MINUTES

Of the

Food Advisory Committee's

AD HOC TASK FORCE on INFANT FORMULA

Meeting

November 18-19, 2002

USDA Building Conference Center
4700 River Road
College Park, MD

Members present: Cutberto Garza, M.D., Ph.D., Chair; James Anderson, Ph.D.; Robert D. Baker, M.D., Ph.D.; Margaret Briley, Ph.D., R.D., L.D.; Scott Denne, M.D.; Goulda A. Downer, Ph.D., R.D., C.N.S.; James E. Heubi, M.D.; Lawrence N. Kuzminski, Ph.D.; Laurie J. Moyer-Mileur, Ph.D., R.D., C.D.; Madeleine J. Sigman-Grant, Ph.D.; Virginia A. Stallings, M.D.; and Patti Thureen, M.D.

Acting Industry Representative: Roger Clemens, Dr. P.H., Director, USC School of Pharmacy, Laboratory for Research and Services in Complementary Therapeutics, University of Southern California

Food and Drug Administration (FDA) representatives: (Center for Food Safety and Applied Nutrition – CFSAN), Jeanne Latham, Executive Secretary; Elizabeth A. Yetley, Ph.D., Lead Scientist for Nutrition, CFSAN; Christine Taylor, Ph.D., Director, Office of Nutritional Products, Labeling, and Dietary Supplements (ONPLDS); Dr. Susan Walker, Associate Director for Clinical Affairs, ONPLDS; Ms. Mary Ann Killian, Program Integrity Adviser, Office of Human Resources and Management Services

Guest speakers: Duane Benton, Ph.D.; Dennis M. Bier, M.D., Professor of Pediatrics, Baylor College of Medicine, USDA/ARS Children's Nutrition Research Center; W. Cameron Chumlea, Ph.D., Fels Professor, Department of Community Health, Wright State University School of Medicine, Lifespan Health Research Center; Kenneth J. Ellis, Ph.D., Professor of Pediatrics, Baylor College of Medicine, USDA/ARS Children's Nutrition Research Center; Samuel J. Fomon, M.D., Emeritus Professor, University of Iowa; Edward A. Frongillo, Jr., Ph.D., Associate Professor of Public Nutrition, Division of Nutritional Sciences, Cornell University; Lawrence M. Grummer-Strawn, Ph.D., Branch Chief, Division of Nutrition and Physical Activity, Maternal and Child Nutrition Branch, Growth Chart Working Group, Centers for Disease Control and Prevention; Jon Tyson, M.D., Ph.D., Michelle Bain Distinguished Professor of Medicine and Public Health, University of Texas - Houston Medical School

Public speakers: Barbara Heiser, R.N., B.S.N., IBCLC, Executive Director of the National Alliance for Breastfeeding Advocacy, Inc. (NABA); Russell Merritt, M.D., Ph.D., Director, Medical Affairs, Nutritional, Ross Products Division of Abbot Laboratories; Jose M. Saavedra, M.D., Medical & Scientific Director, Nutrition Division, Nestle USA; Jon A. Vanderhoof, M.D., Vice President, Global Medical Affairs, Mead Johnson Nutritionals, A Bristol-Myers Squibb Company

Summary Conclusions

The Infant Formula Task Force was asked to consider seven questions under two issues—1. criteria for the adequate evaluation of normal physical growth during the first six months as an indicator of the nutritional adequacy of new infant formulas, and the 2. types of changes in infant formulas that should be accompanied by a clinical study in order to provide assurances of normal physical growth. Questions 1 through 6 addressed the specific criteria covered under the first issue about criteria for evaluating normal physical growth. The types of changes in infant formulas that should be accompanied by a clinical study were addressed in Question 7

Question 1: Considering the values and merits individually, and in combination, please group the following metrics in terms of their clinical usefulness as endpoints for assessing normal physical growth.

- *Body weight,*
- *Recumbent length,*
- *Head circumference,*
- *Skin fold thickness,*
- *Bioelectrical impedance,*
- *Stable isotope, dual energy x-ray absorptiometry, or*
- *Other physical body measurements or body composition measurements*

Recommendation: Task force members reached consensus that body weight, recumbent length, and head circumference are the three metrics that are extremely useful indicators of infant growth. Skin fold thickness was designated a metric of moderate use. Bioelectrical impedance, stable isotope, dual energy x-ray absorptiometry, and other physical body measurements or body composition measurements were deemed to be in the research stage and therefore task force members felt they were unable to comment on the effectiveness of these metrics in determining normal physical growth in infants. Several task force members indicated that there was no basis for the use of bioelectrical impedance or stable isotope metrics in infants six months or younger.

Task force members also considered the question in regard to preterm infants. Their determinations were the same for this population, with body weight, recumbent length, and head circumference identified as the three metrics that should be used as mandatory measures of infant growth. Most committee members agreed that head circumference is critical in preterm infants. Skinfold thickness measurements are of moderate interest. There is no role for bioelectrical impedance. Stable isotope and other physical body measurements are in the research area.

Question 2: Which of the above anthropometric and/or body composition measures are necessary for adequate clinical evaluation of normal physical growth of infants between birth and 6 months of age consuming new infant formula?

Recommendation: The task force reached consensus that body weight, recumbent length, and head circumference are necessary for adequate clinical evaluation of normal physical growth of infants consuming new infant formula between birth and six months of age for both term and preterm infants.

Question 3a: The metrics above can be evaluated as attained (absolute growth) or velocity (rate of change) measures. Please comment on the distinguishing values and merits of each static or variable method in the assessment of normal physical growth.

Recommendation: Task force members reached a consensus that for infants in a study, baseline measurements, i.e., weight, recumbent length, and head circumference, should be taken at birth or no later than 14 days. These three measurements should also be taken at 1 month, 2 months, 4 months, 5 months, and 6 months of age.

The same guidelines were recommended for preterm infants after the infant is discharged from the hospital through 6 months postconceptional age. More frequent measurements—every week—also were seen as necessary for monitoring preterm infant growth while in the hospital.

Question 3b: The outcomes above can also be evaluated as individual infant data or as group comparative data. Please comment on the values and merits of using individual or aggregate data in the assessment of normal physical growth.

Recommendation: Task force members reached consensus that it would be beneficial to see both study data on individual infants and the group comparative data. The value and merit of having individual data are that we would have the ability to assess distributions and potential outliers, as well as other information, in a way that summary data may not lend themselves to easily.

Question 4: For adequate evaluation of normal physical growth, below are examples of clinically distinct reference groups.

- *Concurrent controls (concurrent data or population cohorts for demonstration of bioequivalence)*
- *Reference data used as controls (comparison with previously collected normative data for populations and subpopulations)*
- *Historical controls*
- *Other*

4a: What are the distinguishing values and merits of each type of reference group for the assessment of normal physical growth?

Recommendation: Task force members reached consensus that concurrent data or population cohorts are essential for evaluating growth in infants from birth to 6 months of age, with the

understanding that one concurrent control group could be used for multiple studies. Longitudinal reference data were seen as the second most adequate reference group, with cross-sectional data third. Historical reference data were seen as the least helpful and relied upon only under unusual circumstances.

Task force members reached consensus that concurrent controls are necessary for all studies involving preterm infants. Because of the dynamic nature of the treatment of preterm infants and the center differences that exist, one cannot rely on either reference data or historical controls.

4b: *Please rank these reference groups based upon the ability of the respective control population to contribute to an assessment of normal physical growth in the population intended to consume the formula.*

Recommendation: Task force members ranked the reference groups as follows: longitudinal concurrent data, which will be needed in most circumstances; then cross-sectional data, and finally historical data. One concurrent control group could be used for multiple studies. There should be a comparison in addition to some reference source based on the breastfed infant to try to understand deviations, if any, between growth patterns exhibited by breastfed infants and formula fed infants targeted for study.

4c: What is the role of such a reference group?

Recommendation: Presently available reference data were seen to have comparative value, but were not seen as a standard against which concurrent studies should be evaluated.

Question 5: For the purpose of evaluating normal physical growth of infants fed new formulas, what criteria should appropriate infant growth reference groups meet (e.g., each or selectively, feeding history, gestational age at birth, sex, racial background, socio-economic status, other)?

- *In comparison to the study population?*
- *In comparison to the population intended to consume the formula?*

Recommendation: Task force members determined that the comparison should be to the population intended to consume the formula, e.g., term infants cannot be used to determine effects for preterm infants or that a study of infants older than 6 months cannot be used to determine effects for infants age birth to 6 months. Both the study and control groups should be randomized and matched for criteria such as sex, feeding history, gestational age, and general health. These recommendations apply to studies of both term and preterm infants.

Question 6: Listed below are examples of control feedings (clinical comparators):

- *(current infant formula (IF) + new ingredient) vs. (current IF) vs. (breast milk)*
- *(current IF + new ingredient) vs. (current IF)*
- *(current IF + new ingredient) vs. (breast milk)*
- *(current IF + new ingredient) vs. (formulas fed to historical infant cohort(s), e.g., Iowa data)*

?? *(current IF + new ingredient) vs. (references that may include various type of feedings in such reference populations, e.g., NCHS and WHO)*
?? *(IF + new ingredient)* vs. (any of the above controls)*

**Test formula contains new ingredient but the test formulation matrix differs from the new formula that firm intends to market containing the new ingredient.*

a. What are the most distinguishing values and merits of each of these types of comparisons in infants fed a test formula vs. a comparative feeding for assessing normal physical growth?

b. *Please rank these comparison based upon their potential for generating clinical data, which would be most relevant to an assessment of normal physical growth.*

Recommendation: Task force members considered parts a and b together, reaching consensus that the clinical comparators with the most potential for generating valuable clinical data on normal physical growth would be a comparison that included the current formula, the current formula plus the new ingredient, a breastfed cohort, and three references, e.g., Iowa, CDC, WHO, NCHS, NHANES.

Question 7: With regard to formula composition changes:

a. Describe general principles and criteria that can be used to determine the need for a clinical study intended to provide assurance of normal physical growth.
b. Describe some of the specific changes in infant formula that would reasonably be expected to be accompanied by a clinical study to demonstrate normal physical growth.

Recommendation: Task force members reached consensus that the following specific criteria would trigger the need for a clinical study:

?? Major change in manufacturing process
?? Entirely new formula
?? Use of a substance that has not been tested in children before
?? Major changes in macronutrient content
?? Use of other compounds known to affect hormones, growth factors, genes, or metabolites that regulate growth
?? Formula changes that result in nutrient levels outside established ranges
?? Alterations likely to affect GI function or nutrient bioavailability
?? Studies on extremely vulnerable populations
?? Use in different population than for whom the formula was intended originally

Agenda

Dr. Cutberto Garza, chair of the Infant Formula Ad Hoc Task Force of the Food Advisory Committee, convened the meeting at 8:15 a.m., Monday, November 18, 2002. After welcoming all present, he introduced Dr. Christine Taylor, from the FDA's Center for Food Safety and Applied Nutrition.

Dr. Taylor welcomed members to the second meeting of the task force and provided an overview of the committee members' role. The task force currently operates as the Ad Hoc Task Force for Infant Formula of the Food Advisory Committee. In the future it will be the Infant Formula Subcommittee of the Food Advisory Committee. Dr. Taylor named as temporary voting members Dr. Cutberto Garza, Dr. Virginia Stallings, Dr. James Heubi, Dr. Patti Thureen, Dr. Robert Baker, Dr. James Anderson, Dr. Laurie Moyer-Mileur, and Dr. Scott Denne. Food Advisory Committee members participating were Dr. Goulda Downer, Dr. Lawrence Kuzminski, and Dr. Madeleine Sigman-Grant. Dr. Margaret Briley is the consumer representative and a voting member. Dr. Roger Clemens served as the non-voting "acting industry representative."

Dr. Taylor asked task force members for their scientific input on principles and criteria for evaluating whether a new infant formula supports the normal physical growth of infants under Section 412 of the Food, Drug and Cosmetic Act. The three general topics before the task force are: 1) growth measures and methodologies; 2) role of such measures and methodologies in demonstrating normal physical growth; and 3) principles and criteria to determine the need for a clinical study to provide the agency with an assurance of normal physical growth.

Dr. Taylor then turned the meeting back over to Dr. Garza, who read a letter from Cathy DeRoever to the task forcee pertaining to administrative issues. The letter from Ms. DeRoever asked task force members to provide their best scientific advice in an open and transparent manner in order to avoid the appearance that issues have been decided or conclusions reached outside the meeting. She also cautioned that all contact with members should be routed through the committee Executive Secretary, Ms. Jeanne Latham. Dr. Garza asked task force members if they had questions regarding the content of Ms. DeRoever's letter; they did not.

Ms. Latham identified the temporary voting members appointed by Mr. Levitt, Director of the Center for Food Safety and Applied Nutrition. Issues for discussion were deemed to be of broad applicability with no particular impact on specific products. Under 18 USC Code 208.3, Dr. Garza has been granted a waiver to participate in matters in full. The following reported interests were disclosed for the guest speakers: Dr. Cameron Chumlea has a grant from Nestle; Dr. Samuel Fomon previously was a consultant to an infant formula manufacturer; Dr. Duane Benton owns stock in and receives retirement benefits from Abbott Labs; and Dr. Dennis Bier's employer is the recipient of a 2002 nutritional research grant from Bristol Myers Squibb; no funds go to him personally or for his personal research.

Dr. Garza then reintroduced Dr. Taylor. Dr. Taylor presented regulatory background information pertinent to the task force's discussions, explaining Section 409 and 412 of the Federal Food, Drug and Cosmetic Act.

Thanking Dr. Taylor, Dr. Garza welcomed the committee, guests, and staff. He noted the full agenda and asked if task force members had any questions regarding the agenda or the questions.

Dr. Thureen proposed that the questions regarding the metrics for evaluating normal physical growth be addressed separately for term and preterm infants. There were no objections. Preterm was defined to include low, very low, and extremely low birth weight infants. Dr. Anderson queried if the difference between the first bullet and the asterisk on the last bullet in question 6 meant that the last bullet was a study in which new ingredients would be used independent of their inclusion in a specific formula matrix, making the ingredient a generic ingredient that could be added to any formula. Dr. Walker replied that his interpretation was correct. Dr. Garza clarified that the asterisked bullet was for a more generalized evaluation of a generic ingredient than for a specific formula.

Dr. Garza then proposed that the task force take approximately 30 minutes for each of the six questions and approximately 120 minutes for the seventh question. He indicated that if discussions related to the first six questions warranted additional time, time would be allocated accordingly at the end of the initial review of all seven questions. Time for such discussions would be designated to specific questions from unused assigned times. He noted that he did not want to shortchange any of the seven questions the FDA assigned to the group Task force members agreed to this preliminary schedule. Dr. Garza also asked task force members to spend time Monday evening thinking about a change to an existing formula and the principles and criteria they would use to determine whether or not a clinical study would be warranted, noting that thinking about this in advance of the discussion of question 7 would assist the discussion. Task force members agreed to do this homework.

Dr. Garza then proposed that the guest speakers each be given 15 minutes for their presentations followed by a period for questions from task force members. He then introduced each of the guest speakers in the following order: Dr. Cameron Chumlea, Dr. Kenneth Ellis, Dr. Edward Frongillo, Dr. Lawrence Grummer-Strawn, Dr. Samuel Fomon, Dr. Jon Tyson, Dr. Edward Frongillo, Dr. Duane Benton, and Dr. Dennis Bier.

Following presentations by the guest speakers, Dr. Garza invited the guest speakers to move to the front of the room and asked task force members if they had questions for the speakers. After the question and answer period, Dr. Garza initiated discussion of the questions before the task force.

The session was adjourned at 5:50 p.m.

Dr. Garza reconvened the task force at 8:15 a.m., Tuesday, November 19, 2002. After housekeeping issues, Dr. Garza introduced the four public speakers: Jose M. Saavedra, M.D., Medical & Scientific Director, Nutrition Division, Nestle USA; Jon A. Vanderhoof, M.D., Vice President, Global Medical Affairs, Mead Johnson Nutritionals, A Bristol-Myers Squibb Company; Russell Merritt, M.D., Ph.D., Director, Medical Affairs, Nutritional, Ross Products Division of Abbot Laboratories; Barbara Heiser, R.N., B.S.N., IBCLC, Executive Director of the National Alliance for Breastfeeding Advocacy, Inc. (NABA).

Dr. Garza invited the three public speakers representing the International Formula Council to the front of the room—Dr. Saavedra, Dr. Vanderhoof, and Dr. Merritt. Task force members questioned the speakers regarding their presentations. Dr. Garza then gave task force members

the opportunity to ask the guest speakers questions prior to resuming discussion on the questions addressed to the group by FDA.

Discussion on the remaining questions resumed and continued until the task force reached consensus on all seven questions.

Dr. Garza adjourned the meeting at 12:56 p.m.

FDA Presentations

Dr. Taylor summarized the charge to the task force as follows:

1. Criteria for adequate evaluation of normal physical growth during the first six months as an integrative indicator of the nutritional adequacy of new infant formulas
   - Methods available to measure physical growth
   - Tools available to evaluate the data (bioequivalence and normative references)
   - Usefulness of different types of comparisons

2. Types of changes in infant formulas that should be accompanied by a clinical study in order to provide assurance that a new infant formula supports normal physical growth
   - Interactions affecting potential bioactivity or bioavailablility among individual formula components
   - Interactions of the matrix components with the absorptive surfaces or milieu of the infant

In her presentation on regulatory context, Dr. Taylor discussed the rationale for the infant formula legislation. The Infant Formula Act was passed in 1980, creating Section 412 of the Federal Food, Drug and Cosmetic Act. In 1986 Congress passed amendments to the Infant Formula Act, which increased FDA's capacity to provide assurances of protection for infants fed infant formula. Special provisions for infant formula were warranted because formula is the sole source of nutrition for a vulnerable population. The intent of the legislation was for infant formula to be safe and contain all nutrients required to support growth and health, and should provide them in a bioavailable form.

The provisions of the law require manufacturers to submit notification of their intent to market a new or newly formulated formula 90 days prior to the product being made commercially available. With that notification, the manufacturer may provide whatever information it considers sufficient to assure the agency of the product's nutrient content, its compliance with GMPs and quality control, and its satisfaction of the quality factor requirements. The FDA then reviews the notification package. If the assurances are adequately provided, the agency does not object to the marketing of the formula. If assurances are not adequately provided, the FDA may object, but the manufacturer may market the formula over the agency's objections.

The quality control factors "pertain to the bioavailability of a nutrient and the maintenance of levels or potency of the nutrients during the expected shelf life of the product," according to the 1980 House Committee on Interstate and Foreign Commerce. In the simplest form, Dr. Taylor

said, quality factors are a check on the concern that once the entire product is put together, it works appropriately. Two types of quality factors apply to infant formula: nutrient specific and formulation that results in healthy normal growth.

Dr. Taylor noted that the task force's discussion might be used to inform the scientific review of the 90-day notification conducted by FDA staff, and might be considered in the current rulemaking process. Input from the task force may or may not be relevant to the rulemaking stage for the 1996 FDA proposed rule to implement parts of Section 412, including quality factors. If the input from the task force is pertinent to the rulemaking, FDA has retained the option of re-opening the comment period.

Guest Speaker Presentations

W. *Cameron Chumlea, Ph.D., Fels Professor, Department of Community Health, Wright State University School of Medicine, Lifespan Health Research Center*, discussed the most useful measures of infant growth—weight, recumbent length, and head circumference—in infants from birth to 6 months of age.

Noting that body dimensions increase more rapidly during the first six months of life than at any other time, Dr. Chumlea cited statistics that show, on average, that weight increases about 115 percent, body length about 34 percent, and head circumference about 22 percent during that period. Weight gain is most rapid in the first and second months of life, with an average rate of 1.1 to 1.2 kg/month for boys and girls at age 1 month. The rate slows to about half a kg/month for girls and boys at age 6 months. The rate of growth in recumbent length ranges from 3.5 to 3.9 cm/month for girls and boys at age 1 month, and then slows to approximately 1.8 cm/month at age 6 months.

Accurate and reliable measurements are required to assess growth, according to Dr. Chumlea, who described preferred measurement methods. Two technicians are needed, one to position the infant and take each measurement, and a recorder to help position the infant and equipment and properly record the results. It is preferable that the technicians compare measurement values to ensure that differences fall within allowed ranges. Videos describing measurements are available from NCHS and WHO. They are similar to those in the current NHANES study, NHANES III, and the WHO multicentre Growth Reference Study, as well as methods listed in the *Anthropometric Standardization Reference Manual*.

The preferable method for obtaining an infant weight is to weigh the mother and baby together and then subtract the mother's weight from the total. If the infant is weighed separately, the infant should be weighed nude, or the scale should be tarred to account for the weight of the blanket or diaper. Spring-type bathroom scales and beam balance scales are not accurate for research or clinical purposes.

Dr. Chumlea described the preferred method for obtaining recumbent length. One technician positions the infant's head against the headboard with the infant looking straight up. The other technician then positions the infant down the length of the center of the device with the shoulders

and hips perpendicular to the trunk. This technician straightens the legs and brings the footboard up against the soles of the feet. The technician at the footboard determines the infant's length after ensuring that the head remains stationary and the infant is lying flat.

With the child seated on the mother or caregiver's lap the head circumference is measured with an inelastic tape positioned just over the eyebrows and level across the front of the head, he said. The tape then is moved across the back of the head to locate the greatest circumference.

While weight, recumbent length, and head circumference are the primary measurements for assessing infant growth, Dr. Chumlea cited additional anthropometric measurements that may be useful: crown-rump length taken while the child is seated; chest circumference; limb lengths; and skin fold thickness. He noted that these measurements have a restricted utility, high measurement error, and limited suitable reference data.

Dr. Chumlea noted that BMI in infants is affected by the disproportionality of the head, which is approximately 25 percent of the body length. While the relationship of BMI with direct measures of body composition has not been established, weight for length is descriptive of the relative level of leanness or adiposity in an infant. A high percentile indicates that the infant's weight to length ratio is greater than that of an infant in a lower percentile, implying greater adiposity.

Dr. Chumlea talked about the affect of measurement errors on measurement frequency. He noted that errors are a function of the equipment, its calibration, the technicians, and the infant. Because of the small size of infants, the size of an error is relatively greater. Measurement errors can have a greater impact on measurements of infants, especially on the interpretation of increments. Dr. Chumlea endorsed the use of high-quality equipment that is maintained regularly and calibrated at the same frequency at which measurements are taken, e.g., if measurements are taken daily, the calibration should be performed daily.

Technicians also need to be trained in quality control and standardized measurement techniques, he said, noting that training requires the collection of inter- and intra-observer reliability data. This data also should be collected at one-month intervals as a minimum during the course of the study.

Dr. Chumlea recommended that at a minimum a baseline, interim, and final growth measurement be taken for an infant. The first measurement should be taken at approximately 10 to 14 days, but no earlier than 8 to 10 days, and not later than one month, so that weight loss after birth has been replaced. If weight is to be adequately measured, Dr. Chumlea recommended weight measurements at 1, 2, 4, and 6 months of age. Dr. Chumlea recommended that recumbent length and head circumference be measured at the beginning and end of the study.

Because growth that is measured at repeated visits produces increments in weight, recumbent length, and head circumference from one visit to another, the increments are records of the velocity or rate of growth per unit of time. This data can be compared to existing tables of percentiles and charts for weight, recumbent length, and head circumference that are available from birth to 12 months of age. When using incremental growth data, Dr. Chumlea said,

attention must be paid to the data collection methodology to document and keep measurement errors to a minimum. Dr. Chumlea added that it is important to have two technicians, reliability data needs to be collected, and existing increment charts should be used until WHO charts are available.

Discussion: When asked to discuss skin fold measurements, Dr. Chumlea noted that they are extremely difficult to collect in infants in this age range, that error rates are very high, and that reference data are limited. Technicians must be very careful he said, noting that this measurement is more practical in small studies.

If body fat is the reason for collecting skin fold data, he recommended the use of other methods, such as DXA, which gives data on fat, weight, and bone. Fat measurement is the most important for measuring excess growth, he said. If excess growth is suspected, additional measurements should be taken to ascertain if an infant has excess velocity and is maintaining high velocity when other infants' growth rates are slowing down. He acknowledged that DXA has limited availability.

Dr. Chumlea indicated that incremental data are the gold standard, noting that repeated measurements from the same infants give information with status value in reference to the data used, including percentile levels and distribution rates at which people grow. These measurements are most important, he said, but must take into account errors inherent to repeated measures.

Kenneth J. Ellis, Ph.D., Professor of Pediatrics, Baylor college of Medicine, USDA/ARS Children's Nutrition Research Center, Houston, presented a review of body composition assessment in early infancy. Dr. Ellis reported that there is increasing interest in the association between nutritional status during early infancy and childhood and the increased risks for adverse health effects as adults.

Taking direct measurements of the body fat in infants is difficult, he said, noting that body composition refers to tissues and organs, or the physiological systems of the body that make up body weight. Dr. Ellis explained the classic two-compartment (2-C) model for measuring body composition, which divides the body into fat mass (FM) and fat-free mass (FFM). In this simplest model, the body's water, glycogen, and protein mass make up the lean mass obtained using dual-energy, x-ray absorptiometry (DXA), while the FFM is the lean mass plus the body's mineral content.

Used since the 1950s, the FFM in the 2-C model takes into account body density and hydration. For the whole-body counter method, which measures potassium content, there are technical limitations. Underwater weighing is too difficult to be performed with infants, he said. To accurately measure total body water, the patient must swallow all of the tracer, another difficulty for infants. The test also cannot be repeated until the tracer clears, requiring a return visit. (A plasma sample was noted as the best choice for infants.) Finally, most counters are not designed for infants and are not available in a clinical setting. The benefit of the whole-body counter, however, is that it can be repeated as many times as necessary and the infant can move without affecting the results.

The 2-C model also has scientific limitations. The density of FFM in infants is not consistent, according to Dr. Ellis, and the hydration of FFM in infants changes significantly. In addition, the bone accretion rate is not constant and the ICW/ECW ratio changes in infants.

The 2-C model determines the FM by subtracting the FFM from total body weight. The major limitation to this model, Dr. Ellis said, is that the absolute error in mass units for the larger FFM is fully transferred to the smaller FM component. In the newborn infant, FM is about 13 percent to 15 percent of body weight, so an error of 3 percent for FFM becomes 17 percent when FM is calculated. He also noted the hydration content of the FFM does not remain constant during early infancy and may be altered by disease or medications.

Body composition measures normal growth, which implies an appropriate composition of the increment in body weight, according to Dr. Ellis, who quoted from the American Academy of Pediatrics Committee on Nutrition's June 1988 report: "Sequential measurements of various aspects of body composition (e.g., body water, body fat, bone mineral) have the potential of defining changes in body composition." However, in the opinion of the task force, such measurements have not yet reached the stage of precision, non-invasiveness, and convenience that would make them feasible as a part of routine clinical testing of infant formula.

The 2-C model was modified in 1988 into the 4-C model, which further breaks down the FFM in into mineral or ash, water, and protein—the three components that are of interest regardless of a patient's age. Today, absorptiometry techniques such as DXA hold the best promise, he said, describing the DXA 3-C model as the basic model for measuring anyone of any age. The FFM in the 3-C model is comprised of bone/mineral content and lean tissue mass that is non-bone and non-fat.

With DXA technology, infants can be scanned for bone measurements. Scans can be localized or whole body, and taking about three minutes for infants. Dr. Ellis noted that DXA provides good precision and accuracy for measuring bone, fat, and lean mass with a single assay at a very low exposure risk for infants. The disadvantages, he said, include the very low exposure risk, a 2D image instead of a 3D image, and scanners that are not yet optimized for infants. Different equipment can produce different results, he said, and technicians must be trained to deal with the infant population to minimize motion.

Noting that DXA technology has improved greatly since 1988 and that the technology has almost received reference status; Dr. Ellis noted that technology's accuracy is in the 3 percent to 5 percent range, which translates into minimum detectable change. He noted that he has the statistical ability to measure changes in body weight using whole-body DXA at three weeks. Measured changes in composition comparable to 3 grams per day are possible, he said, with relatively small sample sizes and those changes can be seen relatively quickly.

Discussion: In response to a question regarding how he holds infants still, Dr. Ellis said that none of the infants in his studies are ever sedated, but that the technicians have specific training that helps them work with infants to keep them quiet. He said that feeding the infants right before the measurement helps, but that it can take as long as an hour before the infant is quiet enough for

procedures. Two technicians often are needed to position infants. With the 4500A system he uses measurements usually take about three minutes. He also said that artifacts must be handled carefully and deleted from the image.

Asked how DXA has been validated for infants, Dr. Ellis replied that studies of piglets under the weight of 10 kilograms have been used. He noted that pigs are not the best model because their bones are more mineralized than infants. He also said that phantoms (mock-ups of the human body) have been built with phosphate compounds and that about 30 preterm cadavers have been studied, including neutron activation analyses to measure calcium and other minerals. Accuracy has been within about 5 percent.

Use of a common phantom throughout the length of a study is recommended, he said. He also recommended that for a multi-center study all scans be sent to a central reading site to help assure uniformity in subjective judgments.

Edward Frongillo, Ph.D., Associate Professor of Public Nutrition, Cornell University, presented an overview of the World Health Organization (WHO) growth reference. WHO began to collect data for its new growth chart after recommendations that an international growth reference be compiled that allows cross-national comparisons. In support of this research, he noted that a 1974 paper showed that growth curves between developed and developing countries are about the same, and that reference data available since the 1990s also shows that growth trends are similar for the first 12 months when comparing developing countries in Africa, Asia, and Latin America. WHO, working with a cross-national data set also have shown that girls across a number of countries, with the exception of China, showed growth curves that were close together.

The objective of the WHO study is to build a set of growth curves for all children less than 5 years of age, and then to have those growth curves adopted as the new international growth reference for assessing the growth and nutritional status of populations and individuals. For information purposes, he said the new WHO feeding recommendation is that infants should be fed exclusively on breast milk from birth to age 6 months, with breastfeeding continuing up to 2 years of age.

The issues, he said, are that the data is a descriptive versus prescriptive reference, and that maximum growth versus growth for optimal health is not necessarily the same thing. Infants may not be maximum size, but might have optimal growth when breastfed in their first year.

The study design includes multiple, geographically diverse sites. It is a longitudinal study of 300 infants per site that goes from birth to 24 months. An associated cross-sectional study will be conducted from age 18 to 71 months, with 1,400 infants followed per site. In the longitudinal component, infants are measured at birth; during four biweekly visits from 1 to 2 months of age; during 10 visits from 3 to 12 months of age; and then bimonthly up to age 2. Measurements include weight, length and head circumference. Arm circumference and skinfold measurements are also taken using the same schedule starting at 3 months.

The population criteria exclude socioeconomic constraints that would limit growth, and include low mobility, greater than 20 percent willing and able to follow the WHO feeding recommendations for breastfeeding, the existence of breastfeeding support systems, and the local presence of collaborative institutions. The individual criteria include an absence of health, environmental or economic constraints that could affect growth, maternal willingness to follow WHO feeding guidelines, mothers who are nonsmokers, and a gestational age of 37-42 weeks. Other considerations include mean birth weight, maternal height, complementary feeding, health-related behaviors, and funding issues.

The protocol was developed by an international, multi-disciplinary group of individuals in 1995-1996. The study is being run by an advisory group of senior scientists and the WHO Secretariat. The study sites are Pelotas, Brazil; Oslo, Norway; Davis, CA, USA; Muscat, Oman; Accra, Ghana; New Delhi, India. The sites were selected over time, with some starting earlier than others.

Data management is through local data entry. Then the data are sent to the WHO Human Reproductive Program. Decisions and information presentation are the responsibility of a working group on growth reference protocol, a steering committee, and an advisory group. Multiple levels of documentation have been developed, including a generic manual of operations that was adapted to each site, measurement and standardization protocols, protocol for 12-month visits, complementary feeding guidelines, protocol for assessing diet, protocol for the cross-sectional study, a plan for data management, and questionnaires and interview guidelines.

Data collection is expected to be completed in a few months, with the steering committee meeting now to determine the final method for analysis, how data will be depicted, and how it will be used. A reference is expected to be available in 2005.

Discussion: Asked if he felt the WHO growth curves would become the growth standard in the United States, Dr. Frongillo said that required a judgment about which reference set was more applicable, the WHO or other data, e.g., Iowa/Fels data. The advantage to the WHO data, he said, was that it was longitudinal and thus could serve as a basis for a velocity reference. He said that it has been acknowledged that breastfed infants show a different pattern of growth than formula fed infants. So to the extent that one wants a reference that fits that growth pattern for breastfed infants, that would be an added advantage of the new reference.

The WHO study is taking the standard measurements—weight, recumbent length, head circumference—plus skin fold and arm circumference measurements. Dr. Frongillo said the decision was made to include skin fold measurements because of the lack of reference data and the potential future usefulness of the data..

Dr. Frongillo said that a survey was done before each site was selected to ensure that the relationship between socioeconomic status and growth would not constrain growth. He did note, however, that different criteria were used to meet different conditions.

Asked about the prescriptive nature of the study, Dr. Frongillo said that he didn't expect the WHO breastfeeding guidelines to change appreciably in the near future, but that the study would

provide information needed to reconstruct a reference to conform with new feeding guidelines should they be implemented in the future.

Laurence M. Grummer-Strawn, Ph.D., Centers for Disease Control and Prevention, Division of Nutrition and Physical Activity, Maternal and Child Health Branch, presented an overview on the use of NCHS and CDC growth charts in the nutritional assessment of young infants. As background, Dr. Grummer-Strawn noted that the original NCHS charts, released in 1977, were percentile curves, and that those curves were normalized by the CDC and republished in 1987. As a result, the two charts never matched, although they were similar. NCHS was incorporated into the CDC after 1977, and has remained the active player moving the charts forward, he said.

In 2000, the CDC released revisions to the charts. The old and new charts use the same indicators—weight for age, length for age, weight for length, and head circumference for age—are sex specific, and do not separate according to parental anthropometry, race/ethnicity, infant feeding mode, attained size, percentile, and z-scores. These charts are references, not standards.

The 2000 data set includes a broader spectrum of race/ethnicity and socioeconomic status, increased representation of breastfed infants, a pooling of several datasets, changes in smoothing techniques, z-scores that match percentiles, length that extends to 45 cm instead of 49 cm, accessibility of the 3rd and 97th centiles on the clinical charts, and the exclusion of very low birth weight babies.

Data sources for the 2000 curves include the Missouri and Wisconsin length data, National Natality, PED/NESS, Fels, NHANES III (primary data source for 2 to 6 months of age), and NHANES II (primary data source for age 6 months and older). Data sources for the 1977 curves included data from the Fels Research Institute in Yellow Springs, OH, which consisted mostly of Caucasian, middle-class, primarily formula-fed infants, and a longitudinal follow-up study from 1929 through 1975 of 867 infants measured at birth, and again when they were 1, 3, and 5 months of age.

Smoothing the curves across age was done by combining the NHANES data with other data sets that impacted the curves at different ages. Curves for children under 6 months of age were anchored by the NHANES III data to minimize the effect on the curves of the NHANES II data, which were primarily for older infants. Data from two states that routinely collect length data at birth, Missouri and Wisconsin, were analyzed and found to be relatively the same. This data was extrapolated to the NHANES data and the curves connected. A third data set also was used, the data for which was gathered from the CDC Pediatric Nutrition Surveillance System during the first visit to a pediatric clinic up to 5 months of age. Head circumference data relies on Fels data that has been connected to the NHANES data.

The NCHS 1977 curves were normalized using estimated standard deviations above and below the median, a z-score that equaled the measure minus the median divided by the standard deviation, and normalized curves distinct from percentile curves.

The CDC 2000 data incorporated fractional polynomials used in previous growth studies, with weight for length used in the 5th degree polynomial. The data was transformed by a Box-Cox power transformation to make it symmetrical, normalizing the curve.

The CDC 2000 data represents a mixture of breastfed and formula fed infants, but primarily formula-fed infants: with the ratio 50:50 to breastfed infants up to 2 months of age, and less than 10 percent of the infants exclusively breastfed by 6 months of age. In the NCHS 1977 study, virtually all participants were formula-fed infants.

Comparing the old curves to the new ones, Dr. Grummer-Strawn said that breastfed infants have been shown to grow more slowly after about 4 months of age. He noted that WHO has a pooled dataset from six studies of exclusively breastfed children and that when comparing them to the NCHS 1977 and CDC 2000 height-for-age, weight-for-height, weight-for-age curves there was not much difference below the 10th percentile. He also noted that there was the new height-for-age curves were less steep than the older curves.

In conclusion, Dr. Grummer-Strawn said that the interpretation of the CDC 2000 reference data is not widely different than the older reference data. He noted that the WHO reference data under development might provide more substantive change in the interpretation, which may lead to discussions about different ways to think about growth.

Discussion: Asked if the present CDC reference data could be used to determine a pattern of growth for assessing the nutrition management of infants, Dr. Grummer-Strawn said that it would be different because the CDC data is cross-sectional instead of longitudinal and that it does not include a large sample size so there is a fair amount of noise. He noted that the data for the first three months of life might have been smoothed differently if more data had been available.

Dr. Grummer-Strawn noted that the CDC 2000 data is helpful when comparing the growth of a child in the United States to the growth of other U.S. children over the past 10 years. The formula generates a pattern of growth regardless of whether the infant was breastfed or formula-fed. If that is the intent of the comparison, he said, the CDC data are okay. For a more prescriptive comparison that a formula produces a pattern of growth that is most healthy, he said, the CDC data might not be the best.

Samuel J. Fomon, M.D., University of Iowa, presented an overview of the Iowa and Iowa-Fels data, and references for evaluating infant formulas in terms of gains in weight and length. The Iowa and Iowa-Fels data includes primarily Caucasian, term infants. Data is from meticulous measurements, including recumbent length and weight at age three months.

Specifically, the Iowa reference set includes data from formula-fed infants 8 days to 112 days of age. Measurements were taken at 8, 14, 28, 42, 56, 84, and 112 days. The study of 380 male and 340 female infants was fully longitudinal, with no missing data points. In addition, the Iowa data includes a longitudinal study of 203 male and 216 female breastfed infants with the same measurement points and no missing data points. A second study of 165 male and 188 female

formula-fed infants aged 112 to 196 days includes a subsample of 63 male and 74 female infants, who were followed from 8 to 196 days.

The Fels data includes 240 male and 236 female infants representing a wide socioeconomic status. Data points are at ages 1, 3, 6, 9, 12, 18, and 24 months, with few measurements during the first three months. The data is obtained by mathematic curve fitting.

The Iowa-Fels data includes the Iowa data to age 3 months, Iowa-Fels data from age 3 to 6 months, and Fels data from 6 to 24 months of age. The Iowa data include 580 male and 562 female infants; the Iowa-Fels data include 298 male and 298 female infants; and the Fels data include 233 male and 224 female infants.

Dr. Fomon noted that for individual evaluations the greatest interest lies in the outlying centiles, the early detection of growth abnormalities, that weight gain is more important than length gain, and that data is needed for at least the first two years.

Reference data for formula evaluation, he said, should have the following characteristics:

- The most sensitive evaluation of longitudinal growth of a study cohort requires longitudinal reference data.
- Gains in weight and length are more rapid in male infants than in female infants; thus a formula may be adequate for females but not for males.
- The study interval must include at least part of the neonatal growth spurt (8-42 days of age); formula may be adequate for older infants but not for younger infants.
- Length data as well as weight data is needed. Data show that male infants on low-protein infant formula show a p value of <0.05 compared to reference population at 8 and 112 days.
- The reference population must be similar to the study cohort, with matches in infant health, term, and possibly ethnicity.
- The duration of the study should be at least 84 days. Studies from 8 to 112 days or 14 to 112 days are preferable because most formula-fed infants have regained their birth weight by day 8. Studies of infants from 28 to 112 days of age probably are acceptable, according to Dr. Fomon, who noted that the latter study length greatly aids recruiting.

In his recommendations, Dr. Fomon noted that the size of the data set is not relevant, and that data for infants more than 6 months of age are not relevant. The cutoff age actually could be 4 months of age, he said. Noting that the cohort should match the study group, he said breastfed babies are not relevant to studies of formula-fed infants, and that term studies should apply only to term infants.

Discussion: Noting that the Iowa-Fels data are longitudinal and the CDC data are cross-sectional, Dr. Fomon said he did not consider the CDC study to be relevant to formula-fed infants. He also said he preferred to call the NCHS charts size charts instead of growth charts.

When asked if a study requires concurrent contemporary control group, Dr. Fomon said there are circumstances under which concurrent control is essential. If you have good reference data for

comparison, however, he said it is not necessary to have concurrent controls, which greatly reduces the numbers of infants needed for study. From a practical standpoint, he said, the study of new formulas does not require a concurrent control unless specific aspects of the new formulation suggest otherwise; in most instances it should be sufficient to compare the new formula to the old.

Jon Tyson, M.D., M.P.H., University of Texas - Houston Medical School, presented an evaluation of the early growth of preterm infants, including growth rate and health evaluations, from his perspective as a neonatologist and epidemiologist. Early growth, he said, should not be evaluated in isolation from short- and longer-term health and development. He noted that growth curves published by the NICHD Neonatal Research Network are based on birth weight, length, and bent arm circumference. He referred to the Network's website—http://neonatal.rti.org—where an infant's measurements at birth can be entered to generate expected growth curves for that infant.

Dr. Tyson said the measurements were useful tools for assessing the growth patterns of individual infants, but noted that there are differing meanings for "normal values":

- Values that are expected or typical (typical values)
- Values that are not associated with adverse outcomes (low-risk values)
- Values that do not cause adverse outcomes (healthy or optimal values)
- Values for which intervention has not been demonstrated to be beneficial (values that do not warrant treatment)

Growth curves can be useful clinical tools for assessing the growth patterns of individual infants, he said, though it is not a satisfactory basis on which to assess the growth of preterm infants fed a new formula because many factors can compromise the validity and generalizability of observational studies for assessing infants subsequently fed a new formula, including:

- Measurement errors
- Effects of parenteral and enteral nutrition on growth rates
- Temporal changes in care and outcome since the observational studies were conducted
- Ways in which different centers select study participants, including biases that affect the referral of high-risk mothers and infants
- Inter- and intracenter differences in obstetric practice, including the use of steroids and other medications and interventions, aggressiveness of care for extremely small or premature infants, and the routine feeding and care of preterm infants

Dr. Tyson referred to randomized trials as the gold standard, and concurrent cohort studies carefully done, as the silver standard. Historical controls, he said, are the bronze standard. Noting that it is important to avoid all opportunities for bias, particularly when there is a potential financial interest, he recommended randomized assignment to the new formula or conventional preterm formula, blinded caregivers and evaluators, well-standardized assessments, effective procedures to avoid attrition, predefined stopping rules, and an adequate sample size and statistical power.

The feasibility of these studies, he said, is increased with the use of a management trial. Also known as an effectiveness trail, this is a simple study protocol that is appropriate for addressing questions such as the effect of infant formula on growth, development, or health under routine clinical circumstances. The patient/cost in management trials usually is lower than in traditional explanatory, or efficacy, trials, he said, which are designed to assess an intervention under ideal or restricted circumstances or to define the mechanism of its effects.

A representative sample of the study population should be enrolled in the study. It is inappropriate, he said, to exclude small for gestational age (SGA) or sick infants, twins, or others who are members of the target population. It would be appropriate, he said, to exclude, for example, infants with major congenital anomalies, or overt non-bacterial infections.

It is highly desirable, he said, to include infants who receive mother's milk and randomize them to feedings of new or conventional formula when an adequate supply of mother's milk is not available. The inclusion of these infants increases the generalizability of the results of the trial to the large percentage of infants fed some of their mother's milk, and helps compare the effects of formula and human milk feedings for preterm infants, with adjustment to other confounders, he said.

Regarding the assessments that should be performed, Dr. Tyson said the obvious assessments are change in weight, length, and head circumference; skin fold thickness and mid-arm circumference are two other measurements that might also be used. Body composition or biochemical, physiologic, or functional variables also might be needed for all or a sample of infants, he said, depending on the composition of the new formula, its differences from conventional formulas, and the anticipated benefits and potential hazards.

Follow-up assessments of health and developmental and neurologic status are warranted, he said, based on concerns about the effect of early nutrition and growth on the developing brain, and to better define the optimal growth rate for preterm infants. Follow-up assessments also should consider the effect of interventions and direct or indirect mechanisms that may need to be excluded because they have an adverse effect on development. It would be highly desirable, he said, to assess growth to no less than 18 to 24 months adjusted age (past term), which allows identification of potential late developmental or neurological deficits.

Dr. Tyson said that because there are no longitudinal assessments of intrauterine length or weight rates, the optimal rates are estimated from grids relating body weight and length to gestational age. The goal, he said, should be outcomes that achieve postnatal growth rates postnatal comparable to intrauterine growth, which now is estimated to be about 1.1 cm per week in length. Once preterm infants begin full feeding, they begin to catch up, he said, with postnatal growth rates similar to the estimated intrauterine growth rates at the same postconceptional age. Though many remain below the 10^{th} percentile when compared to full-term infants, if weight, length, head circumference, and body proportions are similar to that of term infants at the same adjusted age, no adverse effects on health or development through 18 months of age have been identified.

It is not easy to determine how many infants are needed for a study to assess a new formula, he said, noting that ideal studies would be so large that the formula manufacturers and NIH would not be willing to fund them.

When a new formula is considered sufficiently promising to justify a management trail, there may be strong *a priori* reasons to believe it would be preferential to a conventional formula. For example, he said, a new formula may contain a nutrient that promotes optimal growth, health, or development that is normally received from the mother before birth and in human milk after birth, but that has not been available in conventional formulas. Even in this situation, he said, it would be important to exclude the possibility of unexpected adverse effects, which at the present time include:

- An absolute increase in necrotizing enterocolitis (NEC) of 3 percent to 7 percent or more. NEC is a serious, often life-threatening neonatal illness. A 3-percent increase, he said, would correspond to a "number needed to harm" of 33 infants.
- A reduction in mean growth to nursery discharge or at follow up of 0.25 SD or greater among very low birth weight or extremely low birth weight infants.
- A reduction in mean developmental quotient at 18 months or later of 0.25 SD or greater, which would substantially increase the number of preterm infants with deficient or marginal IQs.
- A reduction of 0.25 SD in length or head circumference at 18 months. A modest decrease in length or head circumference, if not weight, is presumptive evidence of harm (250 g in weight, 1.25 cm in length, 0.5 cm in head circumference).

Using the conventional method for calculating sample size, Dr. Tyson said, would require 315 infants per group for 80-percent power to identify a 0.25 SD difference at 18 months assuming an attrition rate of less than 20 percent. For a 90-percent power to identify the same difference, 421 infants per group would be required. Increasing the p value is considered significant, he said, when evaluating a serious hazard. Using the p value of 0.05 for benefits and hazards is an arbitrary and not well-justified practice, he said. For serious hazards like necrotizing enterocolitis, a high p value is justified because bias often is toward finding benefits rather than looking for harms.

The appropriate p value, according to Dr. Tyson, should not depend on the cost of drawing the wrong conclusion. For a serious hazard like NEC, he suggested selecting a p value of less than 0.30 that would result in a 70 percent chance or higher, and a difference of that magnitude would not occur by chance under the null hypothesis. This higher p value would decrease the sample size requirements needed to address both benefits and hazards.

In conclusion, Dr. Tyson said that the growth of preterm infants should not be assessed in isolation from the effects on health and development. A large trial that evaluates growth, health, and development to 18 months or more is needed to assure that the benefits of any new formula outweigh any hazards in preterm babies, and to better define the effects the new formula on different growth rates.

Discussion: In response to questions, Dr. Tyson said that he would exclude preterm infants with very unusual problems that affect growth and/or development. He said that growth restricted infants should be included because they are a large part of the population, but that they should be stratified for further analyses and that include explanatory evaluations if the formula has a different effect on those babies. He said trials should be randomized, and noted that it would be cumbersome to pair any infants at birth. Ultimately, he said, the goal of the study should be to test the formula in the same way it is intended to be used in the real world.

Asked if the Neonatal Network would sponsor the large studies suggested, Dr. Tyson said that there was no reason not to propose such a study to them. There are a lot of networks throughout the world, he said, and a lot of people out there that can do a full study of multiple questions at a lower cost than a study that answers just one question. One requirement for such a study, he said, would be a predetermined acceptable ratio for the number of babies that may benefit versus those that may be harmed.

Asked if he would recommend evaluating neurological development outcome at 18 months in studies of new term formula, Dr. Tyson responded that he would like an appropriately sized study that looked at the neurodevelopment in term babies to know if new formulations have beneficial or harmful effects.

Regarding NEC, Dr. Tyson was asked if it would be appropriate to monitor adverse events against some fixed standard; for example, infant formula would be unacceptable if it produced a rate of NEC of 10 percent. Dr. Tyson responded that there are periods when the NEC rate goes up and there is no clear understanding why. He also noted that there is variability about what is called NEC and what is not, making such a standard difficult to use as a baseline.

Referring to the question can preterm babies grow too fast, Dr. Tyson said there isn't clear understanding about what is too rapid growth for preterm infants, and that the only option now is to randomize infants to different feeding regimens that produce different growth rates to see which infants turn out to have the best health and development; that might be a different answer for the really sick babies and the healthy ones.

Edward Frongillo, Ph.D., Associate Professor of Public Nutrition, Division of Nutritional Sciences, Cornell University, presented analytic issues related to the evaluation of normal physical growth as an indicator of nutritional adequacy of new infant formulas. Five issues were considered.

1. Sensitivity and usefulness of several types of comparisons. The Academy of Pediatrics Committee issues guidelines for determining physical growth when evaluating a new infant formula: weight gain over the first four months, with measurements taken at 14, 60, and 120 days; and rates of weight gain in grams per day recorded for the period 14-60 days, 60-120 days, and 14-120 days.

Possible comparison groups for infants receiving new infant formulas include infants randomized to receive a standard established infant formula, or alternatively infants whose

growth is represented in a reference, infants whose measurements are in a currently available data set, and infants whose measurements are in a historical data set, according to Dr. Frongillo.

When recruiting for randomized clinical trials, he said, the challenge is unbiased recruiting. The advantages of randomized clinical studies include theoretical close control of factors that might influence the outcome (growth); incorporation of design features to minimize known sources of potential bias, including stratification based on infant characteristics; and a probability statement justified by the randomized design as well as the statistical model. A negative is that the sample size must be twice as large as other approaches, assuming the control group contributes no sampling variability, he said.

By comparing the study group to a known, established reference data set, the sample size can be reduced to one group of infants. The negatives, he said, are that a new cohort may differ in important ways from the reference sample; the reference sample may exhibit somewhat different growth patterns from those of the new cohort; and the reference data is not free of sampling error, which must be taken into account. In addition, he said, current U.S. reference data is cross-sectional and not representative of the variability in growth increments.

2. Potential for evaluating a meaningful difference in growth increments per day. When evaluating a meaningful difference in growth increments per day, the smallest difference is the one that would be substantively important, not a difference expected or previously found, when looking at the population rather than at individual characteristics. Previous recommendations for smallest meaningful difference (SMD) from the 1988 report are 3 grams per day or 318 grams for 14-120 days. This represents about the same difference between the 25^{th} and 50^{th}, and 50^{th} and 75^{th} percentiles in increments as the Iowa and Fels data. The 318 grams is about as big as the difference between high and low altitude birth weights and is 50 percent larger than the effect of prenatal smoking on the birth weight. Data presented by Dr. Frongillo also showed that males grow at a greater rate than females. Noting that a two-tailed study is more appropriate to assess SMD, Dr. Frongillo recommended a power of 90 percent and a larger sample size to detect smaller differences.

3. Impact of transformations from raw data measurements to normalized indices. Explaining the transformation to z-scores, Dr. Frongillo said that measurements should match with reference values for age and sex. The primary purpose of z-scores is descriptive, allowing a combination of ages and sexes, and assumes that a pattern of growth in the sampled population is the same as in the reference population. (Z-score = measurement – reference median / reference standard deviation)

Transforming raw data measurements to normalized indices has application to the evaluation of new infant formulas because age adjustments would not usually be needed if measurements are taken at pre-specified ages. If needed, he said, include covariates for age or interpolate and extrapolate time series rather than converting them to z-scores. Also, males and females would typically be analyzed separately because of differences in the growth response.

4. Advantages and disadvantages of comparing with various reference data sets. Growth references are tools that provide a common basis for comparison, and the reference population

reflects growth expected for children. Dr. Frongillo, noting that reviews by NCHS/CDC and WHO in the early 1990s led to the development of new U.S. and international references, said at this point in time, he would choose the Iowa-Fels longitudinal data for descriptive purposes.

5. Circumstances favoring one type of comparison to another. Dr. Frongillo suggested considering using currently available data, if several new formulas are being tested in a short period of time. It may be efficient to sample from the same population for the whole series of studies, he said, but without having to do repeated sampling for a comparison group. The concern is that characteristics of a later sample might differ from the earlier samples.

In summary, Dr. Frongillo said that for primary analysis, the design should include a randomized, concurrent comparison group. For descriptive purposes, the attained weight for all groups at each measured age should be compared with current (2000) U.S. reference data and the rate of weight gain should be compared with the Iowa data. A sample size per group of 28 is without sufficient power for meaningful differences of even 3 grams per day; and a larger sample size is needed, even when staying with that guideline. The smallest meaningful difference, he said, should be based on the best understanding of biology and also on required regulatory, clinical, and public health decisions. The smallest meaningful different might be 2 grams per day, he said, implying a much large sample size is needed with sufficient power for obtaining meaningful differences.

Discussion: Asked if different SMDs should be determined for male and female infants, Dr. Frongillo said that females' growth rates are slightly less variable than males, with males having a more variable response to feeding loads. This suggests that a smaller sample size is needed for females, he said.

In response to the study design he described, Dr. Frongillo said that there are shifts in the mean between a new and current formula. Some differences, such as a small subset of infants that do not do well, he said, would not be caught by the study. The study would have guidelines in place, however, to indicate when interventions should be started as necessary. The study would assess the period of fastest growth—from birth to age 4 months.

Duane Benton, Ph.D., discussed product composition considerations in clinical studies. Drawing on 25 years of experience with Ross Products, a division of Abbott Laboratories, Dr. Benton called growth a combination of thousands of responses. Studies must look for the most sensitive measure of how an infant is growing to adequately assure that nothing is going wrong, he said.

Breastfed infants grow differently; therefore it does not make sense to compare breastfed infants to formula-fed infants to measure growth, according to Dr. Benton. Instead, studies must compare a new formula to a present formula that is considered safe with historical data showing this.

Studies should begin by the end of the second week of life, he said. The reason, he explained, is to be able to detect a nutrient deficiency, toxin buildup or any other serious problems in the early weeks.

Discussing the factors that would indicate a need for a study, Dr. Benton discussed how infant formula is made. Batching, the process by which the ingredients are put together, can make a difference in the stability of the nutrients and interactions between nutrients, he explained. An ingredient tested for one formula is not necessarily safe for use in others. Heat interaction also is very important because heat interacts with a number of things, he said, from how the product is put in the container to its shelf life. Changing batching processes or heat interaction is potentially a reason for a study.

Changes in protein quality also may indicate the need for a new study, according to Dr. Benton. For example, he said, replacing bovine milk with goat milk, a high-quality protein, in a formula may require a study because it is a complex mixture that might interact with other ingredients in the formula differently than bovine milk.

Food additives also should be tested for use in formula, he said, noting that only carrageenan has been adequately studied for infant formula; other additives have been tested only for adult use. Just because a product is generally recognized as safe, he said, doesn't mean that we know how it reacts metabolically in infants. He said flavoring compounds are similar to food additives, with little known about their effects on infants.

In summary, Dr. Benton listed five criteria that should be used when evaluating whether or not a clinical study is warranted:

1. Ingredients: knowledge of the chemistry and reactivity of all ingredients
2. The infants potential metabolism of new added ingredients
3. Formula processing, batching, etc.
    a. Processing: All the processes required to solubilized the ingredients, homogenize, and set the batch for sterilization or drying in the case of powdered formula; how the batching may have changed or damaged a nutrient
    b. Heat processing: Comparison of the total heat inputs that reflect the potential damage to the product
    c. Packaging: Potential for the packaging to interact with the product, leach materials in to the product, and how the package modifies the heat input
    d. Shelf-life changes: Losses of nutrients and physical or chemical changes during storage may indicate need for clinical evaluation
4. Experience the company has with well-studied infant formula to determine if any historical experience can appropriately be applied to the new formulation
5. Literature relative to any physiological effects that could possibly be elicited by the new formula; where any such effect can be projected, a physical growth study is needed.

Discussion: Asked a question about prebiotics and probiotics, Dr. Benton explained that in most cases those substances will be reactive and that manufacturers generally will have no experience with the effects they would have on infants.

In regard to determining if a formula gives optimal versus good growth, Dr. Benton replied that growth is complex. To assess optimal growth, he said, there are many things that must be evaluated, down to the cellular level, and the infants would have to be studied for a long time

before conclusions could be drawn. There is no clear picture of optimal growth, he said, adding that Ross Laboratories would be interested in participating in studies to determine the optimal rate of growth.

Asked if formula-fed infants are too fat or too big, Dr. Benton discussed Dr. Fomon's studies on different caloric intakes, noting that 11 calories to the ounce is standard. Several factors influence caloric intake, he said, including the calories to the ounce in the formula and the way mothers feed their infants. One way to study this, he suggested, might be to reduce protein content at feedings.

Dennis M. Bier, M.D., USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX, presented five guidelines and six criteria for deciding when a clinical growth study is necessary. His principles and criteria assume that today's formulas contain all the known essential nutrients, and that nutrient deficiencies do not occur as a consequence of formula ingestion, per se. In addition, new additions to formula are likely to be nutritive and non-nutritive substances added for other purposes, he said.

Guideline #1: Measurement of growth is an integral and necessary component of every clinical study. Growth is a fundamental factor of adequacy, he said, because growth occurs only when all other maintenance nutritional needs are met. Growth measurements also are advantageous because of their simplicity, accuracy, non-invasiveness, lack of specificity, and because they are a generic biomarker of unanticipated detrimental effects.

Guideline #2: Infant growth studies require measurement of both length and weight. Factors controlling linear growth, including genetics and gene regulatory molecules, hormones, and growth factors, are different from those responsible for body weight gain, he explained. It would be helpful for research purposes to measure body composition, he said, but it is not necessary because relationships between infant body contents and childhood and adult outcomes have not been established.

Guideline #3: Animal studies are never a sufficient substitute for human growth studies. Animal studies are necessary to show proof-of-principle and pre-clinical assessment of safety, he said. Species differences preclude the use of animal growth to establish normal human growth.

Guideline #4: The presence of a substance in human milk is not, per se, sufficient reason to eliminate the need for a human growth study. Noting that some compounds in human milk are detrimental, such as dioxins and PCBs, Dr. Bier explained that other non-nutritive components are not well characterized. Components of breast milk, such as growth factors in colostrum, he said, have profound biological activity, but their role in human growth and development are poorly understood.

Guideline #5: Data from post-marketing experiences in other countries are not a sufficient substitute for a pre-market growth study. Often the data is anecdotal, with validity dependent on the reporting of adverse events, he said. Also the post-marketing surveillance data is unlikely to detect subtle growth effects.

Criterion #1: An infant growth study is required for all substances added to influence growth. Any claim of an effect or benefit requires demonstration of the claimed effect or benefit, he said.

Criterion #2: An infant growth study is required for macronutrients and other compounds known to affect hormones, growth factors, genes, or metabolites that regulate growth.

Criterion #3: An infant growth study is required for formula changes that result in nutrient levels outside established ranges. There currently are 29 identified nutrients, he said, with maximum levels for 10 of them required in infant formulas according to established consensus ranges. It is time to start establishing a new set of consensus ranges for term and preterm infant formulas, including exempt infant formulas.

Criterion #4: The addition of an entirely new compound to formulas requires a clinical growth study. All new compounds, regardless of whether they are present in human milk, should be tested with a clinical growth study, according to Dr. Bier.

Criterion #5: All entirely new formulas require a growth study, including ones that are not modifications of products already marketed in the United States. The study would be needed to establish proof of nutritional equivalency with, or superiority to, marketed formulas known to support normal infant growth. The criterion applies to ingredients, ingredient sources, processing and production variables, matrix interactions, and absorption and bioavailability. This criterion coincides with the 1988 American Academy of Pediatrics task force position, Dr. Bier said.

Criterion #6: Formula alterations likely to affect GI function or nutrient bioavailability require a growth study. Alterations include changes to the formula matrix; changes to the macronutrient composition, such as fatty acids; enterocyte functions and receptor interactions; and gastrointestinal flora and motility.

Discussion: Asked if assessments of body composition are worthwhile, Dr. Bier said that such measurements would be appropriate even though there are little comparable data. Its purpose now would be to gather data to understand how body composition relates to growth.

Formula-fed infants should be compared to other formula-fed infants and not breastfed infants, Dr. Bier said, adding that the criterion for such studies becomes the growth pattern of the existing formula chosen for comparison. Asked what besides growth should be assessed to determine the adequacy of a new formula, Dr. Bier said that neurodevelopment is very important and would be the next measurement he would consider.

In summary, Dr. Bier said there are few examples where a change in formulation would not necessitate a growth study, though there are gray areas that might be open to interpretation depending on previous testing and the manufacturer's internal data. It is simplest, he said, to say any change requires study, but that is not a practical requirement.

Questions Posed to the Seven Guest Speakers

Task force members were given the opportunity to question all the guest speakers on Monday afternoon and Tuesday morning. The discussion below summarizes both sessions.

Dr. Anderson posed the following question: Imagine I have taken a marketed infant formula and added to it a new substance. The only information I have, beyond safety, is that in a clinical growth study the children who were measured at 14 days were distributed across the median of the CDC standard, and at 1, 2, 4, and 6 months, all were distributed across the median, with 2.5 percent above the 97th percentile and 2.5 percent below the 3rd percentile. Why should that formula not be approved for marketing?

Dr. Fomon responded that ingredients are to be added for some purpose. The results of this hypothetical study did not show that the purpose was achieved, so there is no point in adding the ingredient—or going forward with the formula. To clarify, Dr. Anderson added that if a new ingredient didn't have a desired effect then it would not be appropriate for marketing, and Dr. Fomon stated that was the case.

Dr. Grummer-Strawn reiterated that insufficient evidence was provided on the characteristics of the new formula. Additional information would be needed, he said, including comparisons with an appropriate reference that would aid in interpreting growth along the 50th percentile.

Dr. Frongillo addressed the reference criteria, stating the concern about what is an appropriate reference and if existing reference data may be used as comparators.

Dr. Bier asked for the benefit to risk ratio. Without knowing the benefits, then the risk component is infinite, he said, noting that until the benefit is shown to outweigh the risk, it is inappropriate to take the risk.

Dr. Benton summarized by saying that the industry would not want to add an ingredient that isn't there for some purpose.

Dr. Stallings asked what would be the optimal pattern of growth and how would it be determined. Is the optimal growth of breastfed infants for the first four months the place to start for comparison?

Dr. Chumlea responded that it could be a starting place as with the NHANES III data, which gives status information. He pointed out, however, that if that were the starting place, the problem of obesity would not have been identified. What the data would provide is what is happening currently, with that data not necessarily optimal. To get the information, he said he would create a study that sufficiently represented children of many cultures and genders in a multi-center design to collect status and rate of growth from birth to 6 months of age.

Dr. Fomon, acknowledging that breastfed and formula-fed infants' growth historically is the same at 14 days, said the breastfed cohort could be used for comparison. For a four-month study, he said, the rules have to be that breastfed infants receive only breast milk and formula-fed infants receive only formula between 14 and 112 days of age.

Dr. Grummer-Strawn replied that a prescriptive reference is needed. Any formula that moves toward breast milk is the right formula, he said, noting that research needs to work with the best data available today by studying the population that is feeding the best way today. He added that 50 years from now optimal might be defined differently.

Dr. Bier noted that in the past breastfed infants were the standard, but it has been shown that formula actually can grow babies as well and as healthy.

Dr. Frongillo said that if nutrients are adequate in formula, then is the question about growth itself or about how formula will impact the infant in the future. He acknowledged that from 1 to 4 months of age average growth is about the same for breastfed and formula-fed infants, and then the growth rates diverge, coming back together again at about 24 months of age. There have been issues raised about discrepancies in these patterns, he said, noting that the Iowa data may be able to provide more information about these issues.

Dr. Garza asked: Given the fact that we don't have long-term data to assess the functional consequence of different growth patterns in the first year, should we assume that growth differences, until proven otherwise, are or are not significant, to long-term health at 7, 15, and 30 years of age?

Dr. Frongillo noted there are two aspects to the question, formula and breastfeeding, given there are differences in the formulas infants receive.

Dr. Benton stated that breastfeeding is the interaction between the mother and the infant, with consumption determined in the process. He stated that he is perplexed by the implication that growth patterns have something to do with nutrition.

Dr. Garza followed up by asking if feeding behaviors, or the interactions between mothers and infants, are consistent cross-culturally.

Dr. Fomon said that the consequences of more or less rapid growth during the first four months are unknown and what we have is the comparison of how they grow in the first 4 months. Little in literature, he said, tells how growth in the first four months relates to later in life.

Dr. Garza asked how useful body composition would be in assessing differences and similarities between breastfed and formula-fed infants.

Dr. Fomon replied that it would give very little information.

Dr. Frongillo agreed that body composition would not necessarily bring substantial information to the data, adding that weight and length data are most important. He also noted that human milk has a biological function so hormonal substances also may play a role.

Dr. Chumlea said that 15 years ago the sense was that fat babies became fat adults. The Fels data do not support this, however. Measurements at 5 or 6 years of age were more predicative.

Preliminary analysis shows that nutrition from 8-12 years of age and 12-16 years of age does have a predictive relation to bone density, he said.

Dr. Ellis added that environmental and other issues affect what happens later in life.

Dr. Sigman-Grant asked if there were any differences in head circumference or body organ weight between breastfed and formula-fed infants.

Dr. Fomon said that head circumference is proportional to length, so the value of head circumference is in detecting discrepancies between gain in length and gain in head circumference. No one really knows about body organ weight, he said.

Dr. Anderson posed another hypothetical situation: An infant formula is developed that contains GRAS long-chain fatty acids with no growth studies, but evidence of benefit to neurologic development. Would you proceed without a growth study?

Dr. Bier replied that a study would be required because fatty acids have specific effects on metabolism that are known and must be tested.

Dr. Stallings asked the speakers to elaborate on the differences in growth rates between breastfed and formula-fed infants in the first four months of life.

Dr. Frongillo calculated rates for New Finland and Iowa data. From 8-42 days of age there is no difference for male and female infants whether breastfed or formula-fed. From 42-112 days of age, there are differences of about 3 grams per day in weight and .07 mm per day in length, with formula-fed infants growing faster. Essentially, he said, formula-fed male infants grew 3 grams per day and formula-fed female infants grew 2 grams per day. The New Finland data showed that from 0 to 2 months of age breastfed infants, both male and female, grew 3 grams per day faster than formula-fed infants. From 2-4 months of age, formula-fed male infants grew at an average rate of 6 grams per day, while formula-fed female infants grew 3.5 grams per day, rates faster than breastfed infants. The rate was similar, he said for growth from 4-6 months of age. The Iowa data was from 1968 to 1987 and the New Finland data was collected in the early 1990s.

Public Speakers

Jose M. Saavedra, M.D., Medical Director, Nutrition Division, Nestle USA, discussed the industry's current analysis and documentation process for new infant formulas. Industry does not introduce a new infant formula, he said, unless there is a specific benefit, and does notify FDA of all major and minor changes that may affect a new formula's nutritional adequacy. Only if criteria for minor or major changes are not applicable is the FDA not notified.

Clinical trials are conducted to assure the nutritional adequacy of a new formula and to show that it supports normal growth. Clinical trials should be done if they can reasonably predict that change will have an impact on growth, he said. Clinical trials should not be done if they are

redundant, unnecessary, or unethical. The decision whether or not to conduct a clinical trial is based on specific, reasonable, and conservative assessment and evaluation of the change in a process that is transparent between industry and the FDA.

Dr. Saavedra provided the decision tree used by U.S. formula manufacturers to document the nutritional adequacy of a new or changed infant formula. The decision process, he said, considers all aspects of the formula matrix and manufacturing process to determine if a study is required.

A major change to a formula is based on three criteria, he said. A formula is considered to have a major change if it is an entirely new formula by a manufacturer that has not made infant formula previously in the United States; it is a change in a current formula where the manufacturer's experience or theory would predict a possible significant adverse impact on the level of nutrients or the bioavailability of nutrients; or a change that causes an infant formula to differ fundamentally in processing or composition from any previous formula produced by a current U.S. manufacturer. If a manufacturer wants to market a formula with a major change, it provides convincing documentation that demonstrates that the formula will support normal growth, the nature of the change and supporting scientific rationale, and supportive data to the FDA.

Dr. Saavedra noted that industry tries to stay scientifically current and does not wait for guidelines to be updated before making what it considers beneficial changes. Manufacturers continually look for information that might modify outcomes, he said, noting that published literature and guidelines, as well as previous product and testing experience, are used as sources of documentation. The critical factor, he said, is the manufacturer's experience and knowledge, which it brings to bear to understand the affect of the change. In addition, manufacturers conduct internal medical assessments and independent expert reviews on formulas with major changes before bringing them to market.

In the last 10 years, Dr. Saavedra said, there have been 100 minor and 150 major change submissions and 50 growth studies involving 6,000 infants. Manufacturers collectively conducted more clinical and growth studies than any other organization. In conclusion, he noted that since the Infant Formula Act of 1980, not a single nutrition-based problem has resulted from formulation changes in infant formula.

Dr. John A. Vanderhoof, M.D., Vice President, Global Medical Affairs, Mead Johnson Nutritional, Bristol-Myers Squibb, discussed the comparators and end points manufacturers use to measure the nutritional adequacy of infant formulas. Safety is a given, he said, noting that manufacturers have an excellent nutritional adequacy record and that the American Academy of Pediatrics Committee on Nutrition's guidelines have served the industry well.
Growth studies during the first four months are critical, he said, because infant formula is the sole source of nutrition when infants are most vulnerable. This also is the time when growth is very rapid and can be attributed directly to the formula. Calling weight measurements the most important endpoint, Dr. Vanderhoof, described this measurement as the most sensitive indicator of nutritional adequacy. Length is a secondary outcome, he said, because it confirms weight gain and represents growth. Head circumference is a helpful, but not mandatory growth measurement. Measurements should be differentiated by gender and addressed by covariate analysis.

Laboratory measurements and other techniques may be appropriate, he said, if ingredients in the new formulation, such as lipid blends and calcium, might specifically affect individual nutrients.

Randomized, double-blind prospective studies are the gold standard of clinical research, he said. Because growth data are inherently objective measures these studies allow for the evaluation of secondary parameters and may be used for multiple purposes. Studies should be powered to detect a mean difference of 3 grams per day weight gain as the primary outcome variable, based on AAP/CON guidelines, he said, noting that since the standard was adopted no product withdrawals have occurred because of nutritional inadequacy.

Present challenges to conducting randomized studies include the limited availability of subjects, due in part to an increased breastfeeding rate. For a study of 500 infants powered to detect smaller differences, he said, you would have to screen about 24,000 infants.

Other options include historical controls and reference data, according to Dr. Vanderhoof. Reference data minimize drift over time, he said, and historical data could power studies at a higher level while reducing the number of infants in research. He noted that manufacturers possess large volumes of data obtained over time.

In summary, he said, present criteria provide the ideal balance for formula research and infant protection. He noted that concurrent controls usually are desirable but might not be necessary in some situations and that other biochemical or body mass measurements might be indicated in specific circumstances when a particular nutrient is tested.

Russell J. Merritt, M.D., Ph.D., Medical Director, Nutritionals, Ross Products Division, Abbott Laboratories, presented the basis and context for the industry's sample growth trial protocol for healthy term infants, which reflects the industry's experience conducting studies since before the inception of the Infant Formula Act of 1980.

Recognizing that infant formula often is the sole food source for a vulnerable population, he said, the purpose of a regulatory growth study is to determine that a proposed new infant formula performs at least as well as a current commercial formula appropriate for the population under study.

Accepted clinical practices generally require the comparison of a new intervention with a current standard of practice, in this case a marketed infant formula, according to Dr. Merritt. The growth data also will be compared to a reference standard, in effect giving both a concurrent control and an historical control different roles in the growth assessment. The historical control, he said, may be as large and more closely related to the infant formula study population under study and may be an excellent surrogate for a more recognized reference group in some circumstances.

He noted that the use of an exclusively breastfed reference group (historical or concurrent) would assume that the group of formula-fed infants should grow identically to that group. No well recognized standard exists, however, for exclusively breastfed infants, he said, so there is no data showing that the growth of breastfed and formula-fed infants should be identical or that a

specific feeding regimen has unequivocally been demonstrated to be better than others from the perspective of long-term growth and body composition.

Two-sided testing, as usually performed, is less sensitive than one-sided testing, he said, because it dilutes the power. One-sided testing, on the other hand, addresses the critical question of whether the new formula performs at least as well as a current commercial formula.

At this point in research, he said, the best infant growth pattern is not known, especially for individual infants in a study. The de facto standard of 3 grams per day appears to serve us well, he said, noting that very small differences between groups are not necessarily meaningful. He pointed out that some Food Advisory Committee consultants have said it is a misuse of the available science to pretend there is an understanding of the health implications of a gram or two of weight gain per day for a short period of time in a human life, or to define an extraordinarily specific, statistically driven definition of a single rate and pattern of infant growth as the only one that is normal or even acceptable when testing an infant formula.

In conclusion, Dr. Merritt said that the context in which growth studies are conducted is not static. He noted that the nation is coming closer to the Healthy People 2010 goals for breastfeeding rates, which will reduce the number of exclusively formula-fed infants available to participate in clinical trials, and that evolving ethical standards might further limit the types of studies acceptable in pediatric subjects. These changes, he said, may make it increasingly difficult to complete even the growth study protocol the industry currently uses. Thus, rather than moving to more restrictive protocols with theoretically more infants, other approaches regarding study participation, eligibility, and/or fewer subjects for the use of historic references may need to be considered in the future.

Discussion with the three public speakers representing the International Formula Council: Asked to comment further on one-tailed versus two-tailed studies, Dr. Merritt said that the historical context has been to ensure nutritional adequacy. In that context, in study designs where all the power is on the lower half there is greater sensitivity to detecting problems on the low side. If the power is on the high side, then the comparison generally is to existing formula and against a reference standard. When data are compared to an existing standard there usually is a concurrence of the two answers. If not, he said, manufacturers will do additional thinking and assessment. The comparison to historical standards is used as a crosscheck.

The speakers were asked if there had ever been a major change that did not require a clinical study. Dr. Saavedra replied that there had been a number of such changes. For example, he said, if an ingredient is GRAS then it already has been studied for its impact on other ingredients. The manufacturer knows all the interactions that could happen as a result of the change, so the formula would not change in providing nutritional adequacy and a study would not be required. Each manufacturer knows its matrices and the components to the extent that it can determine the affects of certain changes. In the case of a soy product introduced into a formula, for example, Dr. Saavedra said that if industry were to change a formula's protein source to one that had not been used before, then a study would be needed.

Asked if it is reasonable to assume that clinical trials are required only when there is a reasonable suspicion that there will be a significant adverse impact, Dr. Vanderhoof replied that if there were a reasonable suspicion of adverse impact on growth, the formula would not be tested. Manufacturers would only want to test a formula clinically, if the company was quite certain that it was nutritionally adequate.

The standard of 3 grams per day was discussed, particularly in regard to the possible need to lower the standard. Dr. Vanderhoof pointed out that it is an arbitrary number that strikes a reasonable balance between the study and control groups and allows study groups to be adequately powered. Dr. Merritt noted that the standard has protected the public health historically and that any changes would require time and additional studies to support. Dr. Saavedra summarized, saying that the standard strikes a balance between what is practical, doable, and beneficial.

The speakers then were asked to define normal growth. Normal growth is growth based on historical data—experience associated with reasonable health, according to Dr. Merritt.

Asked if the growth pattern of breastfed infants should be used as the default standard for normal growth, Dr. Vanderhoof replied that breastfed children who fail to thrive are switched to formula and often begin to grow rapidly. A fair number of breastfed infants do not get the number of calories they need, he said, because breastfeeding is a different process that is dependent on cues for baby and mother to determine when the baby has had enough to eat. These factors, he said, may influence differences in weight gain and nutritional factors. If studies try to replicate breastfeeding by changing the nutrient mix in formulas, the result might be that formula-fed infants are deprived of essential nutrients. The only way to replicate formula feeding with breastfeeding would be to make formula with a caloric density lower than breast milk, he said, noting that such a formula would not meet current standards.

The speakers were asked to comment on the reporting of adverse events and about how complaints are handled. Dr. Merritt said that each manufacturer has a process for reporting complaints in accordance with the Infant Formula Act. The complaints are examined by the FDA on an annual basis and manufacturers review them at regular intervals to ensure safeguards are in place.

Asked about the need for independent monitoring board for clinical studies, Dr. Merritt said that there is enough experience and guidance on how studies are conducted, particularly for most changes in formulas, that independent monitors would be extreme oversight. An external advisory board, however, might be indicated for special situations, e.g., when special populations are targeted or if one is making novel interventions, he said.

Dr. Saavedra was asked to describe the preclinical studies, in vivo and in vitro. He responded that in vitro studies are done on ingredients to measure the stability of the components and whether they remain the way they are supposed to, chemically and structurally, and if they interact with other ingredients. These studies, he said, provide reassurance that there is no nutritional inadequacy before clinical studies begin. In vivo pre-clinical studies are usually animal based, he said.

Barbara Heiser, R.N., Executive Director, National Alliance for Breastfeeding Advocacy, described exclusive breastfeeding as the optimal feeding standard for the first six months of life and called for it to become the standard for what is best for babies. In addition to the standard growth measurements of weight, length, and head circumference, she requested that brain and neurodevelopment be added to the assessments that are routinely monitored to obtain a broad picture of infant health.

Ms. Heiser described the trust the public has in the FDA to assure such high-quality, safe infant formula, noting that many mothers believe formulas go through the same testing and approval procedures as drugs. She pointed out that marketing claims are powerful tools that may influence mothers' thinking, and that they often cause confusion about the benefits of the elements in formula with those in breast milk. She also questioned the use of post-market surveillance data and what is being done with the information.

In summary, she said mothers expect safety from the government and expect great products from the companies. She encouraged the task force to find out what is best for babies on infant formulas.

Discussion: In her presentation, Ms. Heiser spoke of anecdotal information on explosive diarrhea as a result of formula feeding. Asked to elaborate on this information, she explained that her organization and nurses encourage mothers to report this information via MedWatch and to the manufacturers, and that her organization has begun questioning health professionals in hospitals about what they are seeing.

Asked to elaborate on the public's misunderstanding that formula is tested and approved similarly to drugs, Ms. Heiser said that health care providers treat formula manufacturers more like drug vendors, especially in hospitals. She also noted that mothers are told not to give regular milk, so formula is considered different and more important.

Discussion of Seven Questions

Task force members began their discussion of the seven questions Monday afternoon after hearing from the guest speakers. Discussion continued on Tuesday morning following the public speakers. Below is a summary of those discussions.

Question 1: Considering the values and merits individually, and in combination, please group the following metrics in terms of their clinical usefulness as endpoints for assessing normal physical growth.

- *Body weight,*
- *Recumbent length,*
- *Head circumference,*
- *Skin fold thickness,*
- *Bioelectrical impedance,*

?? *Stable isotope, dual energy x-ray absorptiometry, or*
?? *Other physical body measurements or body composition measurements*

Two discussions were held on this question, one concerning term infants and another concerning preterm infants. Dr. Garza, with the permission of the task force members, asked that each metric be designated as extremely useful, of moderate use, of no use, or in research and inappropriate for comment at this time.

The discussion of DXA determined that DXA needs additional study and to be available more readily before it is considered useful. It was acknowledged that there is great interest in the measurements that can be obtained using DXA and that it is an easier-to-use technology than stable isotopes. Most children's hospitals now have the technology or should have it in the near future. In addition, most level-3 nurseries in adult hospitals also have the technology. Private pediatric practices probably don't have the technology, though they should have access through the hospitals. In summary, DXA technology is available generally in middle size and larger cities.

Data obtained from DXA also were deemed to be of limited value because there are little corresponding reference data. An NIH-funded, five-center study is underway for older children. A similar longitudinal study in infants would be needed to provide reference data. Such a study would provide an opportunity for follow-up studies to begin gather data that relate data from younger ages to older ages.

Following the discussion, task force members reached consensus that body weight, recumbent length, and head circumference are the three metrics that are the most useful indicators of infant growth. Skin fold thickness was designated a metric of moderate use. Bioelectrical impedance, stable isotope, dual energy x-ray absorptiometry, and other physical body measurements or body composition measurements were deemed to be in the research stage and therefore task force members felt they were unable to comment on the effectiveness of these metrics at this time. Several task force members indicated that there was no basis for the use of bioelectrical impedance or stable isotope metrics in infants six months or younger.

Task force members also considered the question in regard to preterm infants. Their findings were the same for this population, with body weight, recumbent length, and head circumference the three metrics that are the most useful indicators of infant growth.

Question 2: Which of the above anthropometric and/or body composition measures are necessary for adequate clinical evaluation of normal physical growth of infants between birth and 6 months of age consuming new infant formula?

Referring to the discussion regarding Question #1, the task force reached consensus that body weight, recumbent length, and head circumference are necessary for adequate clinical evaluation of the normal physical growth of infants consuming new infant formula between birth and 6 months of age for term and preterm infants.

Question 3a: The metrics above can be evaluated as attained (absolute growth) or velocity (rate of change) measures. Please comment on the distinguishing values and merits of each static or variable method in the assessment of normal physical growth.

The three measurements for evaluating normal physical growth in infants 6 months of age or younger—body weight, recumbent length, and head circumference—were determined to be velocity measurements. Task force members reached a consensus that for infants in a study, baseline measurements should be taken at birth or no later than 14 days. These three measurements should also be taken at 1, 2, 4, 5, and 6 months of age.

The same guidelines were recommended for preterm infants. Recognizing that the measurements are more stressful for preterm infants than term infants, task force members offered flexibility to mitigate the stress on individual infants. More frequent measurements—every week—also were seen as necessary for monitoring preterm infant growth because of the rapid growth experienced by preterm infants. The more frequent measurements should occur while the infant is hospitalized; monthly measurements should begin after the infant is discharged through 6 months equivalency.

Question 3b: The outcomes above can also be evaluated as individual infant data or as group comparative data. Please comment on the values and merits of using individual or aggregate data in the assessment of normal physical growth.

Task force members reached consensus that it would be beneficial to see both study data on individual infants and the group comparative data. Members recommended examining individual data to identify outliers and the associated summary comments, and for additional analyses, such as distributions and clustering. The aggregate data were seen as useful for obtaining a measure of central tendency and general distribution.

Question 4: For adequate evaluation of normal physical growth, below are examples of clinically distinct reference groups.

- *Concurrent controls (concurrent data or population cohorts for demonstration of bioequivalence)*
- *Reference data used as controls (comparison with previously collected normative data for populations and subpopulations)*
- *Historical controls*
- *Other*

4a: What are the distinguishing values and merits of each type of reference group for the assessment of normal physical growth?

Dr. Garza defined reference groups as a comprehensive database with a specific compilation he stressed that reference groups are not necessarily standards. With the approval of the task force members, he separated the types of reference data and asked task force members to rank them in order of preference.

Discussion focused on concurrent controls, with the proposal that the growth standards should be determined by the growth of healthy infants receiving breast milk exclusively through at least 4 months of age. It was noted that the growth patterns of breastfed infants throughout the world are remarkably similar. While they might start at different places, when plotted against the WHO data, they appear to parallel one another. A concurrent study of breastfed infants would begin to develop data from which growth data could be obtained. The goal of the study would be to develop optimal growth patterns based on breastfed infants.

It was noted that to replicate the growth rates of breastfed infants, changes would have to be made for formulas, and that those changes would mean that the formulas would not be in compliance with current nutritional adequacy standards. The claim was made that it is reasonable to compare breastfed infants and formula-fed infants in an academic setting, but with good data on formula-fed infants over many years, it is reasonable perhaps to compare only one formula to another.

In summary, task force members agreed that concurrent, randomized control groups are essential for studies of term infants, with concurrent control groups mandatory for preterm studies. Longitudinal studies were seen as potentially beneficial, followed by cross-sectional studies, and then historical studies. Historical controls were deemed to be the least helpful because they tend to be small populations collected under protocols difficult or impossible to replicate. Temporal changes that occur also make it difficult to match historical controls to current studies. The consensus regarding historical reference data is that such data should be fairly recent because of the many factors that have affected infant growth over the past 10, 20 and 30 years. The population in the historical data also must be relevant to the current study population.

4b: *Please rank these reference groups based upon the ability of the respective control population to contribute to an assessment of normal physical growth in the population intended to consume the formula.*

Referring to the discussion of question 4a, task force members ranked the reference groups as follows: longitudinal concurrent data, then cross-sectional data, and finally historical data.

4c: What is the role of such a reference group?

Referring to the discussion of question 4a, currently available reference data were seen to have comparative value, but were not seen as a standard against which current studies should be evaluated.

Question 5: For the purpose of evaluating normal physical growth of infants fed new formulas, what criteria should appropriate infant growth reference groups meet (e.g., each or selectively, feeding history, gestational age at birth, sex, racial background, socio-economic status, other)?

?? *In comparison to the study population?*
?? *In comparison to the population intended to consume the formula?*

The discussion re-emphasized that comparisons should be done through randomized trials, with the focus on characteristics that predict the pattern of growth, such as gestational age at birth. In addition, the comparisons should match on sex and general health. Cultural and socioeconomic backgrounds are not necessary matches as long as there are no extenuating circumstances that would preclude normal growth. Matching feeding history was seen as important because infants rarely are fed solely formula or breast milk and complimentary solid foods sometimes are introduced, all of which could have an affect on the growth rate. Maternal education and birth rate were two additional factors considered for matching purposes.

Task force members determined that the comparison should be to the population intended to consume the formula, e.g., that term infants cannot be used to determine effects for preterm infants or that a study of infants older than 6 months cannot be used to determine effects for infant age birth to 6 months. Both the study and control groups should be randomized and matched for sex, feeding history, gestational age, and general health. These recommendations apply to studies of both term and preterm infants.

Question 6: Listed below are examples of control feedings (clinical comparators):

- *(current infant formula (IF) + new ingredient) vs. (current IF) vs. (breast milk)*
- *(current IF + new ingredient) vs. (current IF)*
- *(current IF + new ingredient) vs. (breast milk)*
- *(current IF + new ingredient) vs. (formulas fed to historical infant cohort(s), e.g., Iowa data)*
- *(current IF + new ingredient) vs. (references that may include various type of feedings in such reference populations, e.g., NCHS and WHO)*
- *(IF + new ingredient)* vs. (any of the above controls)*

**Test formula contains new ingredient but the test formulation matrix differs from the new formula that firm intends to market containing the new ingredient.*

c. *What are the most distinguishing values and merits of each of these types of comparisons in infants fed a test formula vs. a comparative feeding for assessing normal physical growth?*

d. *Please rank these comparison based upon their potential for generating clinical data, which would be most relevant to an assessment of normal physical growth.*

Dr. Garza, with the permission of the task force members, proposed categories of most value, moderate value, and never consider doing for the clinical comparators. Following the discussion, four comparator groups were defined, which the task force members then ranked in order of preference. The four groups of comparators are:

- Two formulas (old and new), plus three references (including a breastfed reference)
- Two formulas (old and new), plus only a non-breastfed reference
- Two formulas (old and new), plus only a breastfed reference
- Two formulas (old and new) only

Discussion focused on building a breastfed reference data set as an opportunity for analyzing the difference between breastfed and formula-fed infant growth patterns. Comparisons to

longitudinal reference sets were seen as a way to control drift. Again, concurrent longitudinal studies were deemed the gold standard, with recent longitudinal studies second preference, cross-sectional studies third, and historical data last. Comparisons with multiple references were seen as feasible because current information technology makes such comparisons highly doable. It was noted that only a third of the data on breastfed infants in the Iowa study have been published and that the remaining data may be made available.

Task force members considered these questions together, reaching consensus that comparisons with the most potential for generating clinical data on normal physical growth include contrasting outcomes in infants fed the original formula, the original formula plus the new ingredient, human milk, and two references, e.g., Iowa, CDC, WHO, NCHS, and NHANES. The second preference was for comparisons among groups fed the original formula, the original formula plus the new ingredient, and human milk. The breastfed comparison was viewed as a way to begin accumulating data that compared formula-fed infant growth to that of breastfed infants. Because the second preference is included in the first preference, task force members agreed to the more encompassing comparison.

Question 7: With regard to formula composition changes:

c. Describe general principles and criteria that can be used to determine the need for a clinical study intended to provide assurance of normal physical growth.
d. Describe some of the specific changes in infant formula that would reasonably be expected to be accompanied by a clinical study to demonstrate normal physical growth.

Task force members were asked to develop their own scenarios for new infant formulas and then to determine the general principles and criteria they would use to determine if a clinical study of the new formula was warranted. After discussing several scenarios, it was evident that the same basic criteria were found each time a study was warranted.

Task force members reached consensus that an infant growth study is required for:

?? Major change in manufacturing process
?? Entirely new formula
?? Use of a substance that has not been tested in children before
?? Major changes in macronutrient content
?? Use of other compounds known to affect hormones, growth factors, genes, or metabolites that regulate growth
?? Formula changes that result in nutrient levels outside established ranges
?? Alterations likely to affect GI function or nutrient bioavailability
?? Studies on especially vulnerable populations
?? Use in a population distinct from that for whom the formula was intended originally

I certify that I attended the November 18-19, 2002 meeting of the Infant Formula Task Force of the Food Advisory Committee, and these minutes accurately reflect what transpired.

Jeanne E. Latham, M.S., R.D. Date
Executive Secretary

Cutberto Garza, M.D., Ph.D. Date
Chair