Design of Randomized Clinical Trials: Review of Criteria for Patient Inclusion in Trials of Infant Nutrition

Frank L. Iber

Loyola University, Gastroenterology Section, Edward Hines Jr. Hospital, Hines, Illinois, USA

The randomized controlled trial is the most desirable design to test issues of infant nutrition because it is most free of bias and, through experience, has had the greatest influence in changing delivery of health care. In this chapter, I draw on my experience in studies in this field over the last 10 years and emphasize some of the unique problems of conducting studies in infants.

Randomized controlled trials (RCT) of infant nutrition vary widely in their goals, which in turn largely influence the study design and the criteria for selection of subjects. Nearly all goals in trials in infant nutrition relate to growth and development. A few trials concerned with novel products focus on adverse effects that are common in infants with selected inborn errors of metabolism and infrequent in others. Table 1 lists common goals that are the basis for undertaking trials. Prominent features of setting and reaching the goal are establishing just what will be measured to determine the outcome and which types of infants will be studied.

Growth and development are among the most common and important outcomes of infant nutritional studies. Changes in body weight, length, and head circumference, as well as motor and psychological maturation, are among the most widely used measures, but many surrogate outcome markers are used in order to dissect some biologically appropriate change of maturation or to shorten the period of observation. The commendable desire to shorten the length of intervention required to establish differences in growth and development has led to unestablished methods being used to define the primary outcome. Experience with the variability and rates of change in normal growth of each of these markers largely determines the duration of a study and influences the size of the sample. Roche and Guo (this volume) provide excellent norms for time of change of body weight and head circumference. Unfortunately, many other markers of growth and development are not well characterized for variability and rate of change, and without such information, protocol planning is impaired.

To accomplish the goal of the study, many secondary issues influence the trial
TABLE 1. Purpose of trials in infants of various nutritional treatments

Normal growth and development (1,2)
Catch-up growth in low-birth-weight and short-gestational-age infants
Catch-up growth in diseased infants (3,4)
Maintenance of growth and development
With or following diarrhea (5)
Operation (6)
With impaired diets (7)
Growth and development with special diets (8)
Growth and development in inborn errors of metabolism (9,10)
Measurement of a highly specific function of development or biochemistry (11)
Assessment of toxicity

design and subject selection. Prominent examples of these are listed in Table 2 and are elaborated in this section.

ISSUES OF TIME

Interventions reaching the outcome in a few days or weeks make it easier to enroll subjects and are often conducted with all interventions tightly controlled in the clinic or study unit, thus leading to better compliance. Much longer interventions require more cooperation of the parents and raise problems of compliance with the intervention. Failure to continue the intervention—or to limit feeding to the desired intervention—and failure to obtain the essential outcome measure increase with the length of study. To lessen the loss of subjects, features of the caregiver become increasingly important as entry criteria. When recruitment is prolonged over many months or

TABLE 2. Secondary considerations influencing entry criteria and study design

Issues of time
- Time to complete a single intervention
- Time to complete recruitment
- Time to compete study
Restrictiveness of entry criteria and exclusions
- High restrictiveness may exclude a responsive group of subjects
- Comparability of study group to population targeted for the treatment
- Availability of sufficient subjects
Issues of parent or caregiver
- Extra requirements for parent
- Pain, annoyance or risk to infant
- What is given up or interdicted by study
- Duration
- Conflict with food lore or societal mores
Issues of compliance
- Adherence to the feeding requirements
- Providing primary endpoints
- Methods of statistical treatment of data
years, there is the possibility that changing routine practice patterns or other factors may make those entering the study in the second half quite different from those enrolled initially. Initial entrants often take part in a study because of the enthusiasm of their doctors and the staff over a novel treatment, but this effect is lost after a year or so. Many studies handle lagging recruitment by relaxing entry criteria or adding centers that may be less committed to the study protocol. Extremely long studies often suffer from changes in study coordinators and other key personnel.

ISSUES OF RESTRICTIVE ENTRY CRITERIA

All studies estimate the magnitude of the outcome that must be observed to be meaningful and use this to set the sample size. The sample size is always smaller when the entry criteria are extremely strict and a very homogeneous group of subjects is studied. Such restriction slows recruitment and may limit the generalizability of the results when the admission features are to be extrapolated to the larger population with the condition, who may indeed be quite different. Such restriction often excludes the most severely impaired, who may have the most to gain from the intervention. The restrictiveness of the entry requirements may force the study to be multicentered in order to be completed in a reasonable time.

ISSUES OF THE CAREGIVER

Infants require extensive care, and investigative protocols always place some additional burden on the caregiver. Administration of the diet, obtaining supplies, and transporting the child to the study site for essential observations cannot be avoided, but the bother to the family can be minimized. The longer the study, the more important the continued enthusiastic cooperation of the caregiver becomes. The outcome measures, particularly those involving blood samples or prolonged immobility, often alarm the caregiver because they cause discomfort to the infant. The overall duration that each infant is in a study directly influences the rate of dropout. Because parents differ greatly in their willingness to undertake studies, parental characteristics often enter into the recruitment criteria for more prolonged studies. Increasingly, longer studies involving parents as coinvestigators in all aspects of the study require frequent communication through newsletters and support groups to ensure continued cooperation. The characteristics of the contact staff and the involvement with the study families are important adjuncts to continued participation. Issues of what the child is giving up by participating in the study assume importance in long interventions. The exclusion of all food supplements, the opportunity to travel with the child using locally obtained formula, participation in family celebrations, and so on often determine participation. Particularly in studies in developing countries or immigrant groups, respect for the traditions and mores of infant feeding may influence recruitment. There is a growing belief that the relationship of the research staff with the caregiver is among the most important factors in
determination of both compliance with and continuation in the study until the primary endpoint is attained.

ISSUES OF COMPLIANCE

A run-in period or short trial of an intervention is one of the proven ways to determine short-term compliance. Markers of diet utilization, frequent home visits or phone calls, and occasionally measurements of unique changes produced by the experimental diets are sometimes employed. To assure endpoints in a high proportion of the subjects, frequent measurements of the outcome variable are obtained, so that subjects may be included for the months that the intervention was employed and the outcome obtained, notwithstanding the fact that some patients may adhere to the protocol longer than others. The best-planned studies define exactly what will be done with missing data. Most studies are analyzed on the basis of endpoints obtained, ignoring those who disappeared; this is a setting for potential bias. Patients who drop out and have no endpoint may be healthier than those who remain or may in other ways be different.

There is a consensus among infant feeding experts that breast-feeding is most desirable and optimal among all feeding choices (12–14). Although this is the gold standard of infant feeding, breast-feeding is seldom used in a randomized controlled trial because of the major maternal factors that enter into the decision on whether to breast-feed. A common design adds a nonrandomized comparison group made up of infants similar to the study group whose mothers choose to breast-feed and are willing to undertake the outcome observations. In the best studies, only a single component is different between the two diets, and double-blinding can be accomplished. As a compromise, many studies compare an established diet derived from industrialized nations with the local customary diet, and these cannot be blinded.

Spilker (15), in reviewing quality-of-study issues largely derived from short-term testing of pharmaceutical agents in infants, identified important issues influencing studies, and these are modified in Table 3 to be appropriate to nutritional studies.

REPRESENTATIVE STUDIES IN NORMAL PRETERM AND FULL-TERM INFANTS

Agostoni et al. (16) addressed neurologic development assessed by the Bru- net–Lezine psychomotor score in 60 well-characterized term newborns. Subjects were randomized to formula with or without polyunsaturated fatty acids (PUFA). A third nonrandomized group of breast-fed infants was included. Data were analyzed on the basis of endpoints obtained in 95% of the subjects, and the breast-fed and PUFA-supplemented subjects were more advanced in development than the controls. Carlson (17) and Uauy et al. (this volume) provide insight into improvement of study design in evaluating PUFA in infant feeding. These designs have been largely
TABLE 3. Considerations in designing protocols and interpreting trials on infants

<table>
<thead>
<tr>
<th>Issues in the newborn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Details of labor and delivery</td>
</tr>
<tr>
<td>Medication, smoking, illicit drug use in pregnancy and labor</td>
</tr>
<tr>
<td>Gestational age</td>
</tr>
<tr>
<td>Apgar scores at 1 and 5 min</td>
</tr>
<tr>
<td>Congenital defects</td>
</tr>
<tr>
<td>Length and weight at birth</td>
</tr>
<tr>
<td>Length and weight at enrollment</td>
</tr>
<tr>
<td>Diet until entry if any</td>
</tr>
<tr>
<td>If breast-feeding under study</td>
</tr>
<tr>
<td>Smoking, medications, drugs used by mother concurrently</td>
</tr>
<tr>
<td>Adequacy or exclusivity of breast milk</td>
</tr>
<tr>
<td>Supplements if any for infant</td>
</tr>
<tr>
<td>Issues during first year of life</td>
</tr>
<tr>
<td>Gestational age</td>
</tr>
<tr>
<td>Birth weight</td>
</tr>
<tr>
<td>Percentile of length and weight at enrollment</td>
</tr>
<tr>
<td>Breast, bottle, and solid feeding durations</td>
</tr>
<tr>
<td>Nature and amount of feedings</td>
</tr>
<tr>
<td>Caregiver availability for feeding, observations, travel to study facility</td>
</tr>
<tr>
<td>Age of siblings, other household members</td>
</tr>
</tbody>
</table>

*Modified from Spilker (15), Tables 87.18 and 87.19.*

responsible for the large body of data strongly supporting the need for these fatty acids in selected infants.

Wasserhess *et al.* (18) investigated the effects of supplemental taurine on the neutral and acidic sterols in stools. Groups of six infants, stratified on gestational age and weight for age, were randomly assigned to taurine supplements in a crossover design. Feedings were of 6 days’ duration, with fecal samples collected on days 4 to 6 and 10 to 12. All studies were completed in a hospital unit, and all 30 subjects completed all outcome observations. The results showed that taurine supplements in both preterm and small-for-gestational-age infants led to a significant increase in bile acid excretion and an increased absorption of long-chain fatty acids. This study was extremely well done.

Pichichero *et al.* (19) tested serum and stool responses to rotavirus vaccine given orally to infants aged 2 to 5 months. Spillage of live virus in the stool and development of protective titers of serum antibody were major endpoints; secondary endpoints included the effect of mode of feeding on these outcomes. These investigators showed that infants fed mothers’ milk had a significantly blunted serum antibody response. Stool samples were supplied in fewer than 70% of the cases. Fuchs *et al.* (20) assayed fecal blood loss in infants fed cow’s milk versus several different formulas for at least 6 months. Blood samples were obtained at enrollment and 9 and 12 months of age. Stools were obtained monthly. Fecal blood measurements and blood iron indices were the outcome measurements. Although none of the diets proved different, blood endpoints were available in 90% and stools in 60% to 70%. Analysis was performed only on the endpoints obtained.
A study of 40 infants with cow’s milk intolerance compared an elemental diet with hydrolyzed whey formula (21). All subjects were full term and enrolled an average of 10 weeks after birth. Food diaries were kept, and growth and plasma proteins were the outcome variables. Only two-thirds of the enrollees completed 24 weeks of the diet and were the basis of the evaluation. Among the subjects completing evaluation at 24 weeks, growth was increased, and plasma albumin was higher for the whey than the elemental diet, but the differences were not significant. The groups studied were highly heterologous, the entry allergy was not confirmed, the compliance was poor, and the outcome measures were limited. This limited study is representative of much in the field of special diets for atopic infants (22).

STUDIES IN ACUTE AND CHRONIC DIARRHEA

Diarrhea is most widely studied in the nonindustrialized world, where it is a particular problem and is often associated with malnutrition. Diarrhea is mostly identified by a history of frequent or watery stools, and studies on diarrhea usually separate cases with duration less than or greater than 7 days. Stool weight is rarely measured. Issues of compliance with the diet under study, availability of the diet, precision of the measurements, and the traditions and mores of the population under study interact to determine the exact nature of the study. Short-term studies are often carried out entirely in a clinic-based study unit, but longer ones involve the home and caregivers.

Maulen-Radovan et al. (23), in a research unit in Mexico City, studied 87 boys aged 5 to 36 months with diarrhea less than 96 hr in duration and associated dehydration. Patients were excluded if they had had diarrhea within the previous 14 days, had systemic infection, or were breast-fed. Children whose postrehydration weight for length was less than 2 SD below the median of the U.S. National Center for Health Statistics were also excluded. Nine percent of the subjects did not complete 5 days of observations. All subjects were rehydrated with WHO solution and then randomly assigned to a puree of chicken, brown beans, and carrots or to a soy formula (Isomil, Abbott Laboratories). Treatment failure was defined as continued or recurrent diarrhea in the 5 days of study and occurred in six patients, all of whom received the soy diet (p < 0.01). The overall duration of diarrhea and the stool volume the first day of feeding were significantly less in the children fed the puree, and the weight gain after 5 days was significantly more. The study was highly conclusive, and the observations were made in a study unit, but commonly encountered diarrhea cases were among those excluded.

Darling et al. (24), working in a study unit in Dar es Salaam, Tanzania, compared three locally available infant feedings given at weaning. The conventional porridge was prepared with 66 g corn flour, 28 g peanuts, and 906 g water to provide 39 kcal/100 g. Germinated grain from sorghum was sun-dried, then mixed with 123 g corn flour and 858 ml water to provide an energy density of 58 kcal/100 g. One preparation was fermented for 24 hr; the other was boiled promptly to destroy the amylase.
Selected children were aged between 6 and 25 months, largely weaned, with a history of diarrhea of <14 days. Only infants requiring a nasogastric tube, those with kwashiorkor, or those studied for less than 1 day were excluded. Children were fed five times daily with all diets, which were weighed and recorded, but many were also breast-fed, and these feedings were not quantified. Of 75 children entered, six were excluded. The groups studied proved similar in weight and the presence of other diseases. The group fed amylase-digested porridge ate 42% more food during the 4 days of study than those fed the conventional porridge, but the duration of diarrhea was similar.

This study, much less restrictive in its entry criteria, was carefully controlled in stool frequency but not stool weight. The information on dietary intake was flawed by the absence of recorded volumes of mothers’ milk. The study was too short in duration to show a beneficial effect on malnutrition of the predigested corn flour. The stunting in growth in diarrhea has been addressed in open parallel studies given supplements and can be partially reversed (25).

A highly important group of patients were studied in Dhaka (Bangladesh) by Rahman et al. (26). Severely malnourished children aged 5 months to 5 years, whose weight for age was less than 60% of the median for the U.S. National Center for Health Statistics, were studied after their diarrhea and associated diseases had been treated in hospital. Two diets were compared: one used a wheat-based porridge, and the other the same porridge preincubated with an amylase-rich flour to make the diet more digestible.

All children and mothers were instructed and observed in the hospital in the preparation and feeding of the assigned diet, which was given for 5 days. Exclusion criteria included diarrhea during the observation period in hospital and children who had bacteremia or pneumonia during the initial evaluation. Eighty-one children were enrolled, and all but three (with severe diarrhea) were considered in the initial 5-day study. The important result is that the total energy ingested in those on the amylase-predigested diet was 33% more than in those on the nondigested diet. The predigested diet was particularly effective in infants aged 6 to 11 months. This study, though short-term, is quite convincing. A much longer follow-up of similar patients is needed.

Persisting diarrhea (defined as symptomatic for more than 2 weeks) was the basis of a study in Dhaka, Bangladesh, reported by Ray et al. (27). Twenty-six boys aged 4 to 18 months passing more than two liquid stools daily for more than 14 days were selected. Exclusions were breast-feeding with no intention to introduce supplements, presence of cholera, Salmonella, or Shigella, systemic infections, and kwashiorkor. Twenty-five age-matched healthy controls who were treated for acute diarrhea at least 4 months previously with no subsequent diarrhea were included. Both groups were fed a rice-based diet containing egg white, soybean oil, and glucose for 7 days of detailed observations. Of the patients with persisting diarrhea, 81% recovered within a median time of 4 days. Among those whose diarrhea ceased, much less energy, nitrogen, and fat was recovered in the stools than in those whose diarrhea did not stop. Those in whom diarrhea stopped improved rapidly in nutrition, whereas
those with continued diarrhea did not. This study showed an important impairment of gut absorption in the malnourished infants, the impairment being greatest in those with the most malnutrition. In marked contrast, an industrialized nation approach to intractable diarrhea involved the development of an enteral feeding plan based on an audit of 29 patients and tested in the next 16, randomized into two treatment groups. This showed a diminution in time needed for parenteral nutritional support and fewer days in hospital (28). In another study, a somewhat similar group was randomized to total enteral versus parenteral treatment, and the authors concluded that enteral treatment was advantageous (29).

Brown, in reviewing the contemporary issues of dietary treatment of diarrhea (30), including study design, emphasizes directions for new studies. Brown's group (31), in a study performed in Lima, Peru, found the addition of fiber produced a prompt reduction in watery diarrhea but did not change any of the other measured variables, including stool weight, nutrient absorption, and dietary intake. Allen and Uauy (32) discuss the common problem of stunting of linear growth and its nutritional treatment. They review the genetic factors, the intrauterine and postnatal nutrient supply, the importance of infection, and the need to investigate and control for each of these. Many suggested controls, enrollment investigations, and even measurements to be made are presented.

STUDIES IN PARENTERAL NUTRITION

Although the studies are usually conducted in the hospital, and many of the outcome observations are the same as those required for the control and safety of the treatment, the clinical trials are often flawed. The infants in whom partial or total parenteral nutrition is required vary markedly in their conditions, duration of the requirement, and in the previous growth and disease characteristics.

Thus, McIntosh and Mitchell (33), in a study of 68 neonates given partial or total parenteral nutrition in the first week of life, stratified them by birth weight above or below 1000 g. Subjects requiring parenteral nutrition less than 5 days were excluded after randomization—a total of 20 patients. Two different compositions of parenteral amino acids were studied; body weight, head circumference, and plasma amino acids were the outcomes. No differences were found, but the loss of 20 of the randomized subjects and the absence of the blood endpoint in one-third of the analyzed subjects were major weaknesses. One arm had sicker infants and more deaths. Rosenthal et al. (34) also compared two amino acid formulations in all infants requiring parenteral nutrition. Seven of 39 infants were excluded after randomization for various reasons. Most infants less than 32 weeks of gestation had idiopathic respiratory distress; most over 32 weeks had surgery. Although differences were identified, the small groups, the marked differences in diseases, and the few infants requiring treatment for 21 days or longer rendered the results unconvincing.

A much better study was completed by Hammerman and Aramburo (35), who studied 42 neonates less than 1750 g in birth weight with respiratory distress
syndrome randomized to parenteral solutions with or without lipid. On days 3 and 5, retinopathy was assessed by electroretinograms, and blood prostaglandins were assayed. The results showed clearly that the lipid-infused infants had a more prolonged requirement for ventilatory support and supplemental oxygen, higher grades of bronchopulmonary dysplasia, and more severe retinopathy. It was concluded that lipids should be withheld in this group. The selection criteria were strong, and the exclusions carefully preplanned. The endpoints were obtained in all of the patients, and meticulous clinical follow-up provided many of the statistically significant differences.

CONCLUSIONS

Table 4 sets forth my conclusions regarding a review of controlled trials in infant nutrition published in the last 10 years. These are organized to educate clinical investigators about the frequent shortcomings in study design from otherwise well-motivated and well-intentioned persons.

There are many studies that simply should not have been done. The chapters in this volume by Whitehead and Kauffman support the notion that a study that cannot answer the hypothesis or purpose for any reason should not be carried out. In various publications, the primary outcome measurement is simply too new for there to be established criteria of reliability and variability, condemning the study to failure. In a desire to provide measurement of growth or development within a week or two

| TABLE 4. Comments on randomized controlled infant nutrition trials published in the last decade |
| Studies that should not have been done |
| Unclear hypothesis |
| Primary outcome measure not established as valid |
| No estimate of sample size; reasonable estimates indicate no possibility of reaching a conclusion |
| Inadequate size of study population |
| Weak and inconclusive studies |
| Relaxed or changed entry requirements |
| Primary endpoint not obtained in large numbers |
| No measure of adherence to complex regime |
| Rare events discussed in a conclusive manner when sample size is inadequate |
| Common design shortcomings |
| Failure to blind treatments |
| Subjective criteria evaluated by unblinded observers |
| No measure of compliance |
| Analysis does not consider “intention to treat” limited to primary endpoints obtained |
| Inadequate staffing or training to promote retention and compliance |
| Well-designed and -executed trials |
| Substantial statistical input in planning |
| Multicentered |
| Ample support staff |
| Blinded at all levels |
of the intervention, indirect or surrogate markers (36) of growth are used. In many studies, there is no estimate of sample size or discussion of the magnitude of change that is being sought, and reasonable guesses of a meaningful outcome simply exclude useful information based on the size of the samples. With encouragement, consultation with statisticians at the time of planning, and stronger function of human study committees, these should largely disappear.

Slightly stronger studies emerge with very inconclusive results. The two most common reasons are that patient selection is too broadly based to be meaningful and the primary endpoint was unattained in more than one-third of the patients. Both of these problems can be overcome by better planning and appropriate staffing to obtain more primary endpoints. In many cases, each of these solutions requires funding that is often not available but is essential to convert inconclusive studies into those with a much higher likelihood of success.

Many studies that are both adequately designed and of sufficient power to answer the intended goal or hypothesis have shortcomings that introduce potential bias. Some of the common difficulties are noted in the absence of ready solutions. Diets available in developing countries are often compared with formula diets from industrialized countries to be certain of including all reasonable nutrients, but blinding is impossible. There is no satisfactory method of fully handling subjects who fail to provide the primary endpoint.

Finally, there are many state-of-the-art studies. These are most commonly well supported to allow for careful planning of the study and its measurements and adequately staffed to diminish dropouts and to increase compliance. Granting agencies, university research departments, and industry could assist in improving the quality of studies by underwriting the costs of statistical planning with clinical investigators.

We should also reexamine the major commitment to randomized, blinded, controlled trials. A properly carried out study is extremely expensive, and there are other methods of research design that are sometimes more feasible and may allow the inclusion of larger numbers and greater varieties of patient to give an even more conclusive result. Very infrequent events, such as serious side effects, cannot be studied by randomized blinded controlled trials. Other designs include single-patient assignment to randomized treatment of variable duration and crossover designs in small numbers of patients. The advantages and disadvantages of each of these have been discussed (15).

REFERENCES


DISCUSSION

Dr. Hamburger: Could I ask you to comment further on what can be done to neutralize the loss or the negative effect of large numbers of dropouts.

Dr. Iber: One technique that readily identifies the dropout patient is a run-in period. For many studies, a 1- or 2-week period to allow the formula or the intervention to be tried, combined with more than one visit for physical examination, is a way to identify the extremely weak of heart who will drop out very promptly. I think that that should be done in many studies if it can be built into the design. Otherwise, I think that trained staff and proper planning of the study are more important in diminishing dropouts than almost any other feature.

Dr. Lucas: We agreed that it is unethical to do a badly designed study, and one of the aspects of bad design is a high dropout rate, and particularly for certain sorts of study. In neurodevelopmental outcome studies, selective dropout rates can have such a major impact on the interpretation of the results that they become meaningless if the dropout rate reaches, say, 10% to 20%, so for certain types of study, it is important to define circumstances where a low dropout rate is going to be achieved. This means ensuring that patients can be traced, and much more sophisticated tracing techniques are required than are often employed at present—it shouldn’t be left to random chance. There are certain types of study that simply can’t be done in particular countries, where there are very diffuse and mobile populations, if you really want to achieve a good follow-up. So as part of the ethics of designing a study, it is very important to consider whether the dropout rate you are likely to achieve will, in fact, make it an ineffective study; for a neurodevelopmental outcome study, for instance, a 30% or 40% dropout rate is completely unacceptable.

Dr. Iber: I agree entirely. One of the frustrations in planning a study is that often the population you most need information on may be the most difficult both to find at a later time and to influence to continue the intervention. I applaud continued efforts to tackle better studies in the target groups that need better studies. In America, many of the people with poor medical care and living in poor circumstances are avoided like the plague by investigators because of the traditionally very high dropout rate, and yet, information is desperately needed in these groups.
The great success of the data on fertility control came about in large part because it had a built-in safety net and was very effective in people who weren’t completely compliant.

Dr. Guersy: My first point is that you mentioned the tendency to do shorter studies, taking the baby into hospital for 2 weeks to ensure a compliance. This is fine if the baby is admitted because of pathology, but to admit a healthy baby to hospital for 2 weeks just to do a clinical study would not pass our ethics committee. My second point is related to the paper by Agostini that you cited as an example of a good study. As far as I know, the 4-month data were published only as a letter in Lancet. The statistics were doubtful, and Mr. Baumgartner, who is one of our statisticians, found that the statistical significance was obtained only because there were two outliers in the breast-fed groups. Third, and this is my main point, the choice of the endpoint is very important. To use the development quotient at 4 months of age as the endpoint is dubious when you take into account the difficulties of doing such tests—in particular, the huge individual variability. What I say was confirmed last May in Munich when Agostini himself presented results showing that at 1 year of age, there was no difference whatsoever among the three groups.

Dr. Iber: Thank you for your comments.

Dr. Uaay: I tend to agree with Dr. Guersy. However, Agostini’s paper was fully published in Pediatric Research (1). I also agree that long-term effects are what we should be looking for, but we should not neglect the possibility that there may be transient effects. A transient effect is still a valid observation.

Dr. Iber: I would emphasize that I only looked at papers in my review, not letters.

Dr. Haschke: Alan Lucas calculated the sample size that is needed in neurodevelopmental outcome studies to be 80 in each group, and in the Agostini study, we have only 20.

Dr. Lucas: If you take the breast milk model, the advantage in cognitive development for breast-fed over formula-fed babies, which is purported by some people to be related to the LCPs in breast milk, is around a third of a standard deviation, which is about 5 DQ points. If you wanted to test the hypothesis that all of the difference was caused by LCPs, then you would need 144 subjects per group for 5% significance and 80% power, but of course, if you hypothesize that only half of the difference between breast-fed and formula-fed babies was linked to LCPs, then you would need groups of over a thousand.

Dr. Whitehead: It is incredibly difficult in a community-type study to control for everything, and anybody who carries out such studies is fully aware of the shortcomings of their studies and the assumptions that have to be made. It might make it easier for fellow scientists to follow these arguments if the editorial boards of various journals were rather more tolerant of people spelling out the shortcomings of their studies. I know about the big demand for space and so on, but if you look back at very early reports in journals like Lancet, there was a tremendous amount of discussion of this sort of thing. Now you just never see it, and I think that is creating problems—indeed investigators are often blamed for oversimplifying the presentation of their results when all the pressures are on them to oversimplify and it is difficult for them to be truly honest.

Dr. Iber: I think multicenter controlled trials often have methodologic papers that appear in Controlled Trials. The policy of that journal is to recognize the importance of allowing the designers to include extensive discussion about areas of uncertainty, or where they made compromises to keep cost within reasonable limits or widened their expectations so that the power of the study fitted the anticipated budget. I think that that is the only place I have seen such discussion published.

Dr. Salle: You pointed out that it is very important to look at the effect of nutrition on growth parameters. Don’t you think that it is also important to look at the quality of growth,
particularly in premature babies? We have many tools now for measuring quality of growth, DEXA and indirect calorimetry for example, and it seems to me that it is now very important to do this so that we can assess precisely the effect of different low-birth-weight formulas. Would you comment on that?

Dr. Iber: I think you are correct in a certain type of protocol, for example, where you are looking at correction of nutritional injury imposed by surgery or an acute infectious illness, where it is important to document the immediate response to the maximum amount of nutrient that can be accepted. I think that we need to focus on a hypothesis. In certain studies, the issues you raise would be important; in others, they may well not be.

Dr. Hamburger: I would agree with you. Wherever you can add quantitation, your result is going to be easier to interpret.

Dr. Glinsmann: Would you speak a little bit about the intention-to-treat analysis: how often should it be used, and what do you think its usefulness is when you do, in fact, have a poor study with a large dropout rate?

Dr. Iber: I think we have all become addicted to the randomized blinded controlled trial. The emphasis to avoid bias is on a completely blinded randomization, and only by using that blinded randomization in analyzing the data is one continuing the strength of the controlled trial system. As soon as you statistically analyze only the endpoints that you have, ignoring the dropouts other than to record them, you introduce an influence of factors that lead to dropout and factors that may influence efficacy. These may be correlated; there are data suggesting that poor compliance is a precursor to dropout, and therefore, the dropped-out people might be a very different population and will, because you have no endpoint, be ignored. The difficulty is that when you are talking about growth, where you need two points, if you don’t have that second point, you have to make assumptions of what would happen if it had been zero, or what would happen if it had been the same as the rest of the patients. There are statistical adjustments, but they are all relatively poor guesses, and when the dropout rate approaches 50%, as it does in a fair number of studies for the primary endpoint, you really need to think of another design. We need to devote more thought to innovative designs. Crossover studies have great strength, even though the child is developing and growing. Multiple endpoints provide another way of handling the problem, but if you put too much burden on the caregiver and the subject, you will have major recruitment problems.

Dr. Hamburger: Our secret weapon for compliance and dealing with dropouts was our nurse coordinator, who, in several of our studies, developed personal relationships with every single participant so that compliance increased, and when we had the expected dropouts, she was able to find them and follow up adequately.

Dr. Lucas: On the question of dropouts, I think we may have confused what we mean here. Let’s say that we are doing a randomized intervention study of two formulas. Subjects may drop out of the short-term limb because they develop some problem on the formula, but that doesn’t mean that they necessarily drop out of the follow-up. It is extremely important to follow up such “dropouts”; it is only when the persons actually refuse to have the long-term follow-up, which they had originally agreed to do, that you have a serious problem. So what you need to do is to analyze with your primary outcome all the subjects, regardless of whether they have been on the treatment or not. That is very different from what has been described as dropping out.

Dr. Iber: I agree with that.

Dr. Ferry: One of the problems with dropouts is, of course, noncompliance: patients who fall out because they don’t keep appointments, or who don’t stay on whatever the therapeutic product is, or what have you. I think this comes back to the issue of truly adequate and
thoroughly thought out inclusion criteria, because that is the point at which you really need to think about all the variables that may cause patients to be noncompliant. There are many issues you can take into account right from the start, and this will—at least to some extent—help decrease the number of noncompliant patients.

**Dr. Iber:** I would agree that planning and staff training make a tremendous difference in all studies, and certainly in some studies, these issues have not been given enough attention. However, restricting entry to patients who have characteristics that lead to high compliance may well restrict the applicability of the study results when they are finished because individuals who are the most compliant are a modest subset of the population. I don't think that compliance can be the only thing that determines the design of a study.

**Dr. Fishberg:** I am very concerned about the economic impact, especially when you calculate a very high level of dropouts. These subjects will be exposed to the same measures as the others, but the data will be abandoned. Don't you think it would be wiser to take these studies to other countries or populations with less chance of dropout, especially in underdeveloped countries?

**Dr. Iber:** We must consider what you are trying to accomplish. There are some studies that, because of the uniqueness of the illness or the population or the partial deficiency, can be carried out only in the less developed portions of the world, and though you could have a much higher assurance of compliance and a lower dropout rate by doing a study in an industrial country, this would never allow you the numbers you need. We agreed that poor science is unethical, and I think that many of these problems can be anticipated with planning or perhaps, in a very large study, with a preliminary trial. And one can have a factual data base to suggest what the needs are, what the follow-ups are, and so forth. The multicenter studies funded by the National Institutes of Health often require a small feasibility study of maybe 100 subjects, and then, when they undertake the main multicenter study, they will initially fund two or three centers a year ahead of everyone else to work out the problems and allow a bit of experimentation to determine pitfalls and unanticipated problems before the study spreads to 15 centers, or 25, or what have you.

**Dr. Hamburger:** I would like to ask some of the statisticians, and Dr. Walter particularly, how often they are consulted in the design of a study, in contrast to how often they are asked to look at the results and try to bail out some findings that are inexplicable.

**Dr. Walter:** I would like to make two comments about some other issues first. First, on the crossover design that was suggested as a possible solution, it is true that such designs have a lot of appeal because you can compare the treatments within subjects, and that generally leads to a more efficient comparison, but only if you get the observations. I think that the problems of dropout and compliance are going to be more severe in a crossover design because participants have to commit themselves to two treatments rather than one at a minimum if it is a two-period study. So you need to think very carefully about that, quite apart from all the other issues introduced by using a crossover design, such as a carryover effect, which is important. My second comment is that these very high dropout rates that have been mentioned of up to 50% are much larger than any I have experienced in the fields that I work in. But one proposal that has been made in this situation is to look at "drop out" as an outcome, because it may be that the fact of dropping out is some kind of indication of acceptability of the treatment, and that is certainly something you would want to know about if you were thinking about using a particular nutritional therapy in the future. If it is unacceptable to 50% of the people to whom it is offered, then that obviously has some implications about its clinical usefulness. So I think dropout itself can be analyzed as an outcome in some circumstances. And then, to respond to Dr. Hamburger's question, I suppose people's circumstances differ
enormously. In my institution, I think we have our clinicians relatively well trained, and they do tend to come along quite early; we encourage a longer-term relationship with investigators and actively discourage the people who come at the last minute.

Dr. Uauy: Every experimental design is a compromise between the ideal and the real. We have to be ready to do our best in the planning stage, but we also have to be willing to do the best analysis and the most critical review of our results. I would not dismiss data that have been carefully planned just because they did not meet the ideal. Moreover, most ideal studies become unfeasible because one would like to include, for example, mechanisms behind the observations, and usually in nutritional research that means blood drawing, invasive procedures, or whatever, and that interferes with compliance. So in any design, there will be compromise. My plea is that we stick to reality and judge that there is really no ideal study. We learn in the process to do better studies, but we must also judge the level of the field we are working with: initially you are going to have stage 2 studies, which will be limited and do not pursue the aim of changing clinical practice. Where we need very clearly defined standards is when we feel that the level of knowledge has come to a point where it might change a clinical practice or a feeding practice. But many studies are done in a preliminary phase to provide a better hypothesis or mechanism.

REFERENCE