



Girls health Enrichment Multi-Site Studies

## **Pilot Study**

**Adapted version of the Statistical Analysis Plan**

Compiled by James Rochon, Ph.D

The Biostatistics Center  
The George Washington University

Version 2.1: November 26, 2001.

# Girls health Enrichment Multi-Site Studies

## Pilot Study

### Statistical Analysis Plan

#### Table of Contents

	page
<b>1. INTRODUCTION AND OVERVIEW.....</b>	<b>3</b>
1.1 Summary of the GEMS Protocol.....	3
1.2 Specific Aims.....	5
1.3 Collaborative vs. Site-Specific Data .....	5
1.4 Scope of This Analysis Plan .....	5
1.5 References to Specific Items on the Case Report Forms.....	6
<b>2. ANALYSIS POLICIES .....</b>	<b>6</b>
2.1 Separate Analysis by Field Center .....	6
2.2 Analysis Population.....	6
2.3 Type-I Error .....	7
2.4 Interim Analyses for Efficacy .....	7
2.5 “Cutpoints”.....	7
<b>5. SECONDARY OUTCOMES .....</b>	<b>9</b>
5.1 Anthropometric Measures.....	<b>9</b>
5.1.1 <i>Waist Circumference</i> .....	<b>9</b>
5.1.2 <i>DEXA Evaluation</i> .....	<b>9</b>
5.2 Measures of Physical Activity .....	9
5.2.1 <i>GEMS Activity Questionnaire</i> .....	<b>9</b>
5.2.2 <i>CSA Activity Counts</i> .....	10

## 1. INTRODUCTION AND OVERVIEW

### 1.1 Summary of the GEMS Protocol

**Background and Rationale:** In the U.S. obesity is a major cardiovascular disease risk factor that is not showing improvement over the past few decades. The prevalence of obesity (BMI  $\geq$  30) is particularly high in African-American women (36.5%), and may be a contributing factor to their higher mortality from cardiovascular disease and higher prevalence of diabetes and hypertension compared with White and Hispanic women. This increased prevalence is present even during childhood, and the prevalence of obesity is increasing faster in African-American girls than in White girls. Furthermore, tracking correlations of weight, BMI, or weight-for-height are moderately high, so that obese children are at high risk of becoming obese adults.

It is generally agreed that the etiology of obesity is multifactorial. Excess energy intake and/or low energy expenditure contribute to energy imbalance which results in obesity. Eating and physical activity behaviors, psychosocial factors, socioeconomic status, and family and environmental influences are associated with obesity and behaviors related to obesity; genetic factors strongly influence adiposity. However, weight loss in obese children is difficult to achieve, and there are potential safety concerns in growing children. It may compromise children's nutritional status, and dieting has been linked to eating disorders, especially in girls. Nonetheless, the possibility that healthful behaviors learned in childhood may be sustained through adulthood makes prevention activities targeted at children especially appealing and potentially cost-effective.

**Specific Aims and Objectives:** The purpose of the Girls Health Enrichment Multi-site Studies (GEMS) is to develop and test several interventions to prevent excessive weight gain by African-American girls as they enter and proceed through puberty. The specific aim of the Pilot study is for each field center to conduct a 12-week randomized, controlled clinical trial of its selected intervention. These studies deliver the interventions over a 12-week period; demonstrate the ability to take high quality measurements; evaluate the 12-week intervention; and, prepare for Phase 2 of the research program.

**Basic Study Design:** GEMS is a multi-center research program in which four field centers independently develop and test interventions to prevent excessive weight gain by African-American girls as they enter and proceed through puberty. All four field centers enroll girls according to commonly developed eligibility criteria, allowing for some center-specific differences in eligibility criteria to meet the needs of their specific intervention. They conduct common key measurements according to a standardized protocol which requires central training and certification.

The GEMS Pilot study is conducted over the 9-month period from January to September, 2001. A sample of 20-30 girls is recruited into each treatment group in each field center. Interventions are conducted over a 12-week period. Evaluations are performed at baseline and at the end of 12 weeks.

**Study Population:** The target population is 8-10 year-old African-American girls who are preadolescent and are at high risk of developing obesity. The age range of 8-10 years was chosen to intervene prior to a vulnerable developmental period when fat accumulation associated with sexual maturation increases substantially. The eligibility criteria are similar, but not identical, across the four studies. All four field centers target girls who are at least at the 25<sup>th</sup> (Minnesota) or 50<sup>th</sup> percentile (Baylor, Memphis, and Stanford) of age- and sex-specific BMI or have at least one overweight parent/guardian defined as having a BMI  $\geq$  25 kg/m<sup>2</sup> (Stanford). Baylor requires the child to have a computer and internet access because part of the intervention is web-based. Exclusion criteria include medical conditions and medications affecting growth; conditions limiting participation in the interventions (e.g., unable to participate in routine physical education classes in school); conditions limiting

participation in the assessments (e.g., two or more grades behind in school for reading and writing); and other criteria (e.g., inability or failure to provide informed consent).

**Treatment Interventions:** The interventions to be tested in GEMS are developed using the accepted approach in behavior change research of basing intervention approaches on a behavioral theory or conceptual model. Three of the interventions are designed based on Social Cognitive Theory, while the fourth is based on a conceptual model. Strategies focus on behavioral, personal, and environmental factors that interplay to promote behavioral change. The interventions and their theoretical framework are described in greater detail in the GEMS protocol.

**Sample Size Considerations:** Sample sizes for the GEMS Pilot study are driven by feasibility considerations. Rigorous sample size calculations were not performed but are addressed in Phase 2. This implies that each study is not adequately powered to detect “significant” differences among the treatment groups. Nonetheless, it may be possible to detect “trends” in the between-group differences.

**Primary and Secondary Outcomes:** The primary outcome measure for GEMS is body mass index (BMI), a calculated value based on height and weight (i.e.,  $BMI = kg / m^2$ ). BMI was chosen for the following reasons: height and weight can be measured with high accuracy and precision; they have been correlated with body fat in young African-American girls; extensive normative data for height, weight, and BMI exist for children; and, BMI is sensitive to change following interventions in children. Body composition as measured by DEXA (Baylor, Memphis, Minnesota) or triceps skinfold thickness (Stanford) is being taken, but only at baseline. Dietary intake, as measured by two non-consecutive 24-hr dietary recalls; physical activity measured through recall and direct observation using the Computer Sciences and Applications (CSA) accelerometer device; and, a variety of psychosocial questionnaires are being recorded.

**Schedule of Evaluations:** Girls proceed through a Screening & Eligibility phase to verify that they satisfy the eligibility criteria. Baseline measures taken from the parent(s) include age, race, ethnicity, household membership, SES, cultural identity, and their height and weight. Data recorded from the girls consist of the primary and secondary outcomes as described above. Primary and secondary outcomes (with the exception of DEXA and blood collection) are repeated at 12 weeks post-randomization. Additionally, Baylor takes observations immediately following its summer camp.

**Clinical Monitoring and Adverse Events:** Potential intervention-related risks potentially include a reduction in growth (height and/ or weight) due to a negative energy balance resulting from a lower-calorie diet and increased energy output. In contrast to other intervention programs, however, GEMS interventions encourage healthy eating habits without restricting calorie intake, so that the risk that participation will lead to inadequate growth is small. Nonetheless, using recent CDC growth charts as a reference, low height is defined as a height-for-age below the 5<sup>th</sup> percentile; low weight is defined as BMI below the 5<sup>th</sup> percentile. Height velocity is also monitored, and girls growing less than 3.5 cm per year in stature are identified. The risk of participation “causing” a clinically significant eating disorder is low. Screening for eating disorders is not specifically performed except through the weight loss criteria noted above. Moreover, participation in the studies is not expected to “cause” abnormalities in serum insulin, glucose and lipid measures. Nevertheless, “alert” values are defined according to NHLBI recommendations or clinically accepted practice. Adverse events are recorded at baseline to gauge the level of AEs during the previous 12 weeks and again after the 12-week intervention. GEMS staff also ask parents or caregivers specifically about any injuries that have occurred to the girls during this time frame. Serious adverse events are reported promptly to the study sponsor, the DSMB, the Principal Investigators, and their IRBs.

## 1.2 Specific Aims

The Specific Aims of the GEMS Pilot Study are summarized as follow:

1. Conduct a 12-week randomized, controlled pilot study:
  - a. Develop a 12-week intervention and a comparison condition
  - b. Recruit at least 240 girls and parents across the 4 field centers into the pilot study
  - c. Take follow-up measurements on at least 85% of the participants
2. Deliver a 12-week intervention:
  - a. At least 75% of the participants attend at least 80% of the expected site-defined intervention events.
  - b. At least 80% of the planned intervention is delivered.
  - c. At least 70% of participants respond positively on an evaluation form.
  - d. At least 75% of the intervention staff respond positively on an evaluation form
  - e. Generate intervention reports of attendance, core elements of intervention delivery, and the fidelity to the planned intervention
3. Demonstrate the ability to take high quality measurements:
  - a. Generate recruitment reports showing the number of participants randomized by each field center
  - b. Generate reports of follow-up rates for data collection visits
  - c. Generate other quality control reports including measures of height, weight, dietary intakes, and physical activity
4. Evaluate the 12-week intervention:
  - a. Compared to the comparison group, the intervention group will have more healthful dietary intakes and practices as assessed by the mean of two 24-hour dietary recalls and related instruments:
  - b. Compared to the comparison group, the intervention group will increase their level of physical activity and, for Stanford, decrease sedentary activities.
  - c. Compared to the comparison group, the intervention group will demonstrate changes in psychosocial outcomes, as assessed by self-report questionnaires.
5. Prepare Protocol for Phase 2

## 1.3 Collaborative vs. Site-Specific Data

Although there is a considerable amount of commonality in the data collected at the different field centers (FCs), there are significant variations as well. To take these characteristics into consideration, GEMS investigators distinguished between data to be entered into a “collaborative” database being compiled and maintained at the Coordinating Center (CC) versus those compiled in “site-specific” databases. Generally, data collected in exactly the same manner by two or more field centers are considered “collaborative” in nature and entered into the collaborative database. Data collected at one field center, or collected in an idiosyncratic manner by one field center, are considered “site-specific.”

It is the responsibility of the GEMS Coordinating Center to conduct statistical analysis on data entered into the collaborative database. The individual field centers have responsibility to enter site-specific data into their own databases and perform site-specific analyses on these data. This document discusses statistical analysis from the collaborative database only.

## 1.4 Scope of This Analysis Plan

This document specifically elaborates on statistical procedures outlined in the study protocol to describe the analysis population and to evaluate the potential efficacy and safety of the interventions. Procedures for describing (and comparing) the baseline characteristics of the

analysis population; algorithms for creating the primary and secondary outcomes from the specific fields on the Case Report Forms and electronic datafiles; statistical procedures for between-group comparisons on the primary and secondary outcomes and verifying the underlying assumptions; and the evaluation of adverse events and clinical monitoring are described.

Other analysis plans are being written for a variety of supplemental analyses from the GEMS Pilot study. At the time of writing, they include:

- Psychometric analyses on a number of psychosocial questionnaires
- The utility of performing a central review of DEXA parameters
- The value-added from a central review of the dietary data by the Nutrition Coordinating Center.

These analyses will be summarized in separate analysis plans.

## 1.5 References to Specific Items on the Case Report Forms

Mnemonic coding is frequently used in this document to refer to specific questions on the Case Report Forms. “S&E” refers to the Screening & Eligibility form; “BE” refers to the Baseline Evaluation; “PC” refers to the Procedures Checklist; and, “FU” refers to the Follow-up Evaluation. “SE F2” refers to Question F2 on the Screening and Eligibility data form; “FU E1” refers to Question E1 on the Follow-up form, and so on.

## 2. ANALYSIS POLICIES

### 2.1 Separate Analysis by Field Center

The GEMS research program is conducted as four inter-dependent clinical trials, each investigating a different approach to reducing excessive weight gain in African-American girls as they enter and proceed through puberty. They are “inter-dependent” in the sense that they consider similar study populations, follow similar follow-up schedules, use BMI as the primary outcome measure, and evaluate similar secondary outcomes. Nonetheless, GEMS is not a “multi-center clinical trial” in the usual sense – each field center is evaluating its own intervention, and although the design characteristics are “similar,” there are important differences. The goal of the primary analysis plan, therefore, is to apply the most appropriate analytic strategy for each clinical trial, taking their special design characteristics into consideration. Thus, the different studies are analyzed separately and the analytic approach is tailored to specific characteristics of each study.

### 2.2 Analysis Population

Data from these pilot studies are analyzed according to “Intention-to-Treat” (ITT) principles. The following ITT principles are applied:

1. Randomization is adopted as the “gateway” to the statistical analyses. All participants randomized at any field center and who provide data are included in all analyses irrespective of protocol violations and intercurrent events arising post randomization (e.g., a participant failing to complete the intervention).
2. Participants are analyzed according to the treatment group to which they were originally assigned, irrespective of cross-over and drop-in to other treatment combinations.
3. All participants continue to be followed beyond withdrawal, cross-over, and any other intercurrent events until the scheduled end of follow-up at Week 12.
4. No imputation is performed for missing follow-up data in the primary analysis. That is, only those data actually observed or recorded at the follow-up visits are included in the primary analysis. Secondary analyses which impute missing data under a number of scenarios is discussed in Section **Error! Reference source not found.**

### **2.3 Type-I Error**

Two-sided tests of significance are performed. Because each study is independent of the others, overall type-I error is set at  $\alpha = .05$  for each study. In three of the four field centers, a single intervention is compared against a single comparison (control) group so that this comparison can be made at the  $\alpha = .05$  level of significance. The design at the University of Memphis incorporates two active interventions against a control group. An omnibus 2 d.f. test is first performed to detect significant differences across the three treatment arms. If a significant result is found, pairwise comparisons among the three interventions are performed at the .05 level of significance to guide interpretation of the primary test.

### **2.4 Interim Analyses for Efficacy**

Because this is a pilot study designed to prepare for the Phase 2 studies, no interim analyses for efficacy are performed during the GEMS Pilot study.

### **2.5 “Cutpoints”**

In some cases, e.g., age of the parent / caregiver, a continuous random variable is also broken down into discrete categories. This is primarily for descriptive reasons; however, it may also be used this way analytically, especially when we are unwilling to postulate a linear (or, even curvilinear) relationship with the outcome variable. Suggestions on how to derive these “cutpoints” are made in this document. However, the final determination can only be made after an examination of the actual distribution of the variable.

## 5.2 Measures of Physical Activity

### 5.2.1 GEMS Activity Questionnaire

- Summary score for activity “yesterday”
- Summary score for activity “usually” performed.

For each item, the following values are assigned: 0=“None”, 1=“Less than 15 minutes”, and 10=“15 minutes or more”. “MET” values are assigned to 28 of the 29 activities (the 29th is an “Other” activity).

In a preliminary study, however, only 18 of the 29 items were thought to contribute to the overall measure of physical activity. These include Q.1-4, 7, 10-13, 15, 16, 17-21, 23 and 28. Summary scores are limited to these items.

A weighted average is derived across the 18 activities, with the MET values used as the weights. Thus, if the activities are indexed by “*k*”, a “MET-Weighted Summary Score” is computed as,

$$(\sum \text{MET}_k \times \text{Score}_k) \div (\sum \text{MET}_k)$$

The two summations are over the non-missing activities among the 18 items. This implies that the summary score can range from 1 to 10. If 25% or more (i.e., 5 or more) of the activities comprising the summary score are missing, then the summary score is set to missing.

These summary scores are treated as continuous random variables. Inasmuch as they are normally distributed, they are analyzed using techniques appropriate for Gaussian random variables.

### 5.2.2 CSA Activity Counts

An empirical measure of physical activity is being recorded using a CSA accelerometer device. The number of “counts” (i.e., the level of physical activity) in one-minute intervals is recorded by the monitor and stored. There is also a field signifying whether the monitor was worn or not during that interval. From there, the number of counts and the number of minutes worn are aggregated for each of 4 time periods in any day (midnight – 6 am, 6 am – noon, noon – 6 pm, and 6 pm midnight), to a total for that day, and to a grand total across the 3 days. Moreover, from the number of counts in any interval, an intensity level (“Light”, “Moderate”, “Hard”, and “Very Hard”) can be derived for that interval. The total number of minutes at each intensity level can be summed to 4 time periods in any day, to totals for that day, and to a grand total over the 3 days.

If the monitor was not worn for an entire six-hour time period, then the total count and total number of minutes in that time period (and therefore the average count) are set to missing in that time period and for all intensity levels in that time period. If no activity is recorded in any of the 4 time periods in any day, then the total count and total number of minutes (and therefore the average count) are set to missing for that day and for all intensity levels in that day. If for any intensity level, the number of minutes is 0 (but non-missing), then this signifies that no activity occurred at that intensity level in that time period, and the average is set to 0.

A girl is included in the CSA analyses if the belt was worn for at least one day (out of the three) with 800 minutes or more during the period from 6 am to midnight. If her data satisfy this criterion, then all data provided over the three days are included in all CSA analyses.

The “average activity count per minute” at any level is computed as (total number of counts aggregated to that level) ÷ (total number of minutes aggregated to that level), provided the number of minutes in the denominator is non-missing. Note that the aggregations are performed in the numerator and denominator first before the division is performed.

Thus, we have the following analytic variables:

- Average activity counts per minute in a 24-hour period, aggregated over the 3 days;
- average activity counts per minute in the period from 6 am to midnight, aggregated over the 3 days;
- average activity counts per minute in the period from noon to 6 pm, aggregated over the 3 days;
- number of minutes of moderate to vigorous activity from noon to 6 pm, aggregated over the 3 days; and,
- average activity counts per minute in moderate to vigorous activity from noon to 6 pm, aggregated over the 3 days.

These summary scores are treated as continuous random variables. Inasmuch as they are normally distributed, they are analyzed using techniques appropriate for Gaussian random variables.