose tissue stores, which become abundant at about 12 months, may be an important element in less rapid weight changes (8,9). If an infant has an unusual weight at birth, the centile levels for weight during infancy commonly change, so that they become closer to the medians. These changes, which are rapid from birth to 9 months, continue to 24 months (10) and may be large enough to be categorized as decanalization. They will not, however, meet the criteria for failure-to-thrive, which are basically a low centile level combined with a decrease in the centile level for weight.

Conceptually, the length of an infant is the sum of the lengths of many bones, and it reflects the past elongation of these bones during a period of months. Length cannot decrease. Therefore, any negative increments for length that are present in the recorded data must result from measurement errors. There may be changes in the relative levels for length during infancy that are not dependent on the type of feeding. For example, when the midparent stature, which is the average of the statures of the two parents, is large and the infant is short at birth, the centile levels for the
length of the infant increase until about 12 months (11). Changes in the relative levels of length that are opposite in direction occur in infants who are long at birth and whose parents are short.

It is useful to interpret weight and length jointly. At older ages during childhood, weight and stature are frequently combined as the body mass index [BMI, weight (kg)/stature (m²)]. The use of BMI, with the substitution of length for stature, is not recommended during infancy because the accurate measurement of length requires very careful attention to detail, and the effects of any measurement errors are increased when the square of length is calculated. During infancy, weight-for-length is the preferred combination of weight and length. Weight-for-length is described as age-independent throughout infancy and prepubescence, because for all practical purposes, the centile levels in one age group match the corresponding centile levels in the next older age group for matching lengths (12). As a result, weight-for-length centiles can be displayed conveniently in growth charts. This age independence also has advantages when the age of an infant is unknown, as may occur in developing countries. Some recent data indicate, however, that weight-for-length may increase with age from birth to 3 months (13).

When the centile levels of weight, length, and weight-for-length are considered for an infant, the size of the infant for each variable will be compared with reference data such as those of Hamill et al. (1,2). Some infants will be judged, for example, to be underweight and short but within the normal range of weight-for-length. The prevalence of these and other categories of infants can be compared between the experimental and the control groups.

The importance of head circumference derives from its close relationship to brain weight during infancy, when both the scalp and the cranium are thin (14,15). Small head circumferences are associated with reduced mental ability (16), and large head circumferences lead to concerns about possible hydrocephalus, particularly if the increments in head circumference are markedly greater than the median increments (17). When an unusual head circumference is noted in an infant, or there are rapid increases in head circumference, the cause should be sought. In some cases, an unusual head circumference is a familial trait that does not have functional significance.

Many other anthropometric variables could be recorded, but the measurement of these is recommended only for special circumstances. For example, it may be impossible to measure length and difficult to judge weight in an amputee. Alternative measures such as arm span, skinfold thicknesses, and circumferences can be considered. Skinfold thicknesses could be measured in all the trial infants at the triceps and subscapular sites. The measurement errors, relative to the means, would be particularly large from birth to 2 months because the subcutaneous adipose tissue layer is thin. Later in infancy, the relative measurement errors for these skinfold thicknesses may still be unacceptably large despite a thicker layer of adipose tissue. Midarm circumference can be measured with small errors, but this provides limited useful information if it is not combined with triceps skinfold thickness to calculate the cross-sectional area of adipose tissue and the combined cross-sectional area of
muscle and bone at the same level. These estimates require that the triceps skinfold thickness be squared, which increases the effects of any measurement errors.

THE TIMING OF EXAMINATIONS

Because the word growth implies change, analyses of growth require serial data. If examinations are scheduled only at the beginning and the end of a clinical trial, the recognition of adverse effects would be delayed until the end of the trial, and brief effects may not be recognized. With such a protocol, few increments could be calculated for intervals matching those for which there are reference data. The recent North American sets of reference data for growth increments during infancy that are in common use are listed in Table 1. The examinations should be timed so that increments can be calculated for intervals matching those of the reference data. Although increments are important, not all analyses should be based on them.

It is necessary to compare serial recorded measurements with reference data for status to recognize decanalization (21) and to use reference data for both status and for increments to recognize failure-to-thrive (22.23). Decanalization, which is relatively common during infancy and pubescence, occurs when the serial plotted data for an individual cross two or more major centile lines on growth charts for status. This crossing may demonstrate an increase or a decrease in relative level. Decanalization can occur in length and head circumference but is most common in weight. Failure-to-thrive refers to a weight during infancy that is less than the median at one age and is followed by an increment in weight that is less than the fifth centile.

The timing of examinations is partly determined by the rates and patterns of growth during infancy. Growth is rapid in weight, length, and head circumference during early infancy, but these rates soon decelerate markedly. Consequently, the centiles of 1-month increments decrease rapidly from birth to 6 months and then

<table>
<thead>
<tr>
<th>Authors</th>
<th>Variables and intervals</th>
<th>Age range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guo et al. (5)a</td>
<td>W; 1-mo</td>
<td>Birth–6 mo</td>
</tr>
<tr>
<td></td>
<td>W; 2-mo</td>
<td>Birth–12 mo</td>
</tr>
<tr>
<td></td>
<td>L; 2-mo</td>
<td>Birth–6 mo</td>
</tr>
<tr>
<td></td>
<td>W and L; 3-mo</td>
<td>Birth–24 mo</td>
</tr>
<tr>
<td>Guo et al. (6)</td>
<td>HC: 1-mo</td>
<td>Birth–12 mo</td>
</tr>
<tr>
<td>Roche et al. (7)</td>
<td>W and L; 1-mo</td>
<td>Birth–12 mo</td>
</tr>
<tr>
<td>Roche and Himes (18)</td>
<td>W, L, HC; 6-mo</td>
<td>Birth–36 mo</td>
</tr>
<tr>
<td>Baumgartner et al. (19)</td>
<td>W, L, HC; 6-mo</td>
<td>Birth–36 mo</td>
</tr>
<tr>
<td>Roche et al. (20)</td>
<td>W, L, HC; 1-mo</td>
<td>Birth–12 mo</td>
</tr>
</tbody>
</table>

\(^{a}\) Also given in Fomon (4).
\(^{b}\) W, weight; L, length; HC, head circumference; mo, month.
FIG. 1. Selected percentiles for 1-month increments in weight for girls. (From Roche et al., ref. 7, with permission.)

decrease slowly throughout the remainder of infancy, as shown in Figs. 1 to 3. These figures, in which the increments are plotted opposite the ends of the intervals, display only the fifth, 50th, and 95th centiles because the expected measurement errors would make it difficult or impossible to distinguish among a more complete set of centile levels for an individual infant. The fifth centiles of these increments for girls are given in Table 2; these are slightly smaller than the corresponding values for boys, and, in each sex, the fifth centiles are almost the same as the tenth centiles.

It is recommended that the time intervals between examinations be long enough that the fifth centiles of the increments are larger than the technical errors (TE):

\[
TE = \sqrt{\sum_{i=1}^{l} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 / \sum_{i=1}^{l} n_i},
\]

where \(x_{ij}\) is the \(j\)th measurement of the \(i\)th participant for \(i = 1, 2, \ldots, l\) and \(j = 1, 2, \ldots, n_i\). The TE for weight in infancy is about 10 g (S. Guo and A.F. Roche, unpublished data), which could lead to a suggestion that weight be measured daily to 6 months and then each second or third day for the remainder of infancy. There is, however, a biological variation in weight of about 100 to 250 g that is associated
with feeding, defecation, and urination. In the absence of special arrangements to control this variation, intervals of 1 month between weight measurements are appropriate. The TEs for length during infancy are 0.3 cm, and for head circumference 0.2 cm (S. Guo and A.F. Roche, unpublished data). Even if the TEs for length are doubled because of incomplete quality control, the TE will still be less than the fifth centiles of the 1-month increments (Table 2). After 10 months, the TEs for head circumference that are attained with very good quality control are only slightly smaller than the fifth centiles for 1-month increments. Consequently, 1-month intervals between measurements of head circumference after 10 months are justified only if the quality control within the study is very good. In theory, if the TEs for length and head circumference were reduced to values smaller than those found by Guo and Roche, the measurements of these variables could be made at intervals shorter than 1 month. It would be difficult to achieve such reductions in the TEs, and more frequent examinations would be likely to place undue burdens on the mothers. Furthermore, there are no current reference data for intervals shorter than 1 month.

The calculation of increments can assist the statistical analyses and, in addition, can assist the identification of infants whose growth is unusual. One-month increments allow earlier recognition of serious decreases in growth rates than is possible from consideration of serial values for status. This is illustrated in Fig. 4, which displays theoretical serial weight data for two girls. The weight of each girl was
FIG. 3. Selected percentiles for 1-month increments in head circumference for girls. (From Guo et al., ref. 6, with permission.)

<table>
<thead>
<tr>
<th>Age (mo)</th>
<th>Weight (kg/mo)</th>
<th>Length (cm/mo)</th>
<th>Head circumference (cm/mo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–2</td>
<td>0.54</td>
<td>2.21</td>
<td>1.9</td>
</tr>
<tr>
<td>2–3</td>
<td>0.50</td>
<td>2.10</td>
<td>1.2</td>
</tr>
<tr>
<td>3–4</td>
<td>0.44</td>
<td>1.92</td>
<td>0.9</td>
</tr>
<tr>
<td>4–5</td>
<td>0.38</td>
<td>1.72</td>
<td>0.7</td>
</tr>
<tr>
<td>5–6</td>
<td>0.35</td>
<td>1.58</td>
<td>0.6</td>
</tr>
<tr>
<td>6–7</td>
<td>0.31</td>
<td>1.45</td>
<td>0.5</td>
</tr>
<tr>
<td>7–8</td>
<td>0.29</td>
<td>1.35</td>
<td>0.5</td>
</tr>
<tr>
<td>8–9</td>
<td>0.26</td>
<td>1.26</td>
<td>0.4</td>
</tr>
<tr>
<td>9–10</td>
<td>0.24</td>
<td>1.19</td>
<td>0.4</td>
</tr>
<tr>
<td>10–11</td>
<td>0.23</td>
<td>1.12</td>
<td>0.3</td>
</tr>
<tr>
<td>11–12</td>
<td>0.21</td>
<td>1.07</td>
<td>0.3</td>
</tr>
</tbody>
</table>

*Data from Guo et al. (6) and from Roche et al. (7).*
FIG. 4. Alternative plots of theoretical serial weights at 1-month intervals for girls A and B. In the left panel, serial status values are plotted against the NCHS reference data of Hamill et al. (1,2); in the right panel, 1-month increments are plotted against incremental reference data. (From Roche et al., ref. 7, with permission.)

equal to the median at 1 month. Subsequently, the 1-month increments for girl A were slightly greater than the 95th centile, and those for girl B were slightly less than the fifth centile, as shown in the right panel of the figure. The serial status data for the two girls are plotted against the National Center for Health Statistics reference data of Hamill et al. (1,2) in the left panel. The status values for girl A reach the 95th centile at 2 months, when her rapid weight gain can be recognized. The status centiles for girl B decrease slowly with age; her unusually slow growth may not be recognized until 5 months. The 1-month increments in the right panel allow the unusual rates of growth for both girls to be recognized at age 2 months.

It is important that the TEs be as small as possible to increase the likelihood of significant findings from the statistical analyses. The TEs can be reduced by close attention to quality control with an emphasis on (a) the adoption of a standard protocol, particularly if there are multiple sites, (b) recruiting measurers who are interested in this task and anxious to do it as well as possible, (c) the selection of appropriate anthropometric instruments and the items needed for their calibration, (d) central training of physicians and measurers in measurement, calibration, and
recording, (e) repeated training and the collection of reliability data throughout the trial, and (f) review by the pediatrician of the data recorded at each examination and discussion of these data with the measurer and the mother before the mother and infant leave the clinic. The procedures for the specific measurements should match those used by NCHS, which are in close agreement with the recommendations of a Consensus Conference (24,25). Lists of suppliers of anthropometric instruments are given by Lohman et al. (24) and Roche et al. (26).

The timing of examinations is set in the protocol as a series of target (scheduled) ages. In practice, there will, however, be some differences between these target ages and the actual ages at examinations. Although such differences are inevitable, they should be kept small by careful scheduling of appointments that takes into account the convenience of the mother and by rescheduling infants as soon as possible after missed appointments. No matter how large the differences in timing between the target ages and the actual ages at examinations, none of the recorded data should be excluded from consideration during the statistical analyses. The effects of the inevitable variations from the target ages can be overcome in some types of analyses by the use of exact age-specific centile levels that can be obtained using Epi Info software (27), although centiles are not metric. Analyses of ‘‘1-month increments’’ using the recorded data are likely to be misleading if the lengths of the intervals between measurements differ markedly from 1 month, particularly if these differences vary between groups. In such circumstances, the recorded data must be adjusted to the target ages either by fitting a function to the serial data for each infant or by regression of the experimental and control data sets on target ages. After the data have been adjusted to the target ages, they can be used to calculate values for fixed intervals.

RETENTION

Retention, within the context of clinical trials of infant nutrition, relates much more to interactions with the mother than to interactions with the infant, unless the trial is conducted on institutionalized infants. Retention in the trial is usually more difficult for teenage mothers and those who are single or have a language barrier. These mothers should not be excluded, but individual attention should be given to them and to other mothers with special needs, to increase the retention rate. The fathers should be involved in the trial. This can be done through a newsletter and by evening functions at the beginning and about 2 months after the start of the trial, to which both parents and all the siblings are invited. These functions should be pleasant and informative for the whole family and include the presentation of small gifts.

The convenience of travel to the clinic and of the appointments that are made is closely related to retention. Some mothers may desire appointments on Saturdays or evenings. Efforts should be made to meet these desires early in the trial; such efforts should not be delayed until there is a crisis. Preventive measures are always
better than crisis management. The number of examinations will be set by the protocol, but this number should not exceed what is really necessary. The provision of transportation increases the retention rate.

When the mother arrives with her infant for an examination, she should be greeted warmly and receive immediate attention. Ideally, the same few members of the clinic staff will manage all aspects of these examinations so that cordial relationships will develop between these staff members and each mother. Siblings should be made welcome at the clinic, and facilities for their play should be provided.

The mothers must be fully instructed about the trial so that they can give informed consent. In addition, mothers who fully understand the trial, including its aims and the potential benefits to society and to their infant, are more likely to be interested and compliant participants. This requires individual verbal explanations supplemented by a booklet that describes the trial and the changes that are usually noted in the variables being measured. This booklet should be brought by the mother to each examination, so that the pediatrician can make notes of the findings for her infant in spaces that the booklet provides. One excellent way to encourage continuing membership in the trial is to give the mother instant photographs of the infant that she can place in her booklet. At each examination, a stipend and travel expenses should be paid to the mother, and the infant should receive a toy. The distribution of bibs and carrying bags for diapers can help. Usually, the mother is given sufficient free formula to feed the infant until the next examination. This formula may weigh as much as 5 kg. If the mother uses public transport, special arrangements may be needed to move this formula to her home.

The previous paragraphs describe actions that relate specifically to the mother and father. The infant is important also. Invasive procedures should be kept to a minimum and should not precede the anthropometry. It is impossible to obtain accurate body measurements of an infant who is screaming because a blood sample has just been obtained or a vaccination performed. Similarly, infants who are upset because of hunger, thirst, or an uncomfortable diaper require attention to the cause of their distress before they are measured.

Complete retention is unlikely, despite all the efforts that may be made. Some guide to what can be achieved is provided by the data in Table 3. The data from the Fels Longitudinal Study are for infants born from 1985 to 1992. For the sake

### TABLE 3. The percentage retention in some studies of infants

<table>
<thead>
<tr>
<th>Study</th>
<th>Number enrolled (age)</th>
<th>Percentage retained at various ages (mo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fels Longitudinal Study</td>
<td>47 (1 mo)</td>
<td>94 96 96 94 91</td>
</tr>
<tr>
<td>(births 1985–1992)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>An unpublished clinical trial</td>
<td>352 (1 mo)</td>
<td>89 58 — — — —</td>
</tr>
<tr>
<td>Sempé et al. (28)</td>
<td>542 (birth)</td>
<td>— — 75 65 59</td>
</tr>
</tbody>
</table>


of simplicity, the data presented are for a subset of the examinations scheduled for these Fels participants. Retention was good, partly because of some special circumstances. A parent, and commonly a grandparent, of each infant has been included in the Fels Longitudinal Study since birth, and invasive procedures are not used before 36 months of age. One might expect that these advantages would be somewhat offset by the absence from the study of physical examinations and pediatric advice. The retention rate in one unpublished clinical trial of infant nutrition was not as high as in the Fels Longitudinal Study. The mothers of these infants did not have a preexisting commitment to the trial, and the protocol included some invasive procedures. Also, retention in this clinical trial could have been reduced by treatment failure—for example, inability to digest the formula—or by the application of exclusion criteria relating to changes that occurred after enrollment. Retention in the longitudinal study of Parisian infants, which was reported by Sempé et al. (28), was not as good as in the Fels Longitudinal Study. Invasive procedures were not included, but the mothers did not have a previous commitment to the study.

Attrition should be defined before the beginning of the trial, and its prevalence should be recorded for various categories of apparent causes (maternal refusal, change in place of residence, treatment failure, noncompliance in the feeding of the assigned formula or in the age at introduction of solid foods, noncompliance with the scheduled examinations, and maternal or infant diseases).

SUMMARY

The timing of examinations in a clinical trial of infant nutrition should be based on the rates of change in the variables measured, their technical errors, and the burden on the mothers. Consequently, the selection of the variables to be measured must precede decisions about the timing of examinations. The variables selected should be ones that are likely to show effects of differences in infant feeding on major dimensions of the body (weight, length), or on dimensions that have clear relationships to the size and function of an important organ (head circumference). Wise decisions about the timing of examinations and the choice of measures will be of little avail in the absence of unremitting efforts to maintain a high quality of data collection and to retain the mothers and infants in the trial.

REFERENCES


**DISCUSSION**

*Dr. Uauy:* We can have normal infants who, in fact, are not growing for variable periods of time and then have variable growth spurts. How is this incorporated in individual follow-up?
Dr. Roche: I think the model is true. However, it appears to be true only during infancy—at least there is no proof of it being true at other periods. Points plotted on a growth chart do not show short-term pauses and succeeding spurts, because in order to recognize these little jumps, one has to measure the infant daily or every second day. But the fact that the jumps are happening means that they become a confounding variable—they add a little variance to the data. Thus, if you measure at monthly intervals, you are likely to have three or four of these little jumps and periods of stasis in between, so the 1-month increments that you observe will not be altered, but they will be smoother than the truth.

Dr. Rey: You say that it is difficult to detect the difference between the fifth and the tenth centile, but I never asked myself if there is a difference between the fifth and the tenth centile. Why do you ask this question?

Dr. Roche: I did not ask the question. These are the fifth and tenth centiles of increments we are talking about, and I drew attention to them for two reasons. One is to explain why the published graphs do not have a tenth centile—they only have three lines: the fifth, 50th, and 95th. Other centiles were published in tables. The second reason is that, in analyzing the results from a clinical trial, the FDA expects you to look at the increments in the experimental group and the control group, and, having looked at the increments, you can then make a comparison, a t test if you wish, or use confidence limits, between the increments in one group and the increments in the other; and you can also look at how many children in each of the groups are growing at a slow rate during each of the intervals for which you have data. What I am pointing out is that I would suggest you use the fifth centile, which is more meaningful to regulatory agencies. If you use the tenth centile, you will get almost the same answer. It is difficult to separate the fifth-centile children from the tenth-centile children because the interval is so small. In the revised NCHS status charts, it is still uncertain what the lower centiles will be. There will almost certainly be a third centile, but it is not certain whether there will be a fifth centile. They are likely to be too close together for the difference to be meaningful.

Dr. Walter: You have described the difficulty of estimating single increments in particular children or subjects, and this is really a matter of reliability in the measurement—the measurement variance is of the same order of magnitude as the day-to-day fluctuation and other things that will affect the actual weight or other variable. I wonder whether, instead of using single increments, you could perhaps look back at the history of the subject and use all the accumulated information that you have on that particular person up to that point—that would improve the reliability so that, for instance, you might observe a series of small increments that seem to deviate from the norm, which individually don’t amount to very much but collectively could be very meaningful. I am thinking here of some of the indices referred to a tracking in cohort studies, where you can actually measure the tendency for individuals to stay on the same centile in the distribution over time or to move to a different point in the distribution.

Dr. Roche: These are very good points, and, in fact, we do that. We do fit functions to serial data for individuals and make comparisons between groups on the basis of the parameters of those functions if they fit the data well. We also look at the goodness of fit between the control and experimental groups. This is a similar approach to what you were suggesting.

Dr. Saavedra: Could you comment on the use of standard deviation scores, particularly when it comes to comparing populations across sex or across age, given the fact that the coefficients of variation change for the centiles over a period of time.

Dr. Roche: In the context of infancy, most of these growth variables are normally distributed, even for weight, though that is not the case later. However, in clinical trials, the popula-
tions are usually not general populations of infants—they typically exclude preterm infants, or else they are studies of preterm infants and low birth weight infants only. However, if all infants are included, one has nearly normal distributions. So, in these clinical trials, if we are dealing with a truncated distribution—truncated at 2500 g or 37 weeks of gestation or both—then, it is not appropriate to use standard deviations. A partial solution is to take the distribution and cut it in half, obtaining an SD for the upper half and an SD for the lower half and applying those, depending on whether the value you are describing is in the upper half or in the lower half. Those standard deviations will not be equal, which leads to statistical complications. Standard deviations have an advantage, however, in that, in theory, you can go to the extremes of the distribution. You can talk about children who are 4 or 5 standard deviations below the mean. Some inaccuracy is involved because such children are extremely uncommon in general samples, and most reference data are based on 200 to 500 per age and sex group.

Dr. Yetley: If you had two large enough groups of infants, one formula-fed and one exclusively breast-fed, would you see the same pattern of weight increments, or will they be different?

Dr. Roche: There are not much data that are convincing. The available data indicate that the breast-fed infants grow more slowly in weight for the first 12 months than formula-fed infants. The difference is not apparently as great as it probably was 20 to 30 years ago, when mothers using formula were giving larger amounts than they give now. There does not appear to be a difference in length or head circumference. The difference in weight appears to be subcutaneous adipose tissue rather than muscle. The best of the studies are those by Dewey (1), but some confounding variables were not taken into account.

Dr. Yetley: What ramifications does have that for using a breast-fed infant as control in a clinical trial of an infant formula, where your primary outcome measure is the weight increment?

Dr. Roche: The main problem with the breast-fed group is that it won’t have been obtained by random selection. So if you do have only two groups, your experimental group and your breast-fed group, you would need to adjust for a lot of intervening variables, and, if you are dealing with small samples and many intervening variables, those adjustments will be quite uncertain.

Dr. Van’t Hof: I may add in this respect that in a study on 2000 infants in Europe, we found that up to the age of 4 months, breast-fed infants grew better than formula-fed infants; later on, the formula-fed children were longer and heavier than the breast-fed children up to the age of 1 year. We don’t yet know what happens after that.

Dr. Haschke: Is it recommended that one should measure length at intervals of 2 months, and the weight in monthly intervals?

Dr. Roche: For an individual, yes; for a group, I measure length monthly as well.

Dr. Rey: What is the minimum size of the sample, if we wish to demonstrate that there is no difference in growth between two groups of infants supplemented or not supplemented with anything?

Dr. Roche: I can’t give you an exact answer to that. One would have to calculate it by picking certain powers and $\alpha$ values that you wanted to use. What you are asking for is a power analysis, and it could be done. The other thing is that you have to take into account the period of the study and whether you are going to analyze the sexes separately. Some people like to combine data for the two sexes, and this can be done if you get exact centiles for every status value, using a computer program that does exist. The problem then is that you base the further analyses on centiles, but centiles are not metric values—they are not
evenly distributed. The difference between the 20th and 21st centile is not the same as the
difference between the 80th and the 81st. So what you are asking about is a power analysis,
and what you are calculating is how many you need at the end of the study, not how many
you need at the beginning. You also have to decide what the outcome variable is going to
be, whether it is the total growth from entry to exit, or whether you want to know whether
this growth difference is significant month by month.

Dr. Walter: This is usually thought of in terms of an equivalent study design. I think Dr.
Roche is correct to say that we need to factor in the usual thinking about type I and type II
errors, but the inherent difference in that situation is that you must think about the smallest
effect between the two groups that you would not like to miss—the smallest difference of
clinical interest—and once you have done that, then you can indeed go through a formal
sample size calculation. It is usually, but not always, the case that the result in sample size is
larger in an equivalence design than in the usual situation where you are trying to demonstrate a
difference.

Dr. Clarke: I would like to know to what extent you think that growth data and reference
ranges should be ethnic-specific, and how specific? We are, at the moment, looking at growth
data from groups of infants from the Indian subcontinent, and we are looking separately at
the moment at Pakistani, Bangladeshi, and Indian infants. I recognize that differences may
be culturally determined in terms of feeding patterns, but I am interested to know what you
think about them being genetically determined and requiring separate examination.

Dr. Roche: I am aware of some of the British literature that shows big differences between
the ethnic groups for individuals living in Britain. I don’t know what the reason for those
differences is—whether it is genetic or whether it is in some sense environmental. Your
opinion as to whether it is genetic or environmental largely determines whether you support
the use of ethnic-specific growth charts or one growth chart for all groups. If the cause is
environmental, and there are differences between groups, the ideal is to improve the economic
circumstances of the group that is growing poorly until they grow like the ones who are
growing better. If the cause is genetic, you are imposing a hurdle that they are not likely to
surmount. In the United States, it is believed that there are no clear signs of genetic differences
among the major ethnic groups—and by that I mean the whites, the blacks, and the Hispan-
ics—with one exception, and that is that during pubescence (but more markedly in girls than
in boys), the Hispanics drop back quite markedly in stature in terms of centile level. It is not
clear whether that is a selection effect or whether it is a real phenomenon. It seems to occur
to a lesser extent in Mexico City, and evidence about whether this is a real phenomenon or
a sampling artifact will come when the NHANES-III data are analyzed. Considering this in
another context, it has been decided that the revised National Center for Health Statistics
growth charts will not be ethnic-specific. The ethnic differences in the United States are not
sufficiently large to justify different charts, irrespective of what the cause of the differences
might be.

Dr. Van’t Hof: You advised that, in clinical trials, one should measure weight, length and
head circumference, but pediatricians do not see these as unidirectional variables: that is,
excessive length gain may point to overestimation, high weight may point to obesity, and a
very large head circumference may point to malformations. In such a situation, does the mean
value have any meaning in the nutritional comparison between different groups?

Dr. Roche: I think the mean does still have some implications, but it is important to go
beyond the mean and look at how many individuals there are in the group who are outside
what is usually called the normal range and are, therefore, at risk of some of the conditions
to which you refer. But that has to be done separately, variable by variable.
Dr. Van't Hof: So this doesn't mean that now that we have new measures such as impedance, we need to get rid of these simple measurements like length, weight, and head circumference?

Dr. Roche: No, because I don't know how you can use impedance in the absence of some of these measurements. You can't use impedance in the absence of height data, and you need other anthropometric values, in particular weight and, if you did get a predictive equation for infants, I would expect head circumference to be included, because head size is such a big factor in the size of an infant. When you talk about additional measures such as impedance, you are going beyond what I had contemplated that people would want to do in nutritional trials during infancy. There are some problems with impedance because with the usual tetrapolar placement of electrodes, one has to have the receiving electrode and the current-providing electrode separated by about 8 cm. It is difficult to do that in the wrist or ankle area of an infant; you just haven't got enough space.

Dr. Meng: China has its own national standards for growth and development. Are the NCHS standards good for the children in different countries, even in China? The national standards in our country are put together from populations from the different regions of the country. In your opinion, which would be best for evaluating the growth and development of children?

Dr. Roche: On the basis of population size, clearly the Chinese data have to be the best of all. These National Center for Health Statistics data are not standards; the word "standard" to me, and to a lot of other people, means a rather narrow range within which an individual ought to be. These are not standards, or limits to which people should be constrained; they are reference data, and they are describing what was the situation in the population of the United States in the late 1960s and early 1970s. It was a surprise to the people involved in constructing those charts when the World Health Organization encouraged other countries to apply them. WHO did that because they recognized them as reference data. They wanted a common set of reference data applied in all different countries, so they could be able to say that in such and such a country, 10% of the children are under the fifth centile, in another, 15% of the children are under the fifth percentile. It has no particular physiological significance. It is, however, true that children in the central part of the range, and you can define that as you wish, but somewhere from the tenth to the 90th or from the fifth to the 95th centile, are less likely to have diseases that interfere with growth or environmental circumstances that interfere with growth. Beyond that, they are nonspecific, and they are not standards. There are, however, data from about half a dozen countries indicating that upper socioeconomic groups come close to the NCHS reference data, which would suggest that the differences between the countries are more dependent on the environment than on genetics. But there is certainly still a genetic effect as well.

Dr. Rey: I am not sure that it is very useful to make these length or weight charts. What are we doing? We are just photographing a group, and if we try to make charts with different groups, we shall mix populations with low growth rate and populations with high growth rate. In the United States, for example, the percentage of obesity in children is too high. Should we use these people to establish weight standards for the United States or WHO? I think this would be a mistake. We should try to draw the ideal growth chart and try to compare populations with this ideal growth chart, if there is one. Finally, for a particular child, it is absolutely without interest to compare his growth with any chart; it is only interesting to compare the growth velocity of this child with his own genetic potential and to see if there is a deviation at one time or another. During the first month of life, the environment plays a very large role in weight increase and probably also in growth in length; we may have to
wait 8 or 10 months before the child escapes from the environment. So it is difficult in infancy to be sure whether there is pathologic deviance of growth rate in a particular child; in any case, it is worthless to draw his particular growth rate on any chart.

Dr. Roche: If we talk about status values, you can plot serial data for a child and see whether a child is tracking—and by tracking, you mean: are those points running parallel to a nearby centile line on the chart? If you didn’t have a chart, I wonder how you would recognize tracking for the individual child. Tracking in another sense means age-to-age correlations for groups. Now when you talk about the genetic potential, the only way in which this gets incorporated into the use of growth charts is to calculate what is sometimes called a target height or a target stature—you take the average of the two parents and adjust it for sex and then indicate the point on the chart that you hope the child will reach at 18 years. There is no equivalent for weight, and there is a lot of uncertainty—the confidence limit of getting to that target height is about plus or minus 15 cm.

REFERENCE