COMMON DESIGN ELEMENTS OF THE GIRLS HEALTH ENRICHMENT MULTI-SITE STUDIES (GEMS)

James Rochon, PhD; Robert C. Klesges, PhD; Mary Story, PhD; Thomas N. Robinson, MD; Tom Baranowski, PhD; Eva Obarzanek, PhD; Megan Mitchell, MPH

The Girls health Enrichment Multi-site Studies (GEMS) was a multi-center research program created for the purpose of testing interventions designed to prevent excess weight gain by African-American girls, as they enter and proceed through puberty. However, GEMS was not a ''multi-center clinical trial'' in the usual sense. Although these studies applied similar eligibility criteria, observed a similar follow-up schedule, and followed a similar measurement protocol, important differences existed, as well. Each field center developed its own intervention(s) and corresponding control, and tailored its study to the specific hypothesis being tested. Therefore, the study populations were somewhat different, with recruitment strategies that varied accordingly, and supplemental evaluations appropriate to the specific interventions were conducted on a site-specific basis. The purpose of this paper is to describe the common design elements of the GEMS Phase 1 pilot studies. This report presents the basic study design, a brief overview of the interventions, the measurements taken and their rationale, and procedures both for compiling the collaborative database, and performing site-specific analyses. (*Ethn Dis*. 2003; 13[suppl1]:S1-6–S1-14)

Key Words: Obesity Prevention, African-American Girls, Eligibility, Outcome Evaluations, Data Management, Statistical Analysis Plan

From the Biostatistics Center, George Washington University, Rockville, Maryland (JR, MM); Center for Community Health, University of Memphis, Memphis, Tennessee (RCK); Division of Epidemiology, University of Minnesota, Minneapolis, Minnesota (MS); Division of General Pediatrics and Center for Research in Disease Prevention, Stanford University, Palo Alto, California (TNR); Children's Nutrition Research Center, Baylor College of Medicine, Houston, Texas (TB); Division of Epidemiology & Clinical Applications, National Heart, Lung, and Blood Institute, Bethesda, Maryland (EO).

Address correspondence and reprint requests to James Rochon, PhD; Duke Clinical Research Institute; P.O. Box 17969; Durham, NC 27715; 919-668-8123; 919-668-7055 (fax); james.rochon@duke.edu

BACKGROUND AND RATIONALE

The Girls health Enrichment Multi-site Studies (GEMS) was a multi-center research program created to develop and test interventions designed to prevent excess weight gain by African-American girls, as they enter and proceed through puberty. Epidemiologic evidence indicates that the prevalence of obesity, ie, body mass index (BMI) $\geq$30, is particularly high in African-American women.[1] This excess body weight is present during childhood, and appears to increase more rapidly in African-American girls, compared to White girls.[2] Due to the possibility that healthful behaviors learned in childhood may be sustained through adulthood, prevention activities targeted at children are especially appealing, and potentially cost-effective.

GEMS was divided into 2 phases. Phase 1 was a 2¾-year development phase, during which 4 field centers conducted formative assessment studies consisting of focus groups, interviews, and surveys. This component was designed to address key conceptual issues in the design and conduct of these studies in an African-American population. Phase 1 culminated in randomized pilot studies in which the interventions were delivered over a period of 12 weeks. Evaluations were performed at baseline, and again at follow up, to provide a preliminary measure of the effectiveness of these interventions. In the second phase, a full-scale implementation of the derived methodology will be conducted, and final preparations are being made for the implementation at the time of this writing.

This issue of *Ethnicity & Disease* is devoted specifically to Phase 1 of this research program. The formative assessment process is described in Kumanyika et al.[3] In this paper, we focus on the common design elements of the GEMS pilot studies. The basic study design is presented, including a brief overview of the interventions and rationale for the measurements taken, and procedures for compiling the collaborative database, and for performing site-specific analyses. We focus on the common elements of the GEMS pilot studies and highlight important differences across the field centers.

BASIC STUDY DESIGN

There were 4 field centers, located at the University of Memphis, the University of Minnesota, Stanford University, and the Baylor College of Medicine. Their role was to design, implement, and evaluate interventions to prevent excess weight gain in 8- to 10-year-old African-American girls. A Coordinating Center (CC), located at the George Washington University Biostatistics Center, was responsible for providing support and coordination for key study activities, applying uniform quality control procedures, creating and maintaining a collaborative database, and performing primary statistical analyses for the 4 studies. The National Heart, Lung, and Blood Institute sponsored the study and participated in the program.

The studies were conducted as 4 inter-dependent, randomized clinical trials. The studies were "inter-dependent" in the sense that they considered similar study populations, followed a similar follow-up schedule, used BMI as the

primary measurement for body fatness, evaluated similar measurement variables, applied a common measurement protocol, imposed common quality control and data management procedures, and adopted a common analytic model. However, GEMS was not a "multi-center" clinical trial in the usual sense. A multi-center trial approach was not followed, due to lack of the experience and knowledge required to determine the best intervention to implement in a common protocol. Instead, each field center developed its own intervention (and control) programs. Although their design characteristics were similar, the field centers also had important differences. The target populations were somewhat different across the field centers, recruitment strategies were tailored to the specific needs of these groups, and supplemental evaluations appropriate to the specific interventions were performed on a site-specific basis.

Each study was reviewed by the Institutional Review Boards at the participating field centers, and at the coordinating center. Informed consent was obtained from parents/caregivers; individual girls provided their "assent" to participate.

STUDY POPULATION

The target population was 8- to 10-year-old, pre-adolescent, African-American girls, at risk of developing obesity. A girl was considered to be "African-American" if her parent/caregiver identified her race as "African-American" or "Black," regardless of whether any other races or ethnicities were identified. Moreover, the Memphis, Minnesota, and Stanford studies specifically targeted low-income participants, while the study at Baylor was directed more toward middle- and upper-income girls.

This age range was chosen specifically to intervene prior to a vulnerable developmental period, during which fat accumulation associated with sexual maturation increases substantially. For intervention effects to be most discernible, it was important for the target population to include girls at risk for developing obesity. Because body mass index (BMI) tracks moderately well over time, children at higher BMI levels are at higher risk of becoming overweight and obese adults, compared to children at lower levels of BMI.[4,5] In addition, the risk of obesity increases when there is one or more obese parent.[6] Therefore, all 4 field centers targeted girls who were at least at the 25th (Minnesota and Memphis) or 50th percentile (Baylor and Stanford) of age- and sex-specific BMI,[7] or who had at least one overweight parent/guardian, defined as having a BMI $\geq$25 kg/m^2 (Stanford). Because part of its intervention was Internet-based, Baylor required girls to have a computer and access to the Internet.

A detailed list of the GEMS exclusion criteria is provided in Table 1. In general, these included medical conditions and medications affecting growth; conditions limiting participation in the interventions (eg, unable to participate in routine physical education classes in school); conditions limiting participa-

Table 1. Detailed listing of GEMS exclusion criteria

Medical conditions affecting growth:
- Diagnosed with a genetic or metabolic disease/syndrome associated with obesity (Alstrom-Hallgren syndrome, brain tumor/craniopharyngioma, Carpenter syndrome, Cohen syndrome, Cushing syndrome, Down syndrome, growth hormone deficiency, hypothyroidism, insulin secreting pancreatic tumor, Laurence-Moon-Diedl [Bardet-Biedl] syndrome, polycystic ovary [Stain-Leventhal] syndrome, Prader-Willi syndrome, Pseudo-hypoparathyroidism type I, Turner syndrome).
- Type I diabetes, or type 2 diabetes taking medication (if not taking medication then eligible).
- Chronic gastrointestinal diseases (Gluten-induced enteropathy, Celiac Disease or Sprue, inflammatory bowel disease, Crohn's disease, ulcerative colitis, hepatitis currently or in the past year, cystic fibrosis, short bowel syndrome, liver transplant).
- Chronic kidney (renal) diseases (nephrotic syndrome, nephritis, kidney transplant).
- Structural heart disease/congenital heart disease (uncorrected), heart failure, or heart transplant.
- Anorexia nervosa or bulimia nervosa or binge eating disorder (present or past).
- AIDS and/or HIV infection.
- Pregnancy.

Medications affecting growth:
- Steroids taken by mouth or by injection more than 2 weeks in the past year (if less than or equal to 2 weeks in past year then eligible).
- Insulin injections.
- Oral anti-diabetic drugs.
- Thyroid hormones.
- Growth hormone injections.

Conditions limiting participation in the interventions:
- Unable to participate in routine physical education classes at school.
- Conditions requiring oxygen supplementation for exertion.
- Developmental or physical disability preventing participation in interventions.
- Girls or mothers who cannot medically participate in mild dietary restriction and/or increased physical activity.

Conditions limiting participation in the assessments:
- Girl or primary caregiver not fluent in English (written).
- Two or more grade levels behind in school for reading and writing.

Other:
- Inability to understand informed consent or no consent.
- Did not complete baseline assessments/missed appointment (randomization after baseline assessments completed).
- Plan to move from geographic area within 2 years.
- Homeless (Stanford only).
- No television where the girl lives the most (Stanford only).
- No access to the Internet (Baylor).

tion in the assessments (eg, unable to speak English, 2 or more grades behind in school for reading and writing); and other criteria (eg, inability or failure to provide informed consent). Medical conditions and medication use were ascertained from the parent/caregiver by interview.

SCREENING AND ELIGIBILITY

Girls and parents/caregivers underwent a screening process to verify that inclusion and exclusion criteria were satisfied. Basic demographic information, including age, gender, and ethnicity, were collected at this time. Eligibility criteria that required physical measurement, eg, BMI, as well as the sexual maturation evaluation, were collected at the baseline visit. After it was determined that all eligibility criteria had been satisfied, parents/caregivers signed the informed consent document and randomization was performed. Girls were the unit of randomization. Randomization was stratified by field center, and an urn randomization procedure[8] was used to generate the treatment allocation sequences. The different sequences were stored on a computer at the CC, and accessed using an interactive voice-response telephone system.

TREATMENT INTERVENTIONS

All interventions tested in the GEMS pilot study used an approach based on a behavioral theory, or conceptual, model. All interventions were 12 weeks in length, and were based on Social Cognitive Theory;[9] one intervention (at the University of Memphis) also drew upon a model for the influences of cultural factors on weight management.[10] Strategies focused on the interplay of behavioral, personal, and environmental factors that promote behavioral change. Details of the specific interventions are provided in the site-specific papers[11–14] in this issue. Here, we provide only a brief summary.

The intervention at Baylor was a 4-week summer day camp, followed by an 8-week Internet-based program, plus one Saturday meeting for the girls. The control group focused on general health issues. The Memphis study was designed as a 3-group study, comparing a child-targeted family intervention (intervention directly with the girls) and a family-targeted child intervention (intervention with parents only). The comparison intervention was a low-intensity program that focused on general health and self-esteem issues. The Minnesota study took place after school, 2 afternoons a week, in neighborhood community centers and schools. The intervention included activities designed to model and teach skills to prevent obesity and associated health problems through regular physical activity and healthy eating. The comparison group received a low-intensity program that focused on general health, self-esteem, and youth development. The Stanford study tested the efficacy of using a dance program to increase the girls' physical activity, along with a family-based intervention to reduce television, videotape, and video game use. The comparison group received an education-based intervention consisting of newsletters and community lectures.

MEASUREMENTS

The GEMS research program was created to develop and test interventions designed to prevent excess weight gain by African-American girls. The purpose of the Phase 1 pilot studies was to perform a "dress rehearsal" of the proposed methodologies, so that the primary and secondary "outcomes" really consisted of process measures, such as recruitment and retention rates, levels of attendance at the interventions, and consistency and reliability in collecting the requisite data. The Phase 1 pilot studies were never powered to detect significant differences with respect to traditional outcomes in obesity studies. Nonetheless, they are of considerable interest in their own right. In this section, therefore, we describe the measurements taken in the pilot studies. A systematic effort to validate these measures in this specific population of girls is being performed, and will be reported elsewhere.

Body Mass Index

GEMS was designed to promote physical activity and healthy dietary practices. The primary measurement for body fatness was body mass index (BMI).[15–17] BMI is a calculated value based on height and weight, ie, BMI= weight/height^2 with units, kg/m^2. BMI was chosen for several reasons. Height and weight can be measured with high accuracy and precision. These measures have excellent reliability, and have been correlated with body fat in young African-American girls.[18,19] Extensive normative data for height, weight, and BMI exist for children,[20] and BMI is sensitive to change following interventions in children.[21,22] BMI is also used clinically, and has been recommended as the measure of choice for adolescent preventive services.[23] Other complementary measures of body fat were also taken (see below).

Height was measured, with the girl wearing socks by using a Shorr infant/child/adult measuring board (Shorr Productions, Olney, Md). Weight was measured with the girl wearing socks and indoor light-weight clothing or gowns. Measurements were made using an electronic scale (SECA, Model 770, Seca Corporation, Culver City, Calif). Height was measured to the nearest 0.1 cm, and weight to the nearest 0.1 kg. Two readings of height and weight were taken on each occasion, and the mean of the 2 was used analytically.

Body Composition

All 4 centers measured waist circumference, as it is straightforward to mea-

sure, and is associated with adverse consequences of obesity in the general population,[24] and in young African-American girls.[25] Waist circumference is also sensitive to change during interventions.[26] Two readings of waist circumference were taken, following the method of Callaway et al,[27] using the umbilicus as a landmark. Measurements were made using a metal metric tape, capable of 1 mm repeatability. The mean of the 2 readings was used analytically.

Percent body fat was estimated by dual-energy x-ray absorptiometry (DEXA), at 3 centers (Baylor, Memphis, and Minnesota). Several studies have supported the use of body composition using DEXA, in this population.[28–30] Although BMI is related to body fat, body fat is a more direct measure of adiposity, because it also reflects, to some degree, height and lean body mass.[31,32] The Baylor and Memphis centers used the Hologic QDR 4500 instrument (Waltham, Mass), while Minnesota used the Lunar instrument (Madison, Wis).

24-hour Dietary Recalls

Dietary intake is viewed as one of the chief proximal mediators of change for BMI. Two 24-hour dietary recalls were performed on non-consecutive days at baseline, and again at the 12-week follow-up visit. One recall was performed in person, while the other was performed by telephone, approximately 1–2 weeks later. Parents were allowed to assist the girls' dietary recalls to improve validity.

This information was processed using the Nutrition Data System for Research (NDS-R), a software program developed at the University of Minnesota's Nutrition Coordinating Center (NCC) for collection, coding, and analysis of dietary data. Primary macro-nutrient variables of interest were total energy intake (kJ/day, kcal/day), and percent of energy derived from fat. The number of servings per day of fruit, juice, and vegetables was derived from the intake data. The approach used to count servings of fruit, juice, and vegetables is well established,[33] and followed the general behavioral model of the 5-a-day studies.[34] In addition, fluid ounces per day were calculated separately for water, soft drinks, fruit drinks, sports drinks, tea, flavored water, and artificially sweetened beverages. The macro-nutrients were averaged over the 2 recalls, while the servings were summed, to provide a better fit with the statistical models.

Dietary recalls also collected information on the activity the girl was performing while eating, with categories of: just eating, watching television, watching videotapes or movies, and other activity. Frequent television viewing is thought to be related to decreased physical activity, and an increased likelihood of obesity, and eating alone also appears to be related to weight gain.[35]

On-site quality control procedures were used to capture obvious errors and resolve local interviewer issues; otherwise, detailed quality assurance reviews were performed by the NCC. These methods have been validated for use in children as young as the third grade,[36,37] include the added value of parental assistance,[38,39] and have been shown to be sensitive to dietary interventions in children.[40,41]

Physical Activity

The empirical level of physical activity was derived using the Computer Sciences and Applications (CSA) monitor. The CSA records the acceleration and deceleration of movement in one direction, and reports movement as "activity counts" on a minute-by-minute basis. The girl was instructed to wear a CSA monitor continuously for a 72-hour period (except while showering or swimming), and to keep a log of the times when the device was off and on. After 3 days, the monitor was returned to the field center, and the minute-by-minute counts were downloaded from the device to a computer at the field center. The total number of counts, and the total number of minutes the monitor was worn, were aggregated to grand totals across the 3 days, and the average activity count per minute was computed as the ratio of the two. The CSA monitor has been used in several studies to assess physical activity in children, and has been assessed for validity and reliability.[42–44]

A self-report of physical activity was performed by using the Gems Activity Questionnaire (GAQ). The GAQ is a modification of the Self-Administered Physical Activity Checklist (SAPAC), which was validated and assessed for test-retest reliability on the same-day[45] and was further evaluated by the GEMS investigators in a validation study (Treuth, unpublished data). The questionnaire contains a checklist of 36 common physical activities, and asks the girl to estimate the length of time she spent engaged in these activities on the previous day (none, less than 15 minutes, or 15 minutes or more), and how often she "usually" engaged in these activities (none, a little, a lot). "Usual" activity was assessed because US adolescents tend to engage in little moderate or vigorous physical activity on any particular day,[46] and there was interest in collecting more "typical" activity. Information on 7 sedentary activities was requested in a similar manner.

Psychosocial Mediators and Modulators

Guided by social cognitive theory,[9] multiple levels of assessment (ie, personal, behavioral, and social) were considered in the selection of psychosocial constructs. The major categories of psychosocial measures were mediators of intervention effects, and potential effect moderators and confounders.

Psychosocial mediators at the family level included parent/caregiver reports of: 1) the availability/accessibility of fruit, juice, and vegetables;[47] 2) availability/accessibility of low-fat food alternatives; 3) home barriers to healthy eating; and 4) food preparation practices, adapted from Kristal's Food Habit Behavior Scale.[48] At

Table 2. Listing of psychosocial instruments by field center

Psychosocial Questionnaire	Source	Baylor	Mem-phis	Minne-sota	Stan-ford
Diet and nutrition related:					
Availability/accessibility of FJV	Parent	X	X		
Availability of low-fat and fat-free alternatives	Parent		X	X	
Home barriers to healthy eating	Parent	X	X		
Food preparation for daughter	Parent		X	X	
Child reported preferences—sweetened beverages	Girl	X	X		
Physical activity (PA) related:					
Child outcome expectancies for PA	Girl	X	X	X	
Self-efficacy for PA	Girl		X	X	
Physical performance self-concept	Girl	X	X	X	
PA preferences	Girl	X	X	X	X
Other psychosocial instruments:					
Optimism/pessimism	Girl	X	X		
Body figure silhouettes	Girl	X	X	X	X
Social desirability	Girl	X	X		
McKnight Risk Factor Survey	Girl	X	X		X

FJV=fruit, juice, and vegetables; PA=physical activity.

the child level, psychosocial mediators of primary importance included the girl's reports of: 1) outcome expectancies for physical activity; 2) self-efficacy for physical activity;[49] 3) physical activity self-concept;[50] and 4) physical activity preferences. Potential effect modifiers, and/or confounding measures, included the girl's reports of: 1) optimism/pessimism; 2) food preferences for sweetened beverages; 3) social desirability;[51] 4) body figure silhouettes; and 5) the McKnight Risk Factor Survey.[52]

Field centers performed different subsets of these evaluations, according to the specific needs of their studies. A detailed listing of the different instruments, and the centers applying them, is provided in Table 2.

SCHEDULE OF EVALUATIONS

Girls proceeded through a screening and eligibility phase to verify eligibility, record basic demographic information, and to document reasons for agreeing (or refusing) to participate in the program. Baseline measures collected from the parent/caregiver included self-reported age, race, ethnicity, cultural identity questions, socioeconomic status indicators, household membership characteristics, and their own height and weight. Data collected from the girls consisted of their stage of sexual maturation, a blood sample for clinical monitoring, as described below, and the primary and secondary measurements, as described above. Adverse events (AEs) were recorded at baseline to gauge the level of AEs pre-existing in this study population. Primary and secondary measurements were repeated at 12 weeks, post-randomization. Additionally, the Baylor center took an observation immediately following its summer camp to evaluate this component of its intervention. Girls were prompted for reports of adverse events at the follow-up visit. Parents/caregivers were specifically queried about any injuries since baseline. A detailed summary of the schedule of evaluations is provided in Table 3.

CLINICAL MONITORING AND ADVERSE EVENTS

Potential intervention-related risks included a reduction in growth due to a negative energy balance, resulting from a lower-calorie diet and increased energy expenditure. In contrast to weight-loss intervention programs, however, GEMS interventions encouraged healthful eating habits without restricting calorie intake, so that the risk of inadequate growth was thought to be small. Nonetheless, using recent Centers for Disease Control (CDC) growth charts[7] as a reference, low height was defined as a height-for-age below the 5th percentile; low weight was defined as BMI below the 5th percentile. Height velocity was also monitored, and girls growing at an annual rate of less than 3.5 cm per year were identified. The risk of participation "causing" a clinically significant eating disorder was also thought to be low in this population, and screening for this behavior was not specifically performed, except through the weight loss criteria noted above. Moreover, participation in the studies was not expected to "cause" abnormalities in serum insulin, glucose, and lipid measures. Nevertheless, "alert" values were defined according to NHLBI recommendations, or from clinically accepted practice,[53,54] as follows: glucose <40 mg/dL or >110 mg/dL; total cholesterol >200 mg/dL; HDL-cholesterol <35 mg/dL; LDL-cholesterol >130 mg/dL; triglycerides >135 mg/dL; systolic blood pressure >95th percentile for age, sex, and height; diastolic blood pressure >95th percentile for age, sex, and height. A report was issued by the CC to the field center involved for any girl exceeding these thresholds.

DATA MANAGEMENT PROCEDURES

Although data collected at the different field centers exhibited a considerable degree of commonality, significant variations were also apparent. GEMS investigators distinguished between data to be entered into a "collaborative" database compiled and maintained at the CC, and those compiled

Table 3. Schedule of evaluations in the GEMS pilot studies

GEMS Evaluation	Source*	Screening	Base-line	Follow-up
Screening and eligibility:				
Girl's age, race, ethnicity	P	X†		
Eligibility criteria	P, G	X		
Motivation for participating/reason(s) for refusing	P, G	X		
Parent/caregiver information				
Parent age, race, ethnicity, cultural identity	P		X	
Parent/caregiver weight and height	P		B, Mi, S	
Household SES—education, income, material possessions, household membership	P		X	
Primary outcome: height, weight→BMI	O		X	X
Other physical measures:				
Sexual maturation stage	O‡		X	
Waist circumferences	O		X	X
DEXA	O		B, Me, Mi	
Blood measurements:				
Fasting insulin and glucose	O		X	
Fasting lipids	O		Mi, S	
Diet and nutrition:				
NDS-R dietary recall (2×24 hr)	G		X	X
Physical activity:				
CSA monitor (3×24 hr)	G		X	X
GEMS activity questionnaire (GAQ)	G		X	X
Various psychosocial instruments	P, G		X	X
Participant safety and adverse events:				
Adverse events	P, G		X	X
Serious adverse events	P, G		When it occurred	

* P=parent; G=girl; O=observed or otherwise collected by a GEMS staff member.
† X=all 4 field centers; B=Baylor; Me=Memphis; Mi=Minnesota; and S=Stanford.
‡ Self-reported by girls at Stanford.

in "site-specific" databases. Generally, data collected in exactly the same manner by 2 or more field centers were considered "collaborative" and were entered into the collaborative database.

Collaborative data were forwarded to the CC from a variety of sources. Basic information, eg, eligibility criteria, demographic information, physical measures, and adverse events, was recorded on paper case report forms (CRFs), forwarded to the CC by parcel delivery service, and entered into the study database. The CSA accelerometer counts were downloaded from the devices into personal computer files at each field center. These files were processed and forwarded electronically to the CC. Twenty-four hour dietary recall data were entered directly into a personal computer at the field center, using the NDS-R software. Data files were forwarded to the NCC at the University of Minnesota for quality control review. An electronic file with the nutrition analysis was then forwarded to the CC. An 8-hour fasting blood sample was drawn from the girls (at baseline only) and sent to a laboratory at the University of Minnesota for analysis. Summary measures, including insulin, glucose, and cholesterol measures, were written to an electronic file and forwarded to the CC. Each field center entered the data from the psychosocial questionnaires using its own hardware and software configuration. These data were exported to SAS datasets, and forwarded electronically to the CC. In all cases, data for any particular girl were identified by her GEMS ID number, initials (except Stanford), and date of birth, so that the data would be attributed to the correct girl in the collaborative database.

A multi-faceted approach was implemented to safeguard the integrity of the data.[55] Preliminary validation checks were integrated into the data entry system. In general, these checks ensured that fields were of the right type (ie, numeric, character, or date), data for all key fields (eg, GEMS ID and date of evaluation) were present, skip patterns were correctly followed, and all data fields satisfied range checks for feasibility values. Next, a detailed set of validation checks (eg, range and consistency checks) was designed to ensure the integrity of the database. In general, these checks were designed to safeguard the completeness, accuracy, and timeliness of the accumulating data. Any discrepancy generated a "query," which was then faxed to the FC for review and resolution. Any changes were initialed and dated, so that an audit trail could be established.

QUALITY CONTROL PROCEDURES

A Manual of Procedures was written to fully describe all study procedures. Detailed training sessions were conducted immediately prior to data collection, and were led by the various subcommittees and working groups. Sessions were conducted on the physical measures (eg, height, weight, sexual maturation), evaluation procedures (eg, dietary recall and the NDS-R software, DEXA, and blood collection and processing), administering the psychosocial instruments, clinical monitoring activities, data collection procedures, and the randomization system. These sessions were designed to teach the various aspects of the study procedures, and to ensure that the procedures were applied

consistently across the field centers. A site visit was conducted to each field center during the pilot studies to ensure this consistency.

A detailed quality control report was reviewed periodically to safeguard the integrity of the accumulating database. In general, these reports considered the following issues:

1. Data management reports: The integrity of the data collection process including, for example, checks for any missing data, and the time lag in forwarding data to the CC.
2. Eligibility violations: This section considered whether any randomized girls were, in fact, ineligible, with respect to inclusion and exclusion criteria, eg, age, race, and BMI.
3. Procedural errors: Errors in performing the study procedures, including computing BMI for eligibility, performing the DEXA scan, taking the 8-hour fasting blood sample, and delays in giving the CSA monitors to the girls and retrieving them afterward.
4. Variability in the replicate evaluations: Discrepancies between the replicate observations of height, weight, and waist circumference, were evaluated; digit preferences in measuring these fields were investigated.
5. Procedures not performed: The frequency with which any study procedure (eg, DEXA, dietary recall) was not performed, together with the primary reason for this failure to perform them.
6. Field-center comparisons: Field-center differences with respect to the primary and secondary measurements were considered, using summary statistics and box-and-whisker plots.
7. Consistency among the variables: Consistency across the different evaluations, scatter plots of one variable against another, with special attention given to outliers.

In general, these procedures ensured a high level of confidence in the quality of the data.

STATISTICAL ANALYSIS PLAN

A common analytic approach was adopted to evaluate the effectiveness of the interventions. Nonetheless, because each study was designed as a single-center study, primary analyses were performed on a site-specific basis.

All studies were analyzed according to "intention-to-treat" principles.[56–61] Two-sided tests of significance were performed. In 3 of the 4 field centers, a single intervention was compared against its control group, and this comparison was made at the $\alpha = 0.05$ level of significance. The design at the University of Memphis incorporated 2 active interventions against a control group. An omnibus 2 *df* test was performed first, to detect significant differences across the 3 treatment arms. If a significant result was found, pair-wise comparisons among the 3 interventions were performed at the .05 level of significance, to guide interpretation of the primary test.

Statistical comparisons were performed to compare treatment groups with respect to demographic characteristics, and important prognostic factors at baseline. For binary and ordinal variables, standard techniques for categorical data were applied, including the Fisher exact test for binary variables, and Pearson χ^2, and Mantel-Haenszel procedures for ordinal data. For continuous variables, the Wilcoxon-Mann-Whitney test was performed for 2-group comparisons, and the Kruskal-Wallis test used for the 3-group comparisons at the University of Memphis. Similar procedures were applied to the baseline value of the measurements to determine any systematic differences among the treatment groups at baseline.

For continuous variables, analysis of covariance (ANCOVA) was the primary method to assess between-group differences in the outcome measures at follow up. The baseline value of the measure, centered around its sample mean, was entered as a covariate, to increase precision and adjust for any imbalances at baseline. Standard techniques were performed to ensure that the data were consistent with the underlying assumptions for Gaussian outcomes, including residual plots and tests for normality. Because the "number of servings" variables (eg, the number of servings of sweetened beverages) represented a "count" measure, Poisson regression models[62] were applied instead. The sum across the 2 recalls was used to provide a better fit with these models.

A between-group comparison of the proportion of randomized girls who did not return for the follow-up evaluation was performed using the Fisher exact test. Moreover, the time elapsed from initiating the intervention to the follow-up evaluation was derived, and then categorized as "early" (ie, <77 days), "on time" (ie, ≥ 77 days and ≤ 112 days) and "late" (ie, >112 days). This was treated as a nominal variable, and a between-group comparison was performed using the Fisher exact test. Incidence of injuries to the girls during the study were compared, using the Fisher exact test. Among girls with at least one injury, the median number of injuries per girl was compared, using a Wilcoxon-Mann-Whitney test (Kruskal-Wallis test, in the case of Memphis). Similar comparisons were performed for the incidence of adverse events, and the median number of AEs per girl.

DISCUSSION

The purpose of the GEMS project was to investigate several approaches for preventing the increase in obesity and overweight among 8- to 10-year-old African-American girls. Each field center was charged with developing and testing its own approach, and the GEMS in-

vestigators responded with different interventions, directed at somewhat different populations. The advantage to this method is that, in the midst of an epidemic, multiple new approaches could be evaluated simultaneously. Interventions could also be tailored to the specific needs of the local community, and this flexibility was considered essential for community-based research with this target population.

However, GEMS was not a "multicenter" clinical trial in the usual sense. It did not develop a common protocol to be implemented uniformly across the field centers. In this sense, GEMS was closer to the "network" paradigm for conducting multiple clinical trials, as widely adopted in the treatment fields of oncology,[63] AIDS,[64] and asthma,[65] for example. Multiple protocols were evaluated simultaneously, albeit with one protocol per center. The disadvantage, of course, is that generalizability for any study was restricted to the specific center in which it was conducted. Moreover, it was difficult to pool results across the 4 studies to determine whether behavioral interventions, in general, are effective in this population of girls. Adjustments for demographic (and geographic) differences in the cohorts must be made. Moreover, given the disparate nature of the individual studies, the only hypothesis that can be tested is whether a treatment difference was observed at one or more of the field centers.

Consistent with the network paradigm, GEMS investigators recognized the need to standardize the design and conduct of these studies as much as possible, and the methodology presented in this paper demonstrates a high level of commonality across the 4 studies. Outcome measures demonstrated broad agreement; a collaborative database was developed, with ongoing data entry; common quality control procedures, including site visits to the field centers, were adopted; and a common analytic model was applied. In this sense, therefore, Phase 1 of GEMS was considerably more than the sum of its parts.

ACKNOWLEDGMENTS

This research was funded by grant numbers UO1-HL62662, UO1-HL62663, UO1-HL62668, UO1-HL62732, and UO1-HL65160, from the National Heart, Lung, and Blood Institute. We thank the editors and reviewers for helpful comments that improved the manuscript considerably.

REFERENCES

1. Fingerhut LA, Warner M. *Injury Chartbook. Health, United States, 1996–97.* Hyattsville, Md: National Center for Health Statistics, 1997.
2. Troiano RP, Flegal KM, Kuczmarski RJ, Campbell SM, Johnson CL. Overweight prevalence and trends for children and adolescents: the National Health and Nutrition and Examination Surveys, 1963 to 1991. *Arch Pediatr Adolesc Med.* 1995;149:1085–1091.
3. Kumanyika SK, Story M, Beech BM, et al. Collaborative planning process for formative assessment in the Girls health Enrichment Multi-site Studies (GEMS): a retrospection. *Ethn Dis.* 2003;S1-15–S1-19.
4. Obarzanek E. Obesity in children, adolescents, and families. In: Fletcher GF, Grundy SM, Hayman LL, eds. *Obesity: Impact on Cardiovascular Disease.* Armonk, NY: Futura Publishing Co; 1999:31–53.
5. Serdula MK, Ivery D, Coates RJ, Fredman DS, Williamson DF, Byers T. Do obese children become obese adults? A review of the literature. *Prev Med.* 1993;22:167–177.
6. Whitaker RC, Wright JA, Pepe MS, Seidel KD, Dietz WH. Predicting obesity in young adulthood from childhood and parental obesity. *N Engl J Med.* 1997;337:869–873.
7. Kuczmarksi RJ, Ogden CL, Grummer-Strawn LM, et al. *CDC Growth Charts: United States.* Hyattsville, Md: National Center for Health Statistics; 2000. Advance Data from Vital and Health Statistics, No. 314.
8. Wei LJ, Lachin JM. Properties of urn randomization in clinical trials. *Control Clin Trials.* 1988;9:345–364.
9. Bandura A. *Social Foundations of Thought and Action. A Social Cognitive Theory.* Englewood Cliffs, NJ: Prentice-Hall; 1986.
10. Kumanyika SK, Morssink CB. Cultural appropriateness of weight management programs. In: Dalton S, ed. *Overweight and Weight Management.* Gaithersburg, Md: Aspen Publisher; 1997:69–106.
11. Baranowski T, Baranowski J, Cullen KW, et al. The Fun, Food, and Fitness Project (FFFP): the Baylor GEMS pilot study. *Ethn Dis.* 2003;S1-30–S1-39.
12. Beech BM, Klesges RC, Kumanyika S, et al. Child- and parent-targeted interventions: the Memphis GEMS pilot study. *Ethn Dis.* 2003; S1-40–S1-53.
13. Story M, Sherwood NE, Himes JH, et al. An after-school obesity prevention program for African-American girls: the Minnesota GEMS pilot study. *Ethn Dis.* 2003;S1-54–S1-64.
14. Robinson TN, Killen JD, Kraemer HC, et al. Dance and reducing television viewing to prevent weight gain in African-American girls: the Stanford GEMS pilot study. *Ethn Dis.* 2003;S1-65–S1-77.
15. Winkleby MA, Robinson TN, Sundquist J, Kraemer HC. Ethnic variation in cardiovascular disease risk factors among children and young adults: findings from the Third National Health and Nutrition Examination Survey, 1988–1994. *JAMA.* 1999;281:1006–1013.
16. Dietz WH, Robinson TN. Use of the body mass index (BMI) as a measure of overweight in children and adolescents. *J Pediatr.* 1998; 132:191–193.
17. Malina RM, Katzmarzyk PT. Validity of the body mass index as an indicator of the risk and presence of overweight in adolescents. *Am J Clin Nutr.* 1999;70:131S–136S.
18. Bouchard C. Validity of anthropometry in classifying youths as obese. *Int J Obes.* 1989; 13:183–193.
19. Pietrobelli A, Faith MS, Allison DB, Gallagher D, Chiumello G, Heymsfield SB. Body mass index as a measure of adiposity among children and adolescents: a validation study. *J Pediatr.* 1998;132:204–210.
20. Rosner B, Prineas R, Loggie J, Daniels SR. Percentiles for body mass index in US children 5- to 17-years of age. *J Pediatr.* 1998; 132:211–222.
21. Killen JD, Telch MJ, Robinson TN, Maccoby N, Taylor CB, Farquhar JW. Cardiovascular disease risk reduction for tenth graders: a multiple-factor school-based approach. *JAMA.* 1988;260:1728–1733.
22. Smith JC, Sorey WH, Quebedeau D, Skelton LJ. Use of body mass index to monitor treatment of obese adolescents. *Adolesc Health.* 1997;20:466–469.
23. Himes JH, Dietz WH. Guidelines for overweight in adolescent preventive services: recommendations from an expert committee. *Am J Clin Nutr.* 1994;59:307–316.
24. Ledoux M, Lambert J, Reeder BA, Despres JP. A comparative analysis of weight to height and waist to hip circumference indices as indicators of the presence of cardiovascular disease risk factors. Canadian Heart Health Surveys Research Group. *CMAJ.* 1997;157(suppl 1):S32–S38.
25. Freedman DS, Serdula MK, Srinivasan SR, Berenson GS. Relation of circumferences and skinfold thicknesses to lipid and insulin concentrations in children and adolescents: the Bogalusa Heart Study. *Am J Clin Nutr.* 1999; 69:308–317.
26. Han TS, Richmond P, Avenell A, Lean ME. Waist circumference reduction and cardiovascular benefits during weight loss in women. *Int J Obes Relat Metab Disord.* 1997;21:127–134.

27. Callaway CW, Chumlea WC, Bouchard C, et al. Circumferences. In: Lohman TG, Roche AF, Martorell R, eds. *Anthropometric Standardization Reference Manual.* Champaign, Ill: Human Kinetics Books; 1988:45–46.
28. Ellis KJ, Abrams S, Wong WW. Body composition of a multiethnic young female population. *Am J Clin Nutr.* 1997;65:724–731.
29. Gutin B, Litaker M, Islam S, Manos T, Smith C, Triber F. Body-composition measurement in 9–11-y-old children by dual-energy X-ray absorptiometry, skinfold-thickness measurements, and bioimpedance analysis. *Am J Clin Nutr.* 1996;63:287–292.
30. Nelson DA, Simpson PM, Johnson CC, Barondess DA, Kleerekoper M. The accumulation of whole body skeletal mass in third- and fourth-grade children: effects of age, gender, ethnicity, and body composition. *Bone.* 1997; 20:73–78.
31. Garn SM, Leonard WR, Hawthorne VM. Three limitations of the body mass index. *Am J Clin Nutr.* 1986;44:996–997.
32. Maynard LM, Wisemandle W, Roche AF, Chumlea WC, Guo SS, Siervogel RM. Childhood body composition in relation to body mass index. *Pediatrics.* 2001;107:344–350.
33. Cullen KW, Baranowski T, Baranowski J, Hebert D, de Moor C. Behavioral or epidemiologic coding of fruit and vegetable consumption from 24-hour dietary recalls: research question guides choice. *J Am Diet Assoc.* 1999;99:849–851.
34. Eldridge A, Smith-Warner SA, Lytle LA, Murray DM. Comparison of 3 methods for counting fruits and vegetables for fourth-grade students in the Minnesota 5-A-Day Power Plus Program. *J Am Diet Assoc.* 1998; 98:777–782.
35. Robinson TN. Television viewing and childhood obesity. *Pediatr Clin North Am.* 2001; 48:1017–1025.
36. Lytle L, Nichaman M, Obarzanek E, Glovsky E, Montgomery D, Nicklas T. Validation of 24-hour recalls assisted by food records in third-grade children. *J Am Diet Assoc.* 1993; 93:1431–1436.
37. Van Horn L, Gernhofer N, Moag-Stahlberg A, Farris R, Hartmuller G, Lasser V. Dietary assessment in children using electronic methods: telephones and tape recorders. *J Am Diet Assoc.* 1990;90:412–416.
38. Van Horn L, Stumbo P, Moag-Stahlberg A, et al. The Dietary Intervention Study in Children (DISC): dietary assessment methods for 8–10-year-olds. *J Am Diet Assoc.* 1993;93: 1396–1403.
39. Eck L, Kleges R, Hanson C. Recall of a child's intake from one meal: are parents accurate? *J Am Diet Assoc.* 1989;89:784–789.
40. Lytle L, Stone E, Nichaman M, Perry C, Montgomery D, Nicklas T. Changes in nutrient intakes of elementary school children following a school-based intervention: results from the CATCH Study. *Prev Med.* 1996;25: 465–477.
41. Writing Group for the DISC Collaborative Research Group. Efficacy and safety of lowering dietary intake of fat and cholesterol in children with elevated low-density lipoprotein cholesterol. The Dietary Intervention Study in Children (DISC). *JAMA.* 1995;273:1429–1435.
42. Janz KF. Validation of the CSA accelerometer for assessing children's physical activity. *Med Sci Sports Exerc.* 1994;26:369–375.
43. Janz KF, Witt J, Mahoney LT. The stability of children's physical activity as measured by accelerometry and self-report. *Med Sci Sports Exerc.* 1995;27:1326–1332.
44. Trost SG, Ward DS, Moorehead SM, Watson PD, Riner W, Burke JR. Validity of the computer science and applications (CSA) activity monitor in children. *Med Sci Sports Exerc.* 1998;30:629–633.
45. Sallis JF, Strikmiller PK, Harsha DW, et al. Validation of interviewer- and self-administered physical activity checklists for fifth grade students. *Med Sci Sports Exerc.* 1996;28:840–851.
46. US Department of Health and Human Services. Patterns and Trends in Physical Activity. In: *Physical Activity and Health: A Report of the Surgeon General.* Atlanta, Ga: US Dept of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion; 1996:186–201.
47. Hearn MD, Baranowski T, Baranowski J, et al. Environmental influences on dietary behavior among children: availability and accessibility of fruits and vegetables enable consumption. *J Health Educ.* 1998;29:26–29.
48. Kristal AR, Shattuck AL, Henry HJ, Fowler AS. Rapid assessment of dietary intake of fat, fiber, and saturated fat: validity of an instrument suitable for community intervention research and nutritional surveillance. *Am J Health Promot.* 1990;4:288–295.
49. Saunders RP, Pate RR, Felton G, et al. Development of questionnaires to measure psychosocial influences on children's physical activity. *Prev Med.* 1997;26:241–247.
50. Harter S. The perceived competence scale for children. *Child Dev.* 1982;53:87–97.
51. Reynolds CR, Paget KD. National normative and reliability data for the Revised Children's Manifest Anxiety Scale. *School Psychol Rev.* 1983;12:324–336.
52. Shisslak CM, Rneger R, Sharpe T, et al. Development and evaluation of the McKnight Risk Factor Survey for assessing potential risk and protective factors for disordered eating in preadolescent and adolescent girls. *Int J Eat Disord.* 1999;25:195–214.
53. Expert Panel on Blood Cholesterol Levels in Children and Adolescents. National Cholesterol Education Program Report. *Pediatrics.* 1992;89(suppl):525–584.
54. National High Blood Pressure Education Program Working Group on Hypertension Control in Children and Adolescents. *Update on the Task Force Report (1987) on High Blood Pressure in Children and Adolescents: A Working Group Report from the National High Blood Pressure Education Program.* Bethesda, Md: National Institutes of Health, National Heart, Lung, and Blood Institute; September 1996. NIH Publication No. 96-3790.
55. McFadden E. *Management of Data in Clinical Trials.* New York, NY: John Wiley & Sons; 1998.
56. Pocock SJ. *Clinical Trials. A Practical Approach.* Chichester: John Wiley & Sons; 1983.
57. Peduzzi P, Detre K, Wittes J, Holford T. Intent-to-treat analysis and the problem of crossovers. An example from the Veterans Administration coronary bypass surgery study. *J Thorac Cardiovasc Surg.* 1991;101:481–487.
58. Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. *Int J Epidemiol.* 1992;21:837–841.
59. Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: is it really an option? *Stat Med.* 1991;10:1595–1605.
60. Weinstein GS, Levin B. Effect of crossover on the statistical power of randomized studies. *Ann Thorac Surg.* 1989;48:490–495.
61. Lavori PW. Clinical trials in psychiatry: should protocol deviation censor patient data (with discussion). *Neuropsychopharmacology.* 1992;6:39–63.
62. McCullagh P, Nelder JA. *Generalized Linear Models.* 2nd ed. London, England: Chapman & Hall; 1989.
63. Carbone PP, Tormey DC. Organizing multicenter trials: lessons from the cooperative oncology groups. *Prev Med.* 1991;20:162–169.
64. Rosendorf LL, Dafni U, Amato DA, et al. Performance evaluation in multicenter clinical trials: development of a model by the AIDS Clinical Trials Group. *Control Clin Trials.* 1993;14:523–537.
65. Kephart DK, Chinchilli FM, Hurd SS, Cherniack RM. The organization of the Asthma Clinical Research Network: a multicenter, multiprotocol clinical trials team. *Control Clin Trials.* 2001;22:119S–125S.