

CENTERS FOR DISEASE CONTROL AND PREVENTION

# Youth Media Campaign Longitudinal Survey 2002-2004

## Data File User's Manual

April 2008



**Prepared for:**

U.S. Centers for Disease Control  
and Prevention  
1600 Clifton Road  
Atlanta, GA 30333

**Prepared by:**

**WESTAT**  
1650 Research Boulevard  
Rockville, Maryland 20850



## TABLE OF CONTENTS

<u>Chapter</u>	<u>Page</u>
1	INTRODUCTION ..... 1
	Background on the Youth Media Campaign ..... 1
2	SAMPLE DESIGN, WEIGHTING, AND IMPUTATION ..... 3
	Sampling ..... 3
	Weighting Procedures ..... 4
	Panel 1 2002 Full Sample Weights ..... 5
	Panel 1 2003 Full Sample Weights ..... 7
	Panel 1 2004 Full Sample Weights ..... 8
	Panel 2 2004 Full Sample Weights ..... 8
	Combined Panel 1 and Panel 2 2004 Full Sample Weights ..... 9
	Panel 1 Longitudinal Weights ..... 10
	Item Imputation ..... 10
	Panel 1 2002 and 2003 Item Imputation ..... 11
	Panel 1 2004 Item Imputation ..... 11
	Panel 2 2004 Item Imputation ..... 12
	Manual Imputation ..... 12
	Imputation of Whole Baseline Child Questionnaires ..... 12
	Imputation Flags ..... 13
3	DATA COLLECTION AND RESPONSE RATES ..... 15
	The YMCLS Interviews ..... 15
	Interview Content ..... 16
	Completed Interviews and Response and Coverage Rates for Panel 1 ..... 16
	Completed Interviews and Response and Coverage Rates for Panel 2 ..... 19
4	CONTENT AND ORGANIZATION OF THE DATA FILE SYSTEM ..... 21
	Naming Conventions ..... 21
	Values ..... 22

**TABLE OF CONTENTS—CONTINUED**

<u>Chapter</u>		<u>Page</u>
	Variables on the YMCLS Data File System .....	22
	Administrative Variables .....	23
	Questionnaire Variables .....	25
	Derived Variables .....	27
	Administrative Derived Variables .....	27
	Parent and Household Characteristic Derived Variables .....	27
	Child Characteristic Derived Variables .....	28
	Physical Activity Derived Variables.....	28
	VERB Awareness and Understanding Derived Variables .....	29
5	USING THE YMCLS FILE SYSTEM FOR CROSS-SECTIONAL ANALYSES .....	31
	Analyses Supported by the File System.....	31
	Analyses of Parent Data.....	31
	Cross-sectional File Layout .....	32
	Computing Point Estimates.....	32
	Estimating Variances .....	34
	Detailed Instructions for Custom Programming of Variance Calculations .....	36
	Detailed Instructions for Using WesVar for Variance Calculations.....	36
	Detailed Instructions for Using SUDAAN for Variance Calculations .....	42
6	USING THE YMCLS FILE SYSTEM FOR LONGITUDINAL ANALYSES .....	49
	Analyses Supported by the File System.....	49
	Longitudinal File Layout .....	49
	Computing Point and Variance Estimates .....	50
	Example of Longitudinal Analysis Using WesVar .....	50
	Example of Longitudinal Analysis Using SUDAAN .....	55
	REFERENCES .....	63

**LIST OF APPENDIXES**

<u>Appendix</u>		<u>Page</u>
A Master	Questionnaire.....	A-1

## TABLE OF CONTENTS

### LIST OF ATTACHMENTS

<u>Attachment</u>		<u>Page</u>
5-1	Output from WesVar Example .....	43
5-2	Output from SUDAAN Example .....	45
6-1	Output from WesVar Example .....	56
6-2	Output from SUDAAN Example .....	59

### LIST OF TABLES

<u>Table</u>		<u>Page</u>
2-1	Panel 1 and Panel 2 sample sizes, by child age in 2004 .....	9
3-1	Unweighted and weighted response rates, Youth Media Campaign Longitudinal Survey, Panel 1 baseline survey, 2002 .....	17
3-2	2003 and cumulative response rates including cases with imputed baseline child interviews, Youth Media Campaign Longitudinal Survey, Panel 1 .....	18
3-3	2004 and total overall response rates, Youth Media Campaign Longitudinal Survey, Panel 1 .....	18
3-4	Unweighted and weighted response rates, youth media campaign longitudinal survey, Panel 2 baseline survey, 2004 .....	20

### LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
5-1	Screen shot for creating a WesVar file .....	38

## LIST OF FIGURES—CONTINUED

<u>Figure</u>		<u>Page</u>
5-2	Screen shot for creating a table in WesVAR .....	39
5-3	Screen shot for viewing a table in WesVAR .....	40
5-4	Screen shot for viewing a table in TableViewer .....	41
6-1	Screen shot for creating a WesVar file .....	52
6-2	Screen shot for creating a regression in WesVAR .....	53
6-3	Screen shot for viewing a regression in WesVAR.....	54

## 1. INTRODUCTION

This manual provides documentation for the Youth Media Campaign Longitudinal Survey (YMCLS) data files from 2002, 2003 and 2004. The manual contains background information on the survey; a description of the sample design; an explanation of weighting and imputation; an account of data collection procedures and response rates; descriptions of the files and the variables in them; and directions for using the files.

### **Background on the Youth Media Campaign**

In June 2002, the Centers for Disease Control and Prevention (CDC) launched a national campaign to encourage 9- to 13-year-olds to be physically active everyday. Using the brand and tagline, “*VERB™ It's What You Do,*” the Youth Media Campaign combined paid advertising with school and community promotions and internet activities to disseminate its health-promotion messages.

Campaign planners identified 9- to 13- year-olds, their parents, and other influencers such as coaches and teachers as the three key audiences; the majority of campaign activity was directed at children. Targeted campaigns intended to augment the general market advertising were created for several racial/ethnic subpopulations.

The campaign employs a brand marketing strategy similar to that used by many commercial products. The brand, VERB, is intended to carry the health promotion message by associating itself with attributes with which the target audience desires to be associated. As members of the target audience seek to identify with VERB, they will incorporate the key campaign message, that they be more physically active, into their lifestyle. All forms of media (e.g., television, radio, print, on-line), as well as intercept marketing techniques and public events, were used. The majority of the advertising was placed on cable television networks with predominantly child audiences. During 2002–2003 the advertising was largely focused on developing awareness of and affinity for the brand. In year two, 2003–2004, the advertising sought to motivate the target audience to engage in physical activity.

To assess the impact of these efforts, Westat administered a telephone survey, the Youth Media Campaign Longitudinal Survey (YMCLS) to children and parents. The first panel of parent-child dyads was sampled through random -digit-dialing and computer assisted telephone (CATI) interviews were administered in spring 2002, prior to the launch of advertising. At that time, the children Panel 1 were aged 9–13, the target for the VERB campaign. Panel 1 was interviewed again in spring 2003 and

2004. In spring 2004, a second panel of parent-child dyads was sampled through random-digit-dialing and interviewed using C ATI. Children in Panel 2 were aged 9–13 in 2004. The same interview was administered to members of Panel 1 and Panel 2 in 2004.

Child interviews focused on participation in and attitudes toward physical activity. Participation in specific activities during the 7 days prior to the interview was measured, along with whether the activities had been organized, part of a team or structured lessons, or engaged in during their discretionary or free time. Children were also asked about hours of television viewing as well as school and community activities. A series of items measured children's attitudes toward physical activities. Subsequent to the 2002 interview, psychosocial items exploring normative social influences and self-efficacy were added, along with perceived body weight.

The YMCLS parent interview contained items measuring the importance of their child's physical activity, their confidence in influencing their child's involvement in physical activity, and their attitudes about child physical activity. Behavioral measures included doing physical activities with their children and their involvement as coach, spectator, or provider of transportation for their children's physical activities. Other items tapped perceived barriers to their children's participation in physical activities, including high cost or lack of time. Demographic characteristics, such as child's age and race/ethnicity, respondent's marital status and level of education, number of household members, and household income, were also collected in the parent interview.

For the 2003 and 2004 surveys, items measuring awareness of VERB were added to both the parent and child interviews. These items were designed to capture awareness without any prompting as well as recognition of the VERB brand when prompted by the interviewer. Children were also asked to relate their understanding of campaign messages. In 2004, children were asked to estimate the frequency with which they encountered VERB advertising on television or radio. Children were also asked if they recognized five brief descriptions of television advertisements, four of which were VERB advertisements, and one of which was not. All campaign-specific questions were asked at the end of the survey, after outcome data had been collected.



## 2. SAMPLE DESIGN, WEIGHTING, AND IMPUTATION

This chapter contains information on the sampling for the YMCLS. It also provides a description of the weighting and imputation procedures.

### Sampling

A list-assisted, random-digit-dialed (RDD) method (Casady and Lepkowski, 1993) was used in the YMCLS 2002 to select a sample of households with telephones that would comprise Panel 1. The list-assisted sampling method reduces the number of unproductive calls to nonworking or nonresidential numbers and produces a single-stage, unclustered sample. In list-assisted sampling, an equal probability random sample of telephone numbers is selected from all telephone numbers that are in 100-banks (i.e., 100 telephone numbers with the same first 8 digits of the 10-digit telephone number) in which there is at least one residential telephone number listed in the white pages directory. The sampling frame for the YMCLS was restricted to these 100-banks. To accommodate evaluation of communities in which more intensive advertising efforts were carried out, the sample was divided into two strata. In stratum 1, 92,760 telephone numbers were randomly selected from the frame at a sampling rate of 1 in 1,899 numbers. In stratum 2, 27,144 telephone numbers were randomly selected from the frame at a sampling rate of approximately 1 in 3,069 numbers.<sup>1</sup>

Panel 1 children were recontacted in 2003 and 2004; however, by 2004, most of the children in Panel 1 had aged out of the campaign target range, and a second panel of parents and children was randomly sampled in 2004 using the same list-assisted method but with a stratification scheme. Stratum 1 consisted of 100-banks for which African Americans and/or Hispanics represented at least 20% of the population. Stratum 2 covered the rest of the country. In stratum 1, 131,000 telephone numbers were randomly selected from the sampling frame at a rate of 1 in 866. In stratum 2, 73,000 telephone numbers were randomly selected from the sampling frame at a rate of 1 in 2,190, giving a total sample size of 204,000 for Panel 2. The total sample size and the oversampling of stratum 1 by a rate of 2.53 were designed to improve the precision of estimates for minorities, while balancing cost and overall precision constraints.

---

<sup>1</sup> Stratum 1 was over sampled by a factor of 1.62 to ensure sufficient sample size in the part of the national sample that was comparable to the high-dose communities.

Special procedures were used to identify as many non-working and business-only telephone numbers as possible prior to data collection. All other telephone numbers sampled in 2002 and 2004 were dialed. When residential telephone numbers were reached by interviewers, a screening interview was administered to a household member age 18 or older. Eligibility for the study was determined by whether any children aged 9–13 were living in the household. If there were one or two children in this age range, both children were selected into the sample with certainty. In households containing three or more children aged 9–13, two children were randomly selected into the sample. A parent/guardian interview was conducted in households containing an eligible child or children prior to conducting the interview with the sampled child or children. If the screener respondent was a parent of the sampled child, he or she was asked to respond to questions about “the physical and social activities” that the child participates in or to identify the resident parent or guardian who could answer those questions. Just over 70 percent of the parent respondents were mothers. The parent interview was always administered before the child interview. More information on data collection procedures is in Chapter 3.

## **Weighting Procedures**

**Cross-sectional** weights were developed for Panel 1 in 2002, 2003, and 2004 and for Panel 2 in 2004. The weighting procedures took into consideration different probabilities of selection, survey nonresponse, and undercoverage of age-eligible children—both in nontelephone households and other households. Cross-sectional weights are appropriate for producing separate estimates for each year of the survey, and for estimating *net* change over time. More information on cross-sectional analyses and how to choose the appropriate weights depending on the panel and/or population of interest is in Chapter 5.

**Combined cross-sectional weights** were created for Panel 1 and Panel 2 cases completed in 2004. These weights support analysis of children ages 9 to 15, the combined ages of children in both panels.

In 2003, special procedures were instituted for Panel 1 to improve the robustness of the sample. All cases with a completed parent interview in 2002 were recontacted in 2003; therefore, so many children responded in 2003 even though they had not responded in 2002. In order to maximize the statistical power to detect change while minimizing bias, these converted cases received positive (greater than zero) 2003 cross-sectional weights. However, these cases do not have positive cross-sectional weights for 2002.

**Longitudinal** weights were also developed in a process that closely paralleled the process for the creation of cross-sectional weights in terms of nonresponse adjustment and adjustment for undercoverage. The Panel 1 2004 longitudinal data file contains 2004 longitudinal weights that will enable longitudinal analysis of cases completed in 2002, 2003, and 2004. Cases completed by the child in 2003 but imputed in 2002 have positive weights on this file.<sup>2</sup> More information on longitudinal analysis is in Chapter 6.

**Replicate weights** were developed for each panel in each survey year in order to produce variance estimations. These weights are described in Chapters 5 and 6, which contain guidelines for the use of all weights.

The following sections of this chapter explain how the components of the full sample weights were calculated and describe the procedures for item imputation.

### **Panel 1 2002 Full Sample Weights**

Cross-sectional base weights for Panel 1 2002 were constructed in stages starting with household-level weights, then intermediary parent-level weights, and finally child-level weights. Residential numbers that were contacted were screened to identify households with children eligible for an interview. Household level weights were calculated to adjust for probabilities of selection and to serve as a basis for computing person-level weights. First, a household base weight was developed to account for the RDD sampling of telephone numbers. This weight was adjusted for screener nonresponse and then adjusted for households that had more than one residential telephone number, hence more than one chance of being included in the sample.

Screener nonresponse adjustment cells were formed according to basic telephone-exchange-level information (percent population of age 0-17, average rent, median income, and metropolitan status) and whether or not the address had been obtained through matching with the telephone number and the household mailed a pre-survey notification letter. For each cell, a separate nonresponse adjustment factor was calculated and applied to the household weights of respondents. Such a nonresponse adjustment can reduce the bias due to nonresponse to the extent that households within a cell tend to provide similar answers to survey questions.

---

<sup>2</sup> Special procedures were used to impute the missing child data from 2002. Please see the section, Imputation of Whole Child Baseline Cases, below.

A special adjustment is needed prior to the nonresponse adjustment because it is not possible to determine whether some telephone numbers are residential or not. (See also Chapter 3, response rates.) This occurs when there are only ring-no-answers on repeated calls or when repeated calls reach only an answering machine. Twenty-seven percent of ring-no-answer calls were estimated to be residential (based on calls made to the telephone company business offices in the 1994 National Immunization Survey, see Shapiro et al., 1995). Sixty-three percent of answering machine calls were estimated to be residential (as estimated in the 1995 National Household Education Survey; see Brick and Broene, 1997). Thus, the weights for ring-no-answer calls were multiplied by 0.27, and the weights for answering machine calls were multiplied by 0.63.

To be classified as a screener respondent, a sample household had to provide a count of age-eligible children. Nonresponse after that point in the data collection process was considered either parent or child nonresponse. Since there was no sampling of eligible parents, the parent base weight is equal to the household weight. So to create parent-level weights, the household weights were adjusted for parental nonresponse to the extended interview. Nonresponse cells were created using the number of children aged 9–13 in the household and the relationship of the parent to the child.

Child-level base weights were created from the intermediary parent weights in the final stage of weighting for the baseline year. The first adjustment at this step was to account for the subsampling of children within households that contained more than two age-eligible children. Next, nonresponse cells for child nonresponse were created using characteristics from the parent interview found to be most associated with child-level nonresponse. For this process, a categorical search algorithm called CHAID (Chi-squared Automatic Interaction Detector) (SPSS, 1993) was used. The characteristics were household income, the frequency with which adult members of the household engaged in physical activities with the child, the extent to which the parent believed his/her child's participation in physical activities was limited due to expense, and whether or not the child was African American.

After adjusting for baseline nonresponse, the child weights were further adjusted for undercoverage by a procedure called “raking,” also known as iterative post-stratification (Särndal, et al., 1992). With this procedure, the weights were controlled to known marginal population totals. Panel 1 child weights were controlled to 2002 Current Population Survey (CPS) totals for children aged 9–13 by race/ethnicity and by age crossed with gender (Deming and Stephan, 1940). After raking, a few very large weights were trimmed. The trimming resulted in estimates that did not agree exactly with control totals, so the raking procedure was repeated to produce the Panel 1 final full sample weights (variable name P1\_CS\_WGT0).

## Panel 1 2003 Full Sample Weights

To create weights for the Panel 1 2003 survey, the final baseline child weights were set aside since they equaled zero for all nonrespondent children to the 2002 survey who completed 2003 interviews. Instead of adjusting those weights for attrition, the parent baseline weights were adjusted for attrition in parent cooperation, and these attrition-adjusted parent weights were then adjusted for the original child sampling in 2002 and child nonresponse in 2003.

First, there was an adjustment for nonresponse of parents to the 2003 interview. Nonresponse cells were determined through application of the CHAID algorithm. The most important characteristics used to define nonresponse adjustment cells for parents were parents' education, child's race, percent of whites in area/exchange, and percent of Hispanics or Latinos in area/exchange.

The next adjustment accounted for the subsampling of children within those households containing more than two age-eligible children. The weight was then adjusted for nonresponse to the 2003 child interview. Children who were interviewed in the baseline year were adjusted separately from those who were not interviewed in the baseline year because of the richer set of covariates available for the adjustment. For each group, nonresponse cells were determined through application of the CHAID algorithm. The most important characteristics used to define cells for children interviewed in the baseline year were whether or not the child belonged to school groups, child's age, parent's agreement with the statement, *There are plenty of opportunities for my child to participate in activities where I live*, child's answer to the question, *Have you heard, read, or seen any ads with the slogan "Just do it"?*, and percent of whites in area/exchange. The most important characteristics used to define cells for children not interviewed in the baseline year were parent's confidence in influencing child's free-time physical activities, how often parent attended events related to child's activities, and percent of whites in area/exchange.

After the attrition adjustments, the 2003 weights were controlled to 2003 CPS totals by race and by age crossed with gender using raking in the same manner as for the 2002 weights in order to correct for undercoverage. After raking to these controls, a few very large weights were trimmed. The trimming resulted in estimates that did not agree exactly with control totals, so the raking procedure was repeated to produce the final full sample weights.

### **Panel 1 2004 Full Sample Weights**

For the Panel 1 survey in 2004, nonresponse adjustments were applied on the 2003 final cross-sectional weights to adjust for the 2003–2004 parent and child attrition. The nonresponse adjustment was performed in two sequential steps. The first step was to adjust the 2003 child weights for parent attrition. The parent and child characteristics collected in the 2003 interviews were used to form the nonresponse adjustment cells. The most important characteristics used to define cells for 2004 parent nonresponse were parent's education level, whether or not child did running/jogging, whether the other parent lives in the household, child's level of organized activity, and whether or not the child played soccer yesterday.

The second step was to adjust for child attrition in the 2004 survey, and this adjustment was only applied to those children whose parents completed the 2004 interview. The parent characteristics collected in the 2003 and 2004 followup interviews and child characteristics collected in the 2003 interviews were used to form the nonresponse adjustment cells. The most important characteristics used to define cells for 2004 child nonresponse were child age; child's response to the statement, *Usually there is no one around for me to do physical activities with*; child's level of free-time physical activity; and parent's response to the question, *How often do you set limits on the amount of time [child] plays video games*.

After the 2004 attrition adjustments, the Panel 1 weights were controlled to 2004 CPS totals for children ages 11 to 15 by race and by age crossed with gender. After raking to these controls, a few very large weights were trimmed. The trimming resulted in estimates that did not agree exactly with control totals, so the raking procedure was repeated to produce the final Panel 1 full sample weights.

### **Panel 2 2004 Full Sample Weights**

Panel 2 2004 weights were developed in the same manner as the Panel 1 2002 weights, that is, they were constructed in stages starting with household-level weights, then intermediary parent-level weights, and finally child-level weights. The initial nonresponse adjustment procedure for Panel 2 was the same as that for Panel 1. However, to address the issue of missing values for the parent variables for forming the nonresponse adjustment cells, a preliminary imputation step was added prior to running the CHAID procedure. The most important characteristics used to define cells for Panel 2 baseline child nonresponse were household income, telephone-exchange-level attributes (percent black, percent white, median home value, and median income), child age, parent's education level, importance of child's

unorganized physical activities to parent, and parent's marital status. With the same raking procedure used for Panel 1, the Panel 2 baseline nonresponse-adjusted child weights were controlled to the 2004 CPS totals for children aged 9–13 by race/ethnicity and by age crossed with gender. After the raking, a few very large weights were trimmed. The trimming resulted in estimates that did not agree exactly with control totals, so the raking procedure was repeated to produce the Panel 2 full sample weight (P2\_CS\_WGT0).

### Combined Panel 1 and Panel 2 2004 Full Sample Weights

Panel 1 and Panel 2 2004 survey data can be pooled to increase sample size and thus the robustness of a cross-sectional or change-over-time analysis. (Table 2-1 shows the age and sample size distribution for the two panels in 2004.) A compositing factor was used to create combined 2004 cross-sectional weights. In choosing the compositing factor, three related issues, sample size, variance, and time-in-sample bias, were examined.

**Table 2-1.—Panel 1 and Panel 2 sample sizes, by child age in 2004**

Age in 2004	Number of respondents		
	Panel 1	Panel 2	Panel 1+ Panel 2
9.....		979	979
10.....		1,065	1,065
11.....	440	1,080	1,521
12.....	441	952	1,393
13.....	493	1,101	1,594
14.....	468		468
15.....	414		414
Total.....	2,256	5,177	7,434

Based on all considerations, a compositing factor was devised such that a smaller (0.3) weight was given to Panel 1 data and a larger (0.7) weight to Panel 2 data to create the combined Panel 1-Panel 2 full sample weights (P1P2\_CS\_WGT0).

### Panel 1 Longitudinal Weights

Longitudinal weights were developed for each round of the Panel 1 surveys, with a process that closely paralleled the process for developing cross-sectional weights in terms of nonresponse

adjustment and adjustment for under coverage. The only difference between the cross-sectional and longitudinal weights was in the control totals used in each round of raking adjustment. The 2002 longitudinal weights were controlled to the 2000 Census totals by race and by age crossed with gender.

In 2003, the longitudinal weights were developed by first adjusting for 2002–2003 parent and child nonresponse and then by raking, again controlled to the 2000 Census totals by race and by age crossed with gender. As noted above, all able-bodied children who completed 2003 interviews, including those children with missing baseline child interviews, were eligible to receive a 2003 longitudinal weight.

The longitudinal weights for 2004 were developed in a similar manner as that for 2003, by first adjusting for 2003–2004 parent and child nonresponse, and then by raking. However, the nonresponse-adjusted 2004 child weights were raked to the 2003 longitudinal weight totals by a number of 2003 variables that were important in the analysis, (including age, sex, race/ethnicity, parent education, household income, sum of the total number of days participated in all organized activities, and sum of the total number of days participated in all free time activities, etc.). Since the 2004 cases were a subsample of 2003 completed cases, agreement on critical variables reduces bias and variance and improves face validity, and the sample-based raking achieved this.<sup>3</sup>

## **Item Imputation**

In the YMCLS, as in most surveys, the responses to some data items were not obtained for all interviews. Some respondents do not respond for various reasons. The item response rate was very high for most of the data items collected in the YMCLS. The median item response rate for imputed items from the first parent and child interviews was 98 percent. Only household income had a significantly higher refusal rate, with about 9 percent of responses not obtained at baseline in Panel 1 and 12% in Panel 2. In subsequent interviews, the median item response rate was higher, about 99.8 percent. Typically, parent and child response rates to the question, *What is the name of the message or advertising?* were relatively much lower. That is, a greater percentage of respondents said “don’t know” when asked that question.

---

<sup>3</sup> Raking to 2002 critical variables was not done in 2003, because some cases not completed in 2002 were included in the 2003 sample. Thus, 2003 was not a subsample of 2002 cases.



### **Panel 1 2002 and 2003 Item Imputation**

A hot-deck procedure was used to impute most missing responses (Kalton and Kasprzyk, 1986). All cases were sorted into cells defined by characteristics of households or respondents that were likely to be associated with differences in response propensities. The characteristics varied somewhat based on the item that was to be imputed but included age of child, gender, household income, number of adults in the household, marital status of responding parent, race of child, and number of children in the household. Donors for missing data were randomly selected within the same cells that the recipient cases were located, and the completed data item from the donor case was used to fill in the missing data item. Typically a donor was used only once to supply missing data, and imputation was done separately for each question, permitting different donors for different questions. However, in some cases, a set of answers were from the same person in order to obtain a consistent response pattern. For example, if a child had not responded to the related statements, *I should probably do more physical activities than I do* and *I'm too busy to do more physical activities than I do*, a single donor was used to impute responses for these statements.

### **Panel 1 2004 Item Imputation**

Different procedures were used to impute the missing responses for Panel 1 2004 ordinal and unordered questionnaire items. For each ordinal item that needed imputation, a regression model was fit using a stepwise ordinary least squares selection procedure. Critical baseline variables, environmental variables, current questionnaire variables were used as eligible predictors for this modeling step. The resulting predicted values from the regression model were used in a nearest neighbor imputation procedure where each case with a missing value was matched to the nearest complete case and distance was defined in terms of squared differences in predicted values for the variable being imputed. This procedure was designed to maintain both longitudinal and cross-sectional relationships. A hot-deck procedure was used to impute unordered items within the following soft boundary cells: age, urbanicity, and the presence of cable television in the home. For each item imputed in these imputation procedures, hard boundaries were defined by preceding items that either controlled the skip pattern for the imputed variable or were contained in the same sequence of questions. Items were imputed in the order defined by the sequence in the questionnaire, and items pertaining to the parent interview were imputed prior to those associated with the child interview.

## **Panel 2 2004 Item Imputation**

As was used for Panel 1, a hot-deck procedure was used to impute the missing Panel 2 2004 responses (Kalton and Kasprzyk, 1986). All cases were sorted into cells defined by characteristics of households or respondents that were likely to be associated with differences in response propensities. The characteristics varied somewhat based on the item that was to be imputed but typically included age of child, gender, household income, number of adults in the household, marital status of responding parent, race of child, and number of children in the household. Donors for missing data were randomly selected within the same cells that the recipient cases were located, and the completed data item from the donor case was used to fill in the missing data item. Typically, a donor was used only once to supply missing data, and imputation was done separately for each question, permitting different donors for different questions. However, in some cases, a set of answers were imputed from the same person in order to obtain a consistent response pattern.

After item imputation had been completed, a few additional changes were made to variables that had not been previously fully imputed. A simplified hot-deck procedure similar to that described above was used to impute those cases.

## **Manual Imputation**

In all survey years, manual imputation was conducted on a small number of variables for which random imputation procedures would not have been appropriate, that is, decisions were made based on logical considerations rather than through an automated procedure. An example is the case of children who answered “yes” when asked *In the past 7 days, since last [DAY], did you do any physical activities?* but “don’t know” when asked to name an activity they had done, despite the fact that a definition of physical activity had been provided to them (“...got your body moving like sports...playing actively with your friends”). Reasoning that children who said they had done a physical activity but could not name any activity they had done likely misinterpreted the first question, that response was reset to “no” and the following question was set to *inapplicable*.

## **Imputation of Whole Baseline Child Questionnaires**

In 278 households in the Panel 1 sample, the parent interview had been conducted in both 2002 and 2003, while the child interview had been conducted in only 2003. In order to make the best use of the data for these households, the 2002 child interview was imputed. The imputation involved

identifying children who had data in 2002 and who were as similar as possible to children who were missing baseline data. The similarity was assessed by responses in the parent interviews, both 2002 and 2003, and responses in the 2003 child questionnaire.

The imputation process involved several steps. First, five key outcome measures were identified. Next, predictive models were developed for the five outcome measures. These models were then used to predict values for these measures for children who were missing baseline data. Clustering algorithms were run on the predicted values to form imputation classes. Finally, hot-deck imputation was used to find donors and impute for the missing baseline data within the imputation classes. Once a suitable donor had been identified, all fields from the missing baseline child questionnaire (i.e., the “whole case”) were transferred from the donor to the recipient.

By imputing the entire set of baseline data from one child, the covariance structure was preserved in a way that could not have been achieved if the missing data had been imputed one variable at a time. In the whole-case imputation approach, the covariance structure is maintained among *all* imputed variables. In addition, through the use of 2003 child responses and 2002 and 2003 parent responses in the modeling that was used to form the imputation classes, at least some of the most important features of the covariance structure across years and a cross family members were preserved. In order to validate the whole-case imputation procedures, the prediction models were refit with and without the imputed outcome values.<sup>4</sup> The difference in standard errors between the “with” and “without” imputed values relative to the “without” standard error were small but nearly all negative, reflecting the artificial reduction in standard errors through the increased sample size due to the additional cases made available through whole-case imputation.

### **Imputation Flags**

Each variable that contains imputed data has a corresponding imputation flag variable in the data file. The values of the imputation flag variables indicate the type of imputation, as follows:

- 0 = The referent variable was not missing and therefore not imputed
- 1 = The referent variable was imputed using the original hot-deck procedure
- 2 = The referent variable was imputed using the simplified hot-deck procedure
- 3 = The referent variable was imputed during whole-case imputation
- 4 = The referent variable was imputed manually

---

<sup>4</sup> For validation, simple linear models were used rather than the more complex generalized models that had been used in the actual prediction.

### 3. DATA COLLECTION AND RESPONSE RATES

A general description of the interview method and content is contained in this chapter. A brief discussion of methods for calculating response rates and information about YMCLS response rates is also presented. More detailed information is contained in the methodology appendixes in the reports of findings from the YMCLS (Potter, et al., 2004 and Potter, et al., 2005).

#### **The YMCLS Interviews**

YMCLS interviews were administered using computer-assisted telephone interviewing (CATI) technology. CDC's Institutional Review Board approved all procedures (Protocol number 3450 for all survey years). OMB approval was also obtained (OMB number 0920-0546 for Panel 1 2002, OMB number 0920-0587 for Panel 1 in subsequent years, OMB number 0920-0646 for Panel 2). Interviews were conducted from mid-April to mid-June each year.

Prior to the Panel 1 2002 and Panel 2 2004 data collections, the sampled telephone numbers were sent to a professional matching service to obtain addresses. Letters informing potential respondents about the survey were mailed to all sampled telephone numbers for which addresses were obtained. The letters contained the telephone numbers of appropriate CDC personnel and a toll-free number for Westat.

As noted in Chapter 2, interviews were attempted in 2003 with all Panel 1 telephone numbers where a parent interview had been completed in the initial year, even if the associated child interview had not been completed. Prior to the start of 2003 data collection, a reminder letter with an incentive payment of \$5 was mailed to those households for which addresses were available. In 2004, an information letter was mailed to 2003 respondents, but it did not include an incentive. In each survey year, if children had moved, they were contacted at new telephone numbers where possible, and a different parent or guardian was interviewed in 2003 if caretaking responsibilities had changed since the previous interview. A variable on the files, MKACNT, indicates whether the original responding parent changed. If calls to Panel 1 telephone numbers did not reach previously sampled households, tracing procedures were followed in an attempt to locate respondents at new telephone numbers.

A parent was always interviewed before attempting an interview with the child. Following completion of the parent interview, the responding parent was asked for permission to interview the sampled child. In all years, if an initial refusal was given by a parent respondent, another call was placed

in an attempt to gain cooperation. In 2002, children who completed interviews were mailed thank-you payments of \$5. In subsequent years, a thank-you payment of \$10 was mailed to children after both their responding parents and they completed interviews.

The YMCLS interviews were administered in English and Spanish. Bilingual interviewers called back into households when other interviewers indicated that the respondent did not speak sufficiently fluent English to understand and answer the questions easily. When bilingual interviewers initially encountered households in which Spanish was spoken, the respondent could choose to converse in either English or Spanish. The variables `PARNENSP` and `CHILDENSP` in each file indicate whether the interview was conducted in English or Spanish.

Some children were on school holiday when they were interviewed, either because they were on spring break or, more rarely, the school year had ended, with consequent implications for the levels of physical activity that they reported for the week prior to the interview. The variable `C1` contains this information.

### **Interview Content**

Interview content was similar in each survey year. (Please refer to the master questionnaire for details.) The 2002 Panel 1 parent interview contained items measuring the importance with which respondents regarded their children's physical activity level, their confidence in influencing their children's involvement in physical activity, and their attitudes and beliefs about physical activity for children in general. Behavioral measures included doing physical activities with their children and their involvement as coach, spectator, or provider of transportation for their children's physical activities. Other items tapped perceived barriers to their children's participation in physical activities, including high cost or lack of time. Demographic characteristics, such as child's age and race, respondent's marital status and level of education, number of household members, and household income, were also collected in the first parent interview.

The focus of the child interview was on children's participation in physical activities and their attitudes, beliefs, and self-efficacy about physical activity. Participation in specific activities during the 7 days prior to the interview was measured, along with whether the activities were organized (part of a team or structured lessons) or activities the children had engaged in during their free time. Children were also asked about hours of television viewing. Several items measured children's attitudes toward physical activities.

For the 2003 and 2004 interviews, items measuring recall of the VERB campaign were added to both the parent and child interviews. These items were designed to capture both unprompted recall as well as recognition of the VERB brand when it was mentioned. Children were also asked to relate their understanding of campaign messages. Items measuring social norms were added in 2003. Access to cable television and questions related to rules for children's viewing TV and playing video games were added to the parent interview. Questions on intention to participate in physical activities with their children or support children's participation in physical activities were also added to the parent interview, as was an item measuring parents' participation in physical activities during the 7 days prior to the interview. In 2004 additional attitude items were added to the child questionnaire. Other items in both parent and child surveys were identical to those fielded in first interview; however, demographic information was gathered only from first-time parent respondents to the YMCLS.

### **Completed Interviews and Response and Coverage Rates for Panel 1**

In 2002, 3,114 parent-child interviews were completed in Panel 1. Of the 2,518 households interviewed, 1,930 had a single child responding and 592 had two children responding (1,184 total children). Thus, approximately 24% of Panel 1 households had two children interviewed. In all but one of the two-child households, one parent answered for both children. In 2003, 2,729 parent-child interviews were completed (1,689 from single-child households and 1,040 from two-child households). This number includes 278 cases in which the parent but not the child had completed the 2002 interview. In 2004, 2,256 parent-child interviews were completed in Panel 1.

Nonresponse is a potential source of nonsampling error in any survey. As discussed in Chapter 2, nonresponse adjustments were made at each stage of the weighting process, so that the weights of respondents who are "similar" to nonrespondents were increased. This weighting tends to reduce, but does not completely eliminate, the bias due to nonresponse, since nonresponding households, parents, and children are likely to be systematically different from respondents on some characteristics for which matching of respondents and nonrespondents is impossible.

The response rate is the percentage of interviews completed, taking all stages of the interview into account. In the YMCLS, households with eligible children were identified through a brief screening interview and children were sampled for the survey. Then parents of sampled children were administered an extended interview, and finally, sampled children were administered an extended interview. The overall response rate for the YMCLS takes into consideration all stages of the interview.

It is calculated as the screener response rate times the parent interview response rate times the child interview response rate.

A household response rate is an estimate of the proportion of dialed residential households that completed the screening interview; thus, it is also called the screener response rate. The first step is to establish the number of residential numbers in the sample. Most numbers can be classified as residential, business, or nonworking when they are dialed. However, in order to calculate a screener response rate, it is necessary to estimate residential status for the proportion of calls that are unanswered or answered only by an answering machine despite repeated attempts. One generally accepted approach to this estimation is called the “business office” method. The residential proportion of unanswered calls was established in the National Immunization Survey (NIS) in which the telephone company business office calculated the residential rate for a sample of unanswered numbers from this RDD sample. Using this method, 27% of numbers never answered are assumed to be residential (Shapiro et al., 1995). Research from the 1995 National Household Education Survey estimated that 63% of numbers answered only by answering machine are residential (Brick and Broene, 1997). The screener response rate for the Panel 1 2002 survey was calculated using the business office estimates of residential numbers that were never answered and estimates from the National Household Education Survey for numbers answered by answering machine.

The screener, parent interview, and child interview response rates for the Panel 1 2002 survey are presented in Table 3-1. Both unweighted and weighted response rates are shown. The unweighted response rate is calculated as a simple ratio where all telephone numbers are counted equally in calculating the rate. The weights that reflect the probability of selection are used in the calculation of weighted response rates. Thus, weighted response rates provide a better measure of the *effective* response rate for the survey. The overall response rate in the baseline year, the product of the screener, extended parent interview, and extended child interview rates, was 43%.

**Table 3-1.—Unweighted and weighted response rates, Youth Media Campaign Longitudinal Survey, Panel 1 baseline survey, 2002**

Response rate	Unweighted response rate	Weighted response rate
Screener.....	60.5%	60.9%
Extended interview—parents.....	88.0	87.3
Extended interview—children.....	81.2	80.5
Overall.....	43.2	42.8

As was mentioned, for the Panel 1 2003 survey, interviews were attempted with all telephone numbers where a parent interview had been completed in the initial year, even if the associated child interview had not been completed. Data for 2002 were imputed on a whole-case basis for those children who did not complete interviews in the baseline year and were converted to respondents in the 2003 survey. These child cases were included as respondents in 2003 response rate calculation.

Table 3-2 contains the unweighted and weighted response rates for the Panel 1 2003 survey, including all cases where the parents completed baseline interviews in 2002 whether or not the child interview was completed. The overall 2003 response rate of 72% is the product of the 2003 parent response rate and the 2003 child response rate. The cumulative 2003 response rate is the product of the 2002 screener and parent response rate and the overall 2003 response rate. This response rate accounts for all stages of nonresponse in both the 2002 and 2003 surveys and is in the 38% to 39% range.

**Table 3-2.—2003 and cumulative response rates including cases with imputed baseline child interviews, Youth Media Campaign Longitudinal Survey, Panel 1**

Response rate	Unweighted response rate	Weighted response rate
2003 extended interview—parents .....	75.8%	75.1%
2003 extended interview—children.....	95.5	95.3
2003 overall.....	72.4	71.6
Cumulative (2002 and 2003).....	38.5	38.1

Table 3-3 shows the unweighted and weighted response rates for the Panel 1 survey in 2004. The overall 2004 response rate is the product of the 2004 parent response rate and the 2004 child response rate and ranges from 82% to 83%. The total overall response rate is the product of the cumulative 2003 and 2004 response rates. This response rate accounts for all stages of nonresponse in the 2002, 2003, and 2004 surveys and is in the 31% to 32% range.

**Table 3-3.—2004 and total overall response rates, Youth Media Campaign Longitudinal Survey, Panel 1**

Response rate	Unweighted response rate	Weighted response rate
2004 extended interview—parents .....	85.4%	84.4%
2004 extended interview—children.....	97.1	97.2
2004 overall.....	82.9	82.1
Total overall (2002, 2003, and 2004) .....	31.9	31.2



In addition to nonresponse, there is a level of undercoverage in most surveys, that is, some people who live in an interviewed household are not reported and some households are missed entirely. Undercoverage of children can occur in the case of joint custody, where neither the mother nor father reports the child as a household member. Furthermore, a screener respondent can intentionally fail to report a child living in the household for various reasons. In a survey like this, where it may be clear to a respondent that the length of the interview can be reduced by stating that there are no children aged 9–13 in the household, there can be a high level of intentional underreporting (Judkins, Chu, DiGaetano, & Shapiro, 1999). Undercoverage also occurs for other reasons, such as children living in households with no telephone or in homeless families.

Bias occurs to the extent that undercoverage is high and the characteristics of missed children are different than those of other children in the target population. The percentage of children aged 9–13 living in households with no telephone is very low, only 3.6%, according to data collected in the March 2003 Current Population Survey. The overall, weighted coverage rate for children 9 to 13 is estimated to be 75% for Panel 1. For African American children who were not Hispanic, the coverage rate was about 59%. Weighting procedures were used to help correct for undercoverage. For instance, by controlling estimates to race/ethnicity totals in the weighting process, the bias due to a lower coverage rate for African Americans was mitigated. However, to the extent that African American children (and their parents) who did not have a chance to be represented in the survey react differently to the media campaign than do other African American children, survey results are biased. Female children ages 9 and 13 also had a lower coverage rate than other age/sex groups.

## **Completed Interviews and Response and Coverage Rates for Panel 2**

In 2004, 5,177 Panel 2 parent-child interviews were completed. The overall response rate for the YMCLS Panel 2 survey in 2004 was calculated in the same way as the Panel 1 2002 survey by taking into consideration all stages of the interview. The screener, parent interview, and child interview response rates for the Panel 2 2004 survey are presented in Table 3-4. Both unweighted and weighted response rates are shown.

**Table 3-4.—Unweighted and weighted response rates, Youth Media Campaign Longitudinal Survey, Panel 2 baseline survey, 2004**

Response rate	Unweighted response rate	Weighted response rate
Screener .....	58.8%	59.9%
Extended interview—parents .....	84.9	84.9
Extended interview—children .....	88.0	88.4
Overall .....	43.9	44.9

The overall, weighted coverage rate for children ages 9 to 13 in Panel 2 is estimated to be 74%, which was very close to the Panel 1 overall coverage rate of about 75%. There was a pattern of undercoverage by race/ethnicity similar to that of Panel 1, with a coverage rate for African American children who were not Hispanic of 55% in Panel 2, which was substantially lower than other groups. Weighting procedures similar to those used for Panel 1 were used to help correct Panel 2 for undercoverage.

#### 4. CONTENT AND ORGANIZATION OF THE DATA FILE SYSTEM

This section describes the content of the YMCLS data file system delivered with this documentation. The **YMCLS Cross-sectional File** (`ymcls_cs_020304.sas7bdat`) contains a *separate record for each completed child interview* between 2002 and 2004, except cases where the child was disabled. The file includes data for both Panel 1 and Panel 2 in 2004, as well as appropriate cross-sectional weights. Records for the three different years are “stacked” on top of each other so that, for example, there will be separate records for 2002, 2003, and 2004 for a child who completed interviews in all three years. The numbers of complete cases for Panel 1 by year are: 3,114 (in 2002), 2,729 (in 2003), and 2,256 (in 2004). The file contains 5,177 complete cases for Panel 2 in 2004. The **YMCLS Longitudinal File** (`ymcls_long_020304.sas7bdat`) has one record for each of the 2,256 Panel 1 youth who completed interviews in 2002, 2003, and 2004. Each record contains data for 2002, 2003, and 2004 and appropriate longitudinal weights. In both cross-sectional and longitudinal files, the parent and child interview data for each case are contained on one record.

For this file delivery, four separate codebooks were created to accompany the cross-sectional file. The four codebooks show frequency distributions for Panel 1 2002, Panel 1 2003, Panel 1, 2004, and Panel 2 2004. One codebook accompanies the longitudinal file. It contains the frequency distributions for all variables for Panel 1 2002, 2003, and 2004.

The codebooks display the following information for all the variables in the data sets: the **variable name**, the unique identifier for each variable in the data file; the **variable label**, a short description of the variable; **comments** about the variable; the **response** options for the variable; the **code** for each response option; the **unweighted frequency and percent** for each response option; and the appropriate **weighted frequency and weighted percent** for each response option.

#### Naming Conventions

All variable names in the longitudinal file have a suffix indicating the year of data collection, for example, “\_02” (underscore 02) is appended to the names of variables containing data collected in 2002. In the cross-sectional file, the suffixes have been removed to create the stacked file; analysts can select data for specific years by subsetting on the variables `PANEL` and/or `YEAR`. (Instructions are provided in Chapter 5.)

Most questionnaire variable names are identical to the question number. (Please see master questionnaire.) Variable names that begin with the letter “P” are from the parent questionnaire; a “C” denotes child variables<sup>5</sup>. An underscore following the initial “P” or “C” indicates that the item was added to the questionnaire after 2002.<sup>6</sup> For instance, P13A (whether parents attended events related to their child’s physical activities) was asked in 2002, but P\_13C (intention to lead a group or coach child in the future) was asked beginning in 2003. (Please see the questionnaire crosswalk for the variables asked in each survey year.) An underscore at the beginning of a variable name indicates a derived variable, for example, \_EXINDEX4. Imputation flag variables begin with the letters “IMP.” The names of weighting variables indicate the Panel and the type of weight. For instance, P1\_CS\_WGT0 is the full sample, cross-sectional weight for Panel 1 and P1P2\_CS\_WGT1-60 are the replicate weights to be used with the combined Panel 1 and Panel 2 2004 data on the cross-sectional file. (More details on using the weight variables are in Chapters 5 and 6.)

## Values

The codebooks contain the response categories for all variables and their values on the data files. A value “-1” for any variable indicates that the question for that case was part of a legitimate skip. For example, if the child respondent answered that he or she had not *seen or heard any messages for kids getting active* at C\_15, then the value for C\_16, *What is the name of the message or advertising?* will be -1 because C\_16 was not asked of children who had not seen or heard messages about physical activity. Note that values of “-1” are not included in the weighted percent column in the codebooks.

Variables that were fielded in some years but not in others are system missing (as indicated by a .) for those years for which the data are unavailable. Because of item imputation, no other data are missing.

## Variables on the YMCLS Data Files

The data items are listed in the following order in the codebooks: administrative variables (system variables and variables attached from outside data sources), questionnaire variables (parent and

---

<sup>5</sup> AC4Z1-AC4Z5, AC5Z1-AC5Z5, and AC5A1-AC5A5 are also child variables. They represent the array of possible answers to variables C4 and C5.

<sup>6</sup> There are four exceptions to this, C8AH, C8AI, C8AJ, and C8AK. As indicated on the crosswalk and master questionnaire, these variables were added in 2004, but the underscore convention was not followed.

child variables), derived variables, imputation flag variables, and weighting and variance estimation variables.

### **Administrative Variables**

The following variables contained in the YMCLS data files were based on data collected in the screening interview, CATI administrative variables, or characteristics associated with the respondent's telephone exchange or geographical location. The comments line in the codebook also identifies the origin of these variables.

**PANEL** identifies the case as belonging to Panel 1 or Panel 2.

**YEAR** is the 4-digit year of the survey.

**CHLDID** is a 12-digit child identification number. The first 8 digits of **CHLDID** indicate the household, thus identifying two children households. Digits 9 and 10 are the number of the sampled child; this number was assigned in the screening interview. Digits 11 and 12 are always 01 in these files, indicating a completed case.

**SEX** of the child was ascertained in the YMCLS screener administered at the first interview for the panel. It appears on the data file for each survey year.

**TOTCHLD** is the CATI count of the number of children in the household eligible for sampling. In households with more than two eligible children, two were randomly sampled for the YMCLS.

**RELATE** contains the relationship of the parent respondent to the sampled child. This information was collected in the screener. If adult respondent changed in any survey year, the value of RELATE will differ in survey years.

The variable **MKACNT** denotes those cases in which the adult respondent to the first survey for the panel and the adult respondent to any subsequent survey are not the same person. Generally, in those instances, caretaking responsibilities for the sampled child had changed. If MKACNT is greater

than 0, the values for variables such as relationship to the child and parent education may vary in different survey years.

**PARNENSP** is a CATI administrative variable indicating whether the parent in interview was administered in English or Spanish. The variable indicating whether the child interview was administered in English or Spanish is **CHLDENSP**. On the longitudinal file, CHLDENSP is missing for the whole-case-imputation cases.

**CHLDDOW**, a CATI administrative variable, specifies the day of the week on which the child interview was conducted. On the longitudinal file, CHLDDOW is missing for the whole-case-imputation cases.

Based on information from the RDD sample provider about the state and county of residence associated with the sampled telephone numbers, the name of the metropolitan area, if any, associated with each phone number was determined. Using information from the Census Bureau, the metropolitan areas of the nation were ranked by size into six groups where the total population of each group was roughly equal. These values are contained in the variable **HEXILE**. A value of 1 indicates that the respondent lives in one of the largest metropolitan areas such as New York or Chicago. A value of 6 indicates that the respondent lives in one of the smallest metropolitan areas such as Enid, Oklahoma, or Pocatello, Idaho. A value of 0 indicates that the respondent lives in a nonmetropolitan area.<sup>7</sup>

Geographic information based on addresses respondents provided during the first interview for the panel (or the updated address for the second interview) was obtained from the Claritas Corporation. Three variables that provide different levels of detail have been provided in these files. **URBAN\_5** is the least detailed and contains the Claritas-defined five levels of urbanization: metro urban, metro suburban, second city, town or exurban, and rural.

The most detailed of the three levels of geographic information is contained in **PRIZM\_CL**, which is drawn from Claritas's market segmentation system that classifies every neighborhood in the United States into 1 of 62 distinct lifestyle types or "clusters."<sup>8</sup>

Claritas groups the 62 clusters into 15 social groups, intended to maximize variation across urbanization and socioeconomic status. This variable is called **SOC\_GRP**. It is a two-character variable.

---

<sup>7</sup> To use HEXILE as an ordinal variable, it would be better to code the nonmetropolitan households as HEXILE=7.

<sup>8</sup> Claritas has a brief description of each cluster on its website, [www.clusterbigip1.claritas.com](http://www.clusterbigip1.claritas.com).

The first character reflects urbanization while the second reflects socioeconomic status. The number 1 indicates the highest socioeconomic level within categories of urbanization, and 3 indicates the lowest level.

## **Questionnaire Variables**

Not all items in the YMCLS questionnaires are represented by specific variables in the data files. In the following sections, discrepancies are noted and the variables corresponding to survey items on engagement in physical activities are explained.

Responses to questions in the parent interview designed to ascertain the sampled child's specific age (P1, P2, and P2a) were combined by the CATI system in a calculation of age contained in the derived variable **\_AGE**.

Two items in the parent interview (P\_D1 and P\_D2) were asked if the parent respondent spontaneously said that the sampled child had a disability. Children with disabilities that prevented them from any participation in physical and social activities were ineligible for the survey. Appropriate questions were administered for children who had a disability but were able to participate in either physical or social activities only. There were too few children in the latter category to provide a basis for meaningful analysis; therefore, these cases are not included in the YMCLS data files.

Child's race was asked of the parent in the first interview about the child. The format of the questions changed slightly after the Panel 1 2002 survey. Consequently, the variable names differ slightly between 2002 and other survey years. However, the classification of child's race is exactly identical in all survey years.

Questionnaire variables are listed in the codebooks in the order in which they were asked. (Please note that some questions were moved in some survey years, so variables are not in a strict numerical order.)

**Variables about Physical Activity.** Variables in the data files corresponding to survey item C3 indicate the specific activity children reported participating in during the week prior to the interview. A binary variable exists for each of the relatively common activities in which children participated; the variables indicate in what number and percentage of cases the activity was mentioned. For example, in 2002, 99 child respondents or 3.2% of the cases said they had done aerobics in the previous 7 days. Those cases were coded "yes" for C3AEROBC, and all other cases were coded "no," including cases in

which children said they had not done a physical activity and skipped question C3. The same convention applies to activities named at C7, activities done the day before the interview.

Responding children could mention up to eight activities they did in the week before the interview. For the first five activities, followup questions were asked to determine whether the activity was done as part of an organized group or in the child's discretionary time. Additional questions captured number of days during the past 7 that the child had done the activity; if appropriate, separate questions were asked about an activity done both as an organized activity and in the child's free time. The number of organized activity sessions was calculated as the number of organized activities times the number of days the child did the activity. Free-time sessions were calculated the same way. For instance, if a child reported that she had played basketball on two days in her free time, had practiced with her team one day, and had played a competitive game with her team one time, then two sessions of organized activity and two sessions of free-time activity would be calculated for her. The variables corresponding to survey item C4, AC4Z1-AC4Z5, indicate whether each of the first five activities named was an organized activity, free-time activity, or both. The variables corresponding to survey item C5, AC5Z1-AC5Z5, indicate the number of organized physical activity sessions in the 7 days prior to the interview; AC5A1-AC5A5 indicate the number of free time activity sessions during the same time. Derived variables, explained below, hold the calculations across all activities named by responding children.

As noted above, children named activities done in the 7 days prior to the interview and activities done yesterday. In 2002, only children who said they had done physical activity the 7 days prior to the interview were asked whether they had done any activity the day before the interview. In 2003, half of the children who said they had not done any activity in the previous week were asked about the previous day. Because some of them said they had done a physical activity yesterday, although they had not said they done an activity in the past week, all children were asked about yesterday after 2003.

**Variables Pertaining to VERB Understanding.** Two open-ended questions on the YMCLS captured children's understanding of VERB messages. The text files containing their verbatim responses were coded into up to 17 categories that captured the range of responses. At each question, respondents were probed to elicit full responses, and up to five ideas could have been coded for each response. The value labels show the possible categories that were coded. \_C18AN through \_C18EN hold the codes for the open-ended responses given by children to C\_18 (*"Tell me in your own words what VERB is all about."*). \_C19AN through \_C19EN hold the codes for the open-ended responses given by children to C\_19 (*"What ideas did VERB give you?"*).



## Derived Variables

This section describes the variables that were derived from two or more questionnaire variables or that were recoded from questionnaire variables. They were created for specific analytical purposes and are provided in the data files as a courtesy to data users. All derived variables begin with an underscore (“\_”). For ease of location, they are grouped together in the codebooks in alphabetical order following the questionnaire variables. They are grouped logically below.

### Administrative Derived Variables

**\_CHLDWKD2** is a recode of CHLDDOW, the day of the week the child interview was administered, into weekday and weekend categories.

**\_REGION** indicates the Census region in which the respondent child and his or her family resides. The categories are Northeast, Midwest, South, and West.

### Parent and Household Characteristic Derived Variables

**\_RELATE4** is a recode of RELATE into four categories representing the most commonly reported parent respondent relationships to child: mother, father, grandmother, and other.

**\_EDUCATE** is derived from P17, parent respondent’s level of education, and P18, attainment of a high school diploma or GED. The values are less than high school, high school, some college, technical school or associates degree, college degree, and advanced degree.

**\_INCOME** is a recode of P23 through P23SET3, level of household income. The 14 categories include \$5,000 increments from \$5,000 or less to \$50,000 (10 categories), \$50,001 to \$60,000, \$60,001 to \$75,000, \$75,001 to \$100,000, and over \$100,000. **\_INCOME2** is a recode of **\_INCOME**. The categories are \$25,000 or less, \$25,001 to \$50,000, \$50,001 to \$75,000, and over \$75,000.

## Child Characteristic Derived Variables

**\_AGE** is based on the calculation of the child's age by the CATI system. In the survey's after the first interview about the child (2002 for Panel 1 and 2004 for Panel 2), the value of **\_AGE** was increased by one. For instance, the value of **\_AGE** for a child who was age 9 in 2002 will be 10 in 2003.

**\_RACEETH** indicates child's race. It captures Hispanic ethnicity regardless of race or the number of race/ethnicities identified. The categories are white, black or African American, Hispanic or Latino, Asian, Native Hawaiian/Pacific Islander, American Indian/Alaska Native, and multiracial/other. It is based on the responses to the P19 series of questions, which differed slightly in wording for Panel 1 and Panel 2 but captured the same information (e.g., P19WHITE in 2002 and P19BWHIT in subsequent years, etc.).

**\_RACEETH2** is a recode of **\_RACEETH**. Child's race is collapsed into four categories: white, black or African American, Hispanic or Latino, and other race.

**\_C14HRS**, **\_C14MIN**, **\_CTVTIME**, and **\_CTVTIME2** are all derived from C14 and C14OV which captured the amount of time the child watched television or played video or computer games the day before the interview. The intermediary variables used to create **\_CTVTIME** and **\_CTVTIME2** are **\_C14HRS** and **\_C14MIN**, the number of hours and number of minutes, respectively. **\_CTVTIME** records the total time in minutes, and **\_CTVTIME2** holds the calculation for minutes converted to hours, collapsed into four categories: 0, 1 hour or less, 1 hour and 1 minute to 2 hours, and more than 2 hours.

## Physical Activity Derived Variables

**\_ACTORGR** is the number of all organized activities the child respondent reported participating in during the 7 days prior to the interview. This variable is calculated as the sum of AC4Z1, AC4Z2, AC4Z3, AC4Z4, and AC4Z5, the variables that classify each of the first five activities the child reported in C3 as organized activities, free-time activities, or both at C4.

**\_DAYSORGR**, the number of organized activity sessions, was calculated from AC5Z1, AC5Z2, AC5Z3, AC5Z4, and AC5Z5. These latter variables are a CATI calculation of the number of days each named activity was participated in as an organized activity (C5), and they were summed to create the continuous variable, **\_DAYSORGR**.

**\_DAYSORGR2** is a recode of **\_DAYSORGR** into two categories, no organized physical activity and organized physical activity.

**\_ACTFREER** is the number of free-time activities the child respondent reported participating in during the 7 days prior to the interview. This variable is calculated as the sum of AC4Z1, AC4Z2, AC4Z3, AC4Z4, and AC4Z5, the variables that classify each of the first five activities the child reported in C3 as organized activities, free-time activities, or both at C4.

**\_DAYSFREER**, the number of free-time activity sessions, was calculated from AC5A1, AC5A2, AC5A3, AC5A4, and AC5A5. These latter variables are a CATI calculation of the number of days each named activity was participated in as a free-time activity (C5), and they were summed together to create the continuous variable, **\_DAYSFREER**.

**\_DAYSFREER2** is a recode of **\_DAYSFREER** into two categories, zero to two free-time activity sessions and three or more free-time activity sessions.

**\_ACTTOTR** is a combination of **\_DAYSORGR** and **\_DAYSFREER**. It gives the total number of activities, whether organized or free time, that the child reported participating in during the 7 days prior to the interview.

**\_DAYSTOTR** is the calculation of the total number of organized and free-time activity sessions in which the child participated in the 7 days prior to the interview. Sessions were calculated as the number of activities times the number of days each activity was done.

**\_TOTALPA2** is a recode of **\_DAYSTOTR** into two categories, no organized and zero to two free-time activity sessions versus at least one organized and/or three or more free-time activity sessions.

### **VERB Awareness and Understanding Derived Variables**

**\_EXFREQ** categorizes child awareness and understanding of VERB combined with the frequency with which VERB advertisements were reported as having been seen. This variable is derived from C\_16 (unprompted or unaided naming of VERB as a “message for kids getting active”), C\_17 (naming VERB after prompting), C\_18UND and C\_19UND, the variables indicating understanding of VERB messages, and C\_21X, how often the child usually saw VERB messages on television or heard them on radio. The values for **\_EXFREQ** are no exposure or no understanding, exposed less than once

per week, exposed once a week, exposed several times a week, exposed every day. Children who had learned about and understood VERB from sources other than television or radio were coded as exposed less than once per week.

**\_EXINDEX5** combines child awareness and understanding of VERB in one variable. It is derived from C\_16 (unprompted or unaided naming of VERB as a “message for kids getting active”), C\_17 (naming VERB after prompting), \_C18UND and \_C19UND, the variables indicating understanding of VERB messages. **\_EXINDEX5** has five values: no recall, aided recall with no understanding of VERB messages, unaided recall with no understanding of VERB messages, aided recall with understanding of VERB messages, and unaided recall with understanding of VERB messages.

**\_EXINDEX4** is recoded from **\_EXINDEX5**. It has four values: no recall, exposure with no understanding of VERB messages, aided recall with understanding of VERB messages, and unaided recall with understanding of VERB messages.

**\_P\_UNAIDED** is derived from P\_14D, parents’ unprompted identification of VERB as a “message for kids getting active”, and P\_15E, parents’ recognizing VERB after prompting. The values for this variable are unaided recall versus aided or no recall.

**\_P\_AWARE** is derived from **\_P\_UNAIDED** and **\_P\_AIDED**. It indicates overall awareness of VERB on the part of parents with values of unaided or aided recall of VERB versus no recall.

**\_C18UND** is a recode of \_C18AN through \_C18EN. If any of the child’s responses to C\_18 coded in \_C18AN through \_C18EN contained accurate VERB messages, **\_C18UND** is coded as understanding of VERB. Accurate VERB messages were defined as response codes of 3, 4, 5, 6, 7, 9, 11, 12, or 16.

**\_C19UND** is a recode of \_C19AN through \_C19EN. If any of the child’s responses to C\_19 coded in \_C19AN through \_C19EN contained ideas related to VERB messages, **\_C19UND** is coded as understanding of VERB. Accurate VERB messages were defined as response codes of 3, 4, 5, 6, 7, 9, 11, 12, or 16.

## 5. USING THE YMCLS FILE SYSTEM FOR CROSS-SECTIONAL ANALYSES

This chapter provides information on how to conduct cross-sectional analyses with the YMCLS file system for 2002 to 2004, including instructions for computing point estimates and their variances.

### **Analyses Supported by the File System**

Westat has constructed a cross-sectional file in a format that is ideally suited to producing cross-sectional profiles of America's youth, in terms of means and percentages. Specifically, data users can produce cross-sectional estimates for youth aged 9 to 13 in the spring of 2002, youth aged 10 to 14 in the spring of 2003, and youth aged 9 to 15 in the spring of 2004. Users should note, however, that an analysis of temporal trends from 2002 to 2004 needs to be restricted to ages 10 to 13 in each year to remove any confounding developmental effects. It is important to understand the layout of the cross-sectional file, an explanation of which is given below. The file system also contains a longitudinal file designed to support longitudinal analyses of two years of the maturation process (see Chapter 6).

The file system cannot be used to replicate Westat's dose-response analysis of the association between VERB exposure and outcomes because the additional weights required for that purpose are not in these files.

### **Analyses of Parent Data**

The YMCLS file system should not be used to profile parents of American youth. The parent interviews were conducted with knowledgeable custodial parents rather than with a random sample of all parents. The result is that fathers are underrepresented. Thus, these data are not representative of the parents of 9- to 15-year-olds in the U.S. In addition, some of the parent data is duplicated in the file system when two children per household were interviewed.

However, if analysis of the data at the parent level is desired, a non-representative, non-random sample can be extracted from the file system. An unweighted data set may be created by unduplicating parent responses in households in which one parent responded for two children. The second case in households with two children can be deleted because the parent data is duplicated for such

households. Note that the first eight digits of **CHLDID** identify the household; the ninth and tenth digits identify the child interviewed within the household.

### **Cross-sectional File Layout**

The cross-sectional file contains a *separate record for each completed child interview* between 2002 and 2004, except cases where the child was disabled, as noted in Chapter 3. Data from the parent interview and the child interview are on one record for each case. The file includes data for both Panel 1 and Panel 2 in 2004. Records for the three different years are “stacked” on top of each other so that, for example, there will be separate records for 2002, 2003, and 2004 for a child who completed interviews in all three years. These records will have the same value of CHLDID but are easily distinguished by the value of the variable YEAR. With the file constructed in this way, there is no need to reference the year of interview in any of the variable names. The variable *\_AGE*, for example, represents: the age of the child in 2002 for records where YEAR = 2002; the age of the child in 2003 for records where YEAR = 2003; and the age of the child in 2004 for records where YEAR = 2004. The same logic applies to the weight variables so that the *name* of the cross-sectional weight for Panel 1 is the same for each year, but its *value* on each record is appropriate for the year determined by the value of the variable YEAR.

Three sets of cross-sectional weights are included in the cross-sectional file: one set of weights for Panel 1; one set of weights for Panel 2; and one set of weights for the two panels combined. The weights for use with Panel 1 are missing for records where PANEL = 2. The weights for use with Panel 2 are missing for records where PANEL = 1. The weights for use with both panels combined are non-missing for all records in the cross-sectional file; this is for user-convenience and is explained below. The names and appropriate usage of these three sets of weights are given in the next section.

### **Computing Point Estimates**

To produce unbiased point estimates, it is essential that the correct sampling weights are used. The appropriate choice of weights for different forms of cross-sectional analysis is explained below. (Refer to Chapter 2 for the methods that were used to compute these weights.)

To prepare **cross-sectional estimates for 2002 only**, the data user should first subset the cross-sectional file to records where YEAR = 2002. The appropriate full sample weight for such analyses

is called P1\_CS\_WGT0 (all 2002 cases belong to Panel 1). Cross-sectional estimates for 2002 are restricted to children aged 9–13.

To prepare **cross-sectional estimates for 2003 only**, the data user should first subset the cross-sectional file to records where YEAR = 2003. The appropriate full sample weight for such analyses is also P1\_CS\_WGT0 (all 2003 cases belong to Panel 1). Cross-sectional estimates for 2003 are restricted to children ages 10 to 14.

There are three different ways to prepare **cross-sectional estimates for 2004 only**, depending upon the survey panel and/ or population of interest. However, in each case the data user should first subset the cross-sectional file to records where YEAR = 2004. If the user wishes to use *only* data from Panel 1, then the cross-sectional file should be further subset to records where PANEL = 1. The appropriate full sample weight for such analyses is P1\_CS\_WGT0 (all cases of interest belong to Panel 1). Cross-sectional estimates for 2004 using Panel 1 only are restricted to children ages 11 to 15. If the user wishes to use *only* data from Panel 2, then the cross-sectional file should be subset to records where PANEL = 2 (after the initial subset by year). The appropriate full sample weight for such analyses is called P2\_CS\_WGT0 (all cases of interest belong to Panel 2). Cross-sectional estimates for 2004 using Panel 2 only are restricted to children aged 9–13. Depending on the population of interest, the user can take advantage of larger sample sizes by combining Panels 1 and 2 to produce cross-sectional estimates for 2004. In this case the cross-sectional file need only be subset by year. The appropriate full sample weight for this combined panel analysis is called P1P2\_CS\_WGT0 (cases from both panels are used). Cross-sectional estimates for 2004 can be produced for children ages 9 to 15 when Panels 1 and 2 are combined. However, the benefit of increased sample size applies only to ages 11 to 13.

For all **analyses involving population shifts between 2002 and 2004**, the cross-sectional file should be subset (using the variable \_AGE) so that it only includes records for children aged 10 to 13 at the time of interview. This ensures that comparable populations are being compared. The appropriate full sample weight for such trend analyses is P1P2\_CS\_WGT0. For records associated with the years 2002 and 2003 this combined-panel cross-sectional weight has been set equal to the Panel 1 cross-sectional weight, while for 2004 it is truly the combined cross-sectional weight that applies to analyses of Panels 1 and 2 together. This ensures that the correct weight is being applied to each case in each year when analyzing change over time. Variance estimates can then be generated using custom programming or the WesVar or SUDAAN software packages, as described below.

If the data user does not want to use values of the analysis variable that have been imputed, these cases can be dropped by using the appropriate imputation flag. Imputation flags names are

composed of the prefix “IMP” followed by the name of the variable. For instance, the imputation flag for the variable P3 is IMP3. Imputation flags are listed in the codebooks after the questionnaire and derived variables and before the weight variables.

## Estimating Variances

Standard errors for the YMCLS were computed through a design-based, jackknife replication variance methodology (JK2) using software package **WesVar** (Westat, 2000)<sup>9</sup>. In this method, subsamples, or replicates, are formed from the original sample. These subsamples were created by first forming 60 pseudo strata, reflecting the order in which the full sample was systematically drawn from the vendor frame, each of which was divided into two half-samples. Sixty replicate weights were formed by setting the weights to zero for one randomly selected half of the sample from one pseudo stratum at a time and doubling the weights of the households and persons in the other half of the sample from the same pseudo stratum. The estimates of interest were then calculated from each of the replicates as well as from the full sample. The variance among the replicate estimates was used to estimate the variance for the full sample.

The standard error is a measure of the variance in estimates produced by the YMCLS sample. Standard errors are used to create confidence intervals. Confidence intervals show the margin of error in point estimates that results from surveying a sample rather than the entire population. If similar independent samples were drawn repeatedly, and the statistic of interest and its 95 percent confidence interval based on the standard error were calculated from each of the samples, 95 percent of those confidence intervals would contain the value of the statistic for the population. For instance, in 2003, 90 percent of 10- to 13-year-olds in the nation reported that they engaged in physical activity in the previous week. This estimate has a standard error of 0.6. Multiplying this standard error by 1.96 produces a 95 percent confidence interval for the percentage of children who engaged in physical activity in the previous week of approximately 89 to 91 percent.

The **replicate weights** for 2002 cross-sectional estimates are P1\_CS\_WGT1 through P1\_CS\_WGT60. The replicate weights for 2003 cross-sectional estimates are also P1\_CS\_WGT1 through P1\_CS\_WGT60. There are three sets of replicate weights for 2004 cross-sectional estimates. For 2004 cross-sectional estimates based on Panel 1 only, the replicate weights are P1\_CS\_WGT1 through

---

<sup>9</sup> WesVar is a proprietary analysis package developed by Westat. Information about purchasing WesVar can be found at <http://www.westat.com/wesvar/licensing/index.html>. Go to <http://www.westat.com/wesvar/about/WV4.2%20Manual.pdf> to download the manual at no cost.



P1\_CS\_WGT60. For 2004 cross-sectional estimates based on Panel 2 only, the replicate weights are P2\_CS\_WGT1 through P2\_CS\_WGT60. For 2004 cross-sectional estimates based on Panels 1 and 2 combined, the replicate weights are P1P2\_CS\_WGT1 through P1P2\_CS\_WGT60. For trend analyses (using data from both panels in 2004), the replicate weights are also P1P2\_CS\_WGT1 through P1P2\_CS\_WGT60.

The replicate weights in the YMCLS files can be used with WesVar or SUDAAN. They can also be used with user-written custom software. Each of the three approaches is described in more detail below. Please note that to the best of our knowledge there are no other statistical software systems that will calculate appropriate variance estimates for this file system. Two children were interviewed in some households, and standard multi-level modeling programs are known to have problems with clusters of size two (Korn and Graubard, 2003).

Some considerations in choosing among the three analytic software options are now discussed, following which, detailed instructions for each option are given.

**Custom code** is the most flexible, but also the most labor-intensive. Most statistics of interest can be produced with appropriate standard errors in both WesVar and SUDAAN.

To estimate variances using **WesVar**, it is necessary to create a WesVar data file. This is a copy of the data of interest stored in a special format. WesVar is an interactive program with a graphical user interface (GUI) that runs in a Windows environment. It can read SAS files of the type delivered with this manual, as well as other types of files. Creating a Wesvar file can be somewhat complicated, but once the file is created, it is easy to specify tables, regressions, and logistic regressions. We recommend that all new derived and composite variables desired by analysts be created in SAS. Although simple recodes can be performed in WesVar, there is no audit trail of the process.<sup>10</sup>

To estimate variances using **SUDAAN**, one must have access to Release 8.0, or higher.<sup>11</sup> Earlier releases will not work with replicate weights. Version 8.0.2 (and higher) will run within SAS, eliminating the need to translate files. SUDAAN is run by programming statements that are similar to SAS statements rather than by using a GUI interface as in WesVar. A disadvantage of SUDAAN is that all table variables must be numeric variables (some in the file system are currently alphanumeric).

---

<sup>10</sup> Also, the process of exporting WesVar files back into SAS would not work well on this data set because of the long variable names.

<sup>11</sup> SUDAAN is a proprietary analysis package developed by the Research Triangle Institute. Information about SUDAAN can be found at <http://www.rti.org/sudaan/>.

variables) and must be coded from 1 to  $m$  with no gaps, where  $m$  is the number of levels for the variable. Also, the number of levels for each variable must be known before a table involving the variable can be created in SUDAAN. These limitations do not apply in WesVar. Neither package produces very attractive tables in their regular printed output, but the SUDAAN output is somewhat more attractive than the WesVar output. Both systems allow the output of statistics to files where they can be manipulated to produce attractive tables. In addition, WesVar has a companion software package called TableViewer that can easily display attractive two-way tables.

For simple statistics involving variables with no missing values, both packages should produce the same results when using the same full sample weight and set of replicate weights. However, there are differences between the two packages in terms of the handling of missing values, the computation of complex test statistics and associated p-values, and the calculation of design effects. Regarding complex test statistics, both procedures are approximations. No broad rules have been established for when one approximation is superior to the other, but both approximations are far better than ignoring the complex design in analysis.

### **Detailed Instructions for Custom Programming of Variance Calculations**

For custom analyses not supported by WesVar or SUDAAN, variance estimates can be calculated as

$$\hat{\sigma}^2 = \sum_{r=1}^{60} (\hat{t}_r - \hat{t}_0)^2,$$

where  $\hat{t}_r$  is the estimate of the statistic of interest calculated using the  $r$ -th replicate weight and  $\hat{t}_0$  is the estimate calculated using the full sample weight.

### **Detailed Instructions for Using WesVar for Variance Calculations**

Since WesVar is a GUI interface, a series of screenshots and sample output is provided in addition to commentary. The process is illustrated in the screenshots (and in the SUDAAN example that follows) is for a cross-sectional analysis of the attitude of 10- to 13- year-olds about the availability of neighborhood places for physical activities. Data from 2002 to 2004 are used.

Figure 5-1 shows a screen shot for the WesVar interface that is used to create the WesVar file. It is for the YMCLS cross-sectional file.<sup>12</sup> Note that the “JK2” option is checked for “Method.” This is the correct choice for the YMCLS survey and is critically important.

Figure 5-2 shows a screen shot for the WesVar interface that is used to request a table. The “RS2” and “RS3” buttons request the two versions of design-based chi-square independence tests suggested by Rao and Scott (1981, 1984), as explained in the WesVar manual. Although not shown, the table request has been subset to children ages 10 to 13 in each year by entering “\_AGE >= 10 AND \_AGE <= 13” in the “Sub Set” field of the WesVar window that first appears when the user requests a table.

Figure 5-3 shows a screen shot for the WesVar interface for viewing an output table within WesVar. This output can be exported to a text file, which is shown in Attachment 1.

Figure 5-4 is a screen shot of the same table displayed in WesVar TableViewer, a separate program that is provided with WesVar.

---

<sup>12</sup> The file name in the figure differs slightly because the program was run on an earlier version.

Figure 5-1. Screen shot for creating a WesVar file

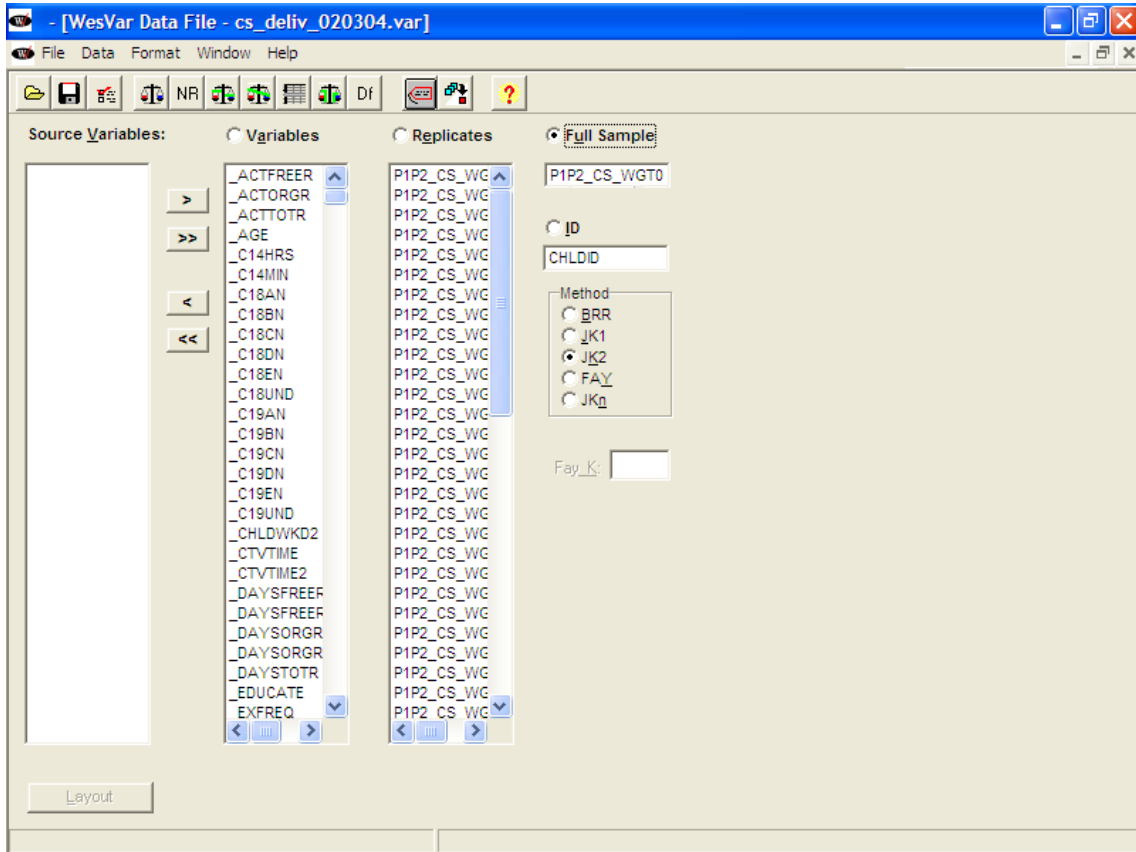
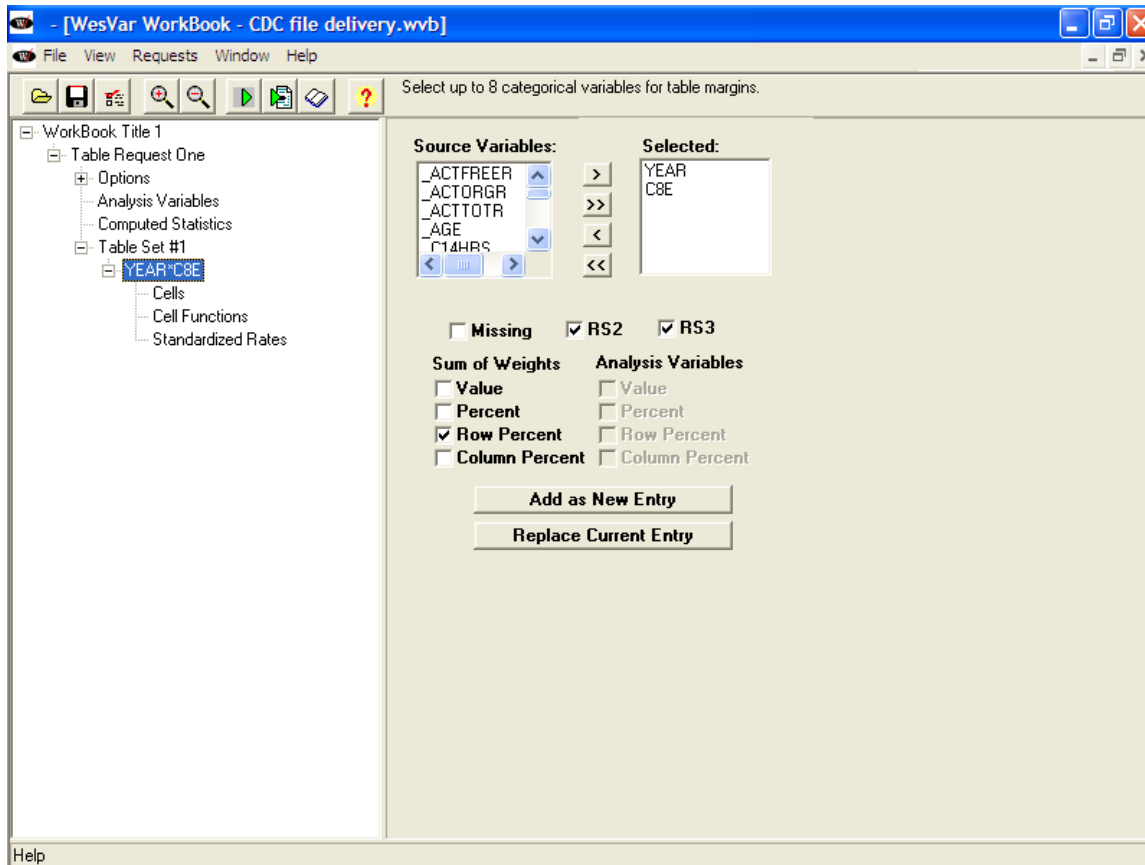


Figure 5-2. Screen shot for creating a table in WesVar



**Figure 5-3. Screen shot for viewing a table in WesVar**

WesVar Output File for Workbook Title 1

File View Help

WorkBook Title 1

- Table Request One
  - Analysis Variables
  - Computed Statistics
  - Table Set #1
    - YEAR \* C8E**
    - Chi-Square

TABLE : YEAR \* C8E

YEAR	C8E	STATISTIC	EST_TYPE	ESTIMATE	STDERROR	CV (%)
2002	1	SUM_WTS	ROWPCT	44.79	1.007	2.
2002	2	SUM_WTS	ROWPCT	28.82	0.909	3.
2002	3	SUM_WTS	ROWPCT	15.39	0.721	4.
2002	4	SUM_WTS	ROWPCT	11.00	0.667	6.
2002	MARGINAL	SUM_WTS	ROWPCT	100.00	.	
2003	1	SUM_WTS	ROWPCT	47.74	1.369	2.
2003	2	SUM_WTS	ROWPCT	29.75	1.050	3.
2003	3	SUM_WTS	ROWPCT	14.19	0.788	5.
2003	4	SUM_WTS	ROWPCT	8.33	0.768	9.
2003	MARGINAL	SUM_WTS	ROWPCT	100.00	.	
2004	1	SUM_WTS	ROWPCT	48.71	0.865	1.
2004	2	SUM_WTS	ROWPCT	28.11	0.727	2.
2004	3	SUM_WTS	ROWPCT	13.82	0.548	3.
2004	4	SUM_WTS	ROWPCT	9.36	0.441	4.
2004	MARGINAL	SUM_WTS	ROWPCT	100.00	.	
MARGINAL	1	SUM_WTS	ROWPCT	47.09	0.716	1.
MARGINAL	2	SUM_WTS	ROWPCT	28.89	0.588	2.
MARGINAL	3	SUM_WTS	ROWPCT	14.46	0.461	3.
MARGINAL	4	SUM_WTS	ROWPCT	9.56	0.384	4.
MARGINAL	MARGINAL	SUM_WTS	ROWPCT	100.00	.	

**Figure 5-4. Screen shot for viewing a table in TableViewer**

WesVar Table Viewer - [C:\CDC file delivery..001]

File Edit Window Help

Table : YEAR \* C8E  
Variable : SUM\_WTS

		C8E				
YEAR	r% Estimate	1	2	3	4	MARGINAL
		r% Std. Error				
2002	44.79 1.007	28.82 0.909	15.39 0.721	11.00 0.667	100.00 .	
2003	47.74 1.369	29.75 1.050	14.19 0.788	8.33 0.768	100.00 .	
2004	48.71 0.865	28.11 0.727	13.82 0.548	9.36 0.441	100.00 .	
MARGINAL	47.09 0.716	28.89 0.588	14.46 0.461	9.56 0.384	100.00 .	

## Detailed Instructions for Using SUDAAN for Variance Calculations

Below is a sample SAS program that uses the SUDAAN CROSSTAB procedure to compare the 2002, 2003, and 2004 distributions of the attitude of 10- to 13-year-olds about the availability of neighborhood places for physical activities. The program starts with the YMCLS cross-sectional file as input. Please note the recoding of the YEAR variable to be numeric with values of 1 and 2; a requirement of SUDAAN.

Attachment 5-2 shows the output produced by SUDAAN. As noted above, computations of statistics can differ between WesVar and SUDAAN; the standard errors on row percentages are the same as in the WesVar output in Attachment 5-1 but the chi-square tests for independence are different in the WesVar and SUDAAN examples.

```
libname ymc '\\YMC\CSfile\SASData';

OPTIONS FMTSEARCH=(ymc)
        NOCENTER NOFMterr LINESIZE=105 PAGESIZE=80 VALIDVARNAME=UPCASE;

DATA temp;
    SET ymcls_cs_020304;
    IF year = 2002 THEN YearNumeric = 1;
    ELSE IF year = 2003 THEN YearNumeric = 2;
    ELSE IF year = 2004 THEN YearNumeric = 3;
    IF _age NOT IN (10,11,12,13) THEN DELETE;
RUN;

PROC CROSSTAB DATA=temp FILETYPE=SAS DESIGN=JACKKNIFE;
    WEIGHT plp2_cs_wgt0;
    JACKWGTS plp2_cs_wgt1-plp2_cs_wgt60 / ADJJACK=1;
    TABLES YearNumeric*c8e;
    SUBGROUP YearNumeric c8e;
    LEVELS 3 4;
    TEST CHISQ LLCHISQ;
    PRINT NSUM WSUM SEWGT ROWPER SEROW;
RUN;
```



## Attachment 5-1. Output from WesVar Example

### Summary Information of Table Request One

WESVAR VERSION NUMBER : 4.2  
TIME THE JOB EXECUTED : 18:32:46 03/15/2006  
INPUT DATASET NAME : \\YMC\CSFile\WesVarData\ymcls\_cs\_020304.var  
TIME THE INPUT DATASET CREATED : 16:42:26 03/15/2006  
FULL SAMPLE WEIGHT : P1P2\_CS\_WGT0  
REPLICATE WEIGHTS : P1P2\_CS\_WGT1...P1P2\_CS\_WGT60  
VARIANCE ESTIMATION METHOD : JK2

OPTION COMPLETE : ON  
OPTION FUNCTION LOG : ON  
OPTION VARIABLE LABEL : OFF  
OPTION VALUE LABEL : OFF  
OPTION OUTPUT REPLICATE ESTIMATES : OFF  
FINITE POPULATION CORRECTION FACTOR : 1.00000  
VALUE OF ALPHA (CONFIDENCE LEVEL %) : 0.05000 (95.00000 %)  
DEGREES OF FREEDOM : 60  
t VALUE : 2.000  
SUBSET CRITERIA : \_AGE >= 10 AND \_AGE <= 13

ANALYSIS VARIABLES : None Specified.  
COMPUTED STATISTIC : None Specified.  
TABLE(S) : YEAR\*C8E

FACTOR(S) : 1.00

NUMBER OF REPLICATES : 60  
NUMBER OF OBSERVATIONS READ : 10297  
WEIGHTED NUMBER OF OBSERVATIONS READ : 50930082.136

TABLE : YEAR \* C8E

YEAR	C8E	STATISTIC	EST_TYPE	ESTIMATE	STDERROR	CV(%)	CELL_n	DENOM_n	DEFF	
2002	1	SUM_WTS	ROWPCT	44.79	1.007	2.247	1140	2515	2515	1.030
2002	2	SUM_WTS	ROWPCT	28.82	0.909	3.155	730	2515	2515	1.014
2002	3	SUM_WTS	ROWPCT	15.39	0.721	4.687	379	2515	2515	1.005
2002	4	SUM_WTS	ROWPCT	11.00	0.667	6.063	266	2515	2515	1.143
2002	MARGINAL	SUM_WTS	ROWPCT	100.00	.	.	2515	2515	.	
2003	1	SUM_WTS	ROWPCT	47.74	1.369	2.867	1068	2210	2210	1.659
2003	2	SUM_WTS	ROWPCT	29.75	1.050	3.530	657	2210	2210	1.167
2003	3	SUM_WTS	ROWPCT	14.19	0.788	5.558	304	2210	2210	1.129
2003	4	SUM_WTS	ROWPCT	8.33	0.768	9.225	181	2210	2210	1.708
2003	MARGINAL	SUM_WTS	ROWPCT	100.00	.	.	2210	2210	.	
2004	1	SUM_WTS	ROWPCT	48.71	0.865	1.777	2679	5572	5572	1.671
2004	2	SUM_WTS	ROWPCT	28.11	0.727	2.585	1607	5572	5572	1.456
2004	3	SUM_WTS	ROWPCT	13.82	0.548	3.967	753	5572	5572	1.406
2004	4	SUM_WTS	ROWPCT	9.36	0.441	4.709	533	5572	5572	1.276
2004	MARGINAL	SUM_WTS	ROWPCT	100.00	.	.	5572	5572	.	
MARGINAL	1	SUM_WTS	ROWPCT	47.09	0.716	1.520	4887	10297	10297	2.117
MARGINAL	2	SUM_WTS	ROWPCT	28.89	0.588	2.034	2994	10297	10297	1.730
MARGINAL	3	SUM_WTS	ROWPCT	14.46	0.461	3.186	1436	10297	10297	1.766
MARGINAL	4	SUM_WTS	ROWPCT	9.56	0.384	4.021	980	10297	10297	1.760
MARGINAL	MARGINAL	SUM_WTS	ROWPCT	100.00	.	.	10297	10297	.	

Chi-Square

CHI-SQUARE	D.F.	VALUE	PROB	
PEARSON		6.00	23.900	0.001
RS2	6.00	18.030	0.006	
RS3	5.45	16.373	0.008	

## Attachment 5-2. Output from SUDAAN Example

The SAS System

16:04 Wednesday, March 15, 2006 4

S U D A A N  
Software for the Statistical Analysis of Correlated Data  
Copyright Research Triangle Institute February 2005  
Release 9.0.1

Number of observations read : 10297      Weighted count : 50930082  
Denominator degrees of freedom : 60

Date: 03-15-2006  
Time: 16:25:20

Research Triangle Institute  
The CROSSTAB Procedure

Page : 1  
Table : 1

Variance Estimation Method: Replicate Weight Jackknife  
by: YEARNUMERIC, LOTS OF PLACES IN NEIGHBORHOOD FOR PA.

YEARNUMERIC		LOTS OF PLACES IN NEIGHBORHOOD FOR PA				
		Total	REALLY AGREE	SORT-OF AGREE	SORT-OF DISAGREE	REALLY DISAGREE
Total	Sample Size	10297	4887	2994	1436	980
	Weighted Size	*****	*****	*****	7364845.96	4867734.11
	SE Weighted	8057.27	364689.98	299207.10	234605.20	195721.47
	Row Percent	100.00	47.09	28.89	14.46	9.56
	SE Row Percent	0.00	0.72	0.59	0.46	0.38
1	Sample Size	2515	1140	730	379	266
	Weighted Size	*****	7544224.16	4853253.41	2591363.53	1853114.89
	SE Weighted	542.22	169530.95	153142.45	121445.76	112356.68
	Row Percent	100.00	44.79	28.82	15.39	11.00
	SE Row Percent	0.00	1.01	0.91	0.72	0.67
2	Sample Size	2210	1068	657	304	181
	Weighted Size	*****	8135311.98	5070221.97	2417651.60	1418792.38
	SE Weighted	13.82	233245.43	178999.15	134368.16	130877.20
	Row Percent	100.00	47.74	29.75	14.19	8.33
	SE Row Percent	0.00	1.37	1.05	0.79	0.77
3	Sample Size	5572	2679	1607	753	533
	Weighted Size	*****	8303612.61	4790877.94	2355830.83	1595826.84
	SE Weighted	7906.54	147513.52	123851.50	93482.64	75179.48
	Row Percent	100.00	48.71	28.11	13.82	9.36
	SE Row Percent	0.00	0.87	0.73	0.55	0.44

# Attachment 5-2. Output from SUDAAN Example

Date: 03-15-2006  
 Time: 16:25:20

Research Triangle Institute  
 The CROSSTAB Procedure

Page : 2  
 Table : 1

Variance Estimation Method: Replicate Weight Jackknife  
 Chi Square Test of Independence for YEARNUMERIC and LOTS OF PLACES IN NEIGHBORHOOD FOR PA

	ChiSq	16.91
	P-value ChiSq	0.0175
	Degrees of Freedom ChiSq	6
	LLChiSq	16.62
	P-value LLChiSq	0.0192
	Degrees of Freedom LLChiSq	6



## 6. USING THE YMCLS FILE SYSTEM FOR LONGITUDINAL ANALYSES

This chapter provides information on how to conduct longitudinal analyses with the YMCLS file system for 2002 to 2004, including instructions for computing point estimates and their variances. With respect to the types of analyses supported by the longitudinal file, many of the same caveats discussed in Chapter 5 apply: parent-level analyses are not recommended and the file cannot be used to replicate Westat's dose-response analysis of the association between VERB exposure and outcomes because the additional weights required for that purpose are not provided.

### Analyses Supported by the File System

Westat has constructed a longitudinal file in a format that is ideally suited to producing estimates of gross change (i.e., change at the individual level) for America's youth between 2002 and 2004, as well as estimates of associations between variables at different times during this period. Specifically, data users can analyze two years of the maturation process for youth aged 9 to 13 in the spring of 2002. It is important to understand the layout of the longitudinal file, an explanation of which is given below.

### Longitudinal File Layout

The longitudinal file contains *one record for each Panel 1 youth who completed an interview in 2004*, and was not disabled in any year. Each record contains data for 2002, 2003, and 2004. For those youth who completed interviews in 2003 and 2004 but not at baseline, the whole-case imputation described in Chapter 2 was used to complete the 2002 data. Nearly all variable names in the longitudinal file have a suffix indicating the year of data collection. The variable C11\_02, for example, represents whether or not the child belonged to a school group (like band, drama club, newspaper, or student government) in 2002; the variable C11\_03 represents whether or not the child belonged to a school group in 2003; and the variable C11\_04 represents whether or not the child belonged to a school group in 2004. The exceptions to this rule include variables whose values are constant over time (CHLDID, SEX, \_RACEETH, \_RACEETH2) and the variable \_AGE which represents the child's age in 2002. The child's age in 2003 (or 2004) can be determined by adding 1 (or 2) to his/her age in 2002. One set of longitudinal weights is included in the longitudinal file.

## Computing Point and Variance Estimates

To produce unbiased point estimates, appropriate sampling weights should be used. For **longitudinal** analyses of the youth in the longitudinal file, the full sample weight to use is P1\_LG\_WGT0 (the longitudinal weight for 2004). The population of inference for a weighted analysis of all cases in the longitudinal file is somewhat difficult to describe. Although the outcomes and other data in the file are relevant to the years 2002 through 2004, the 2004 longitudinal weights were controlled to Census 2000 totals. This means that estimates of totals reflect the size of the 9- to 13-year-old population in 2000, and estimates of means or percentages reflect the composition of the 9- to 13-year-old population in 2000 (in terms of single year of age, sex, race/ethnicity, etc). (Chapter 2 describes the methods that were used to compute the Panel 1 longitudinal weights.)

Variance estimates can be generated using custom programming or the WesVar or SUDAAN software packages, as described in Chapter 5. The replicate weights for longitudinal estimates are P1\_LG\_WGT1 through P1\_LG\_WGT60.

If the data user does not want to use values of the analysis variable that have been imputed, these cases can be dropped by using the appropriate imputation flag. Imputation flags names are composed of the prefix “IMP” followed by the name of the variable. For instance, the imputation flag for the variable P3\_02 is IMPP3\_02. By year, imputation flags are listed in the codebook for the longitudinal file after the questionnaire and derived variables. The weight variables appear at the end of the codebook.

## Example of Longitudinal Analysis Using WesVar

Since WesVar is a GUI interface, a series of screenshots and sample output is provided in addition to commentary. The process is illustrated in the screenshots (and in the SUDAAN example that follows) is for a longitudinal analysis of the overall physical activity level of 11- to 15-year-olds in 2004 in terms of certain attitudes toward physical activity reported by their parents and by themselves in 2002, and their age.

Figure 6-1 shows a screenshot for the WesVar interface that is used to create the WesVar file. It is for the YMCLS longitudinal file. Note that the “JK2” option is checked for “Method.” This is the correct choice for the YMCLS survey and is critically important.



Figure 6-2 shows a screen shot for the WesVar interface that is used to request a regression. Although not shown, a logistic regression has been requested by selecting “Logistic” in the “Model Type” field of the WesVar window that appears when the user highlights “Options” in the left-hand pane. (This is the appropriate choice because the dependent variable is binary.)

Figure 6-3 shows a screen shot for the WesVar interface for viewing a regression output within WesVar. This output can be exported to a text file, which is shown in Attachment 6-1.

Figure 6-1. Screen shot for creating a WesVar file

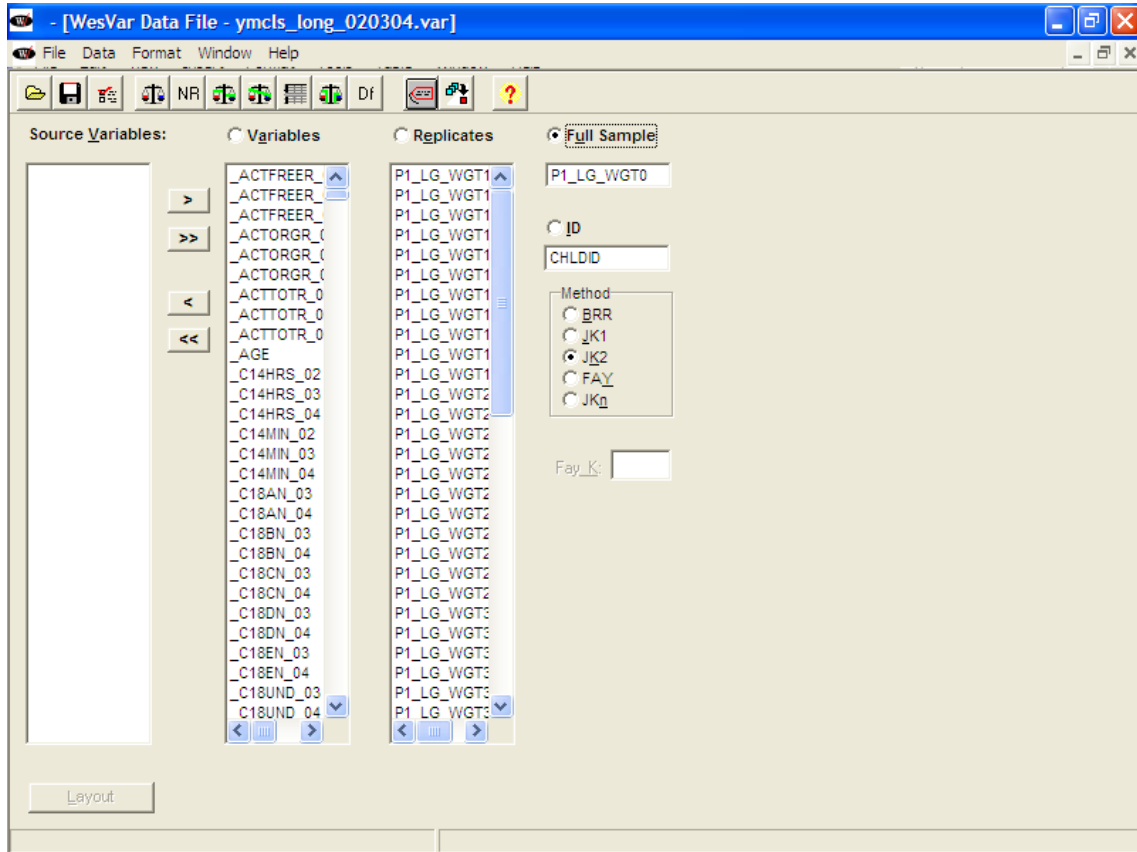
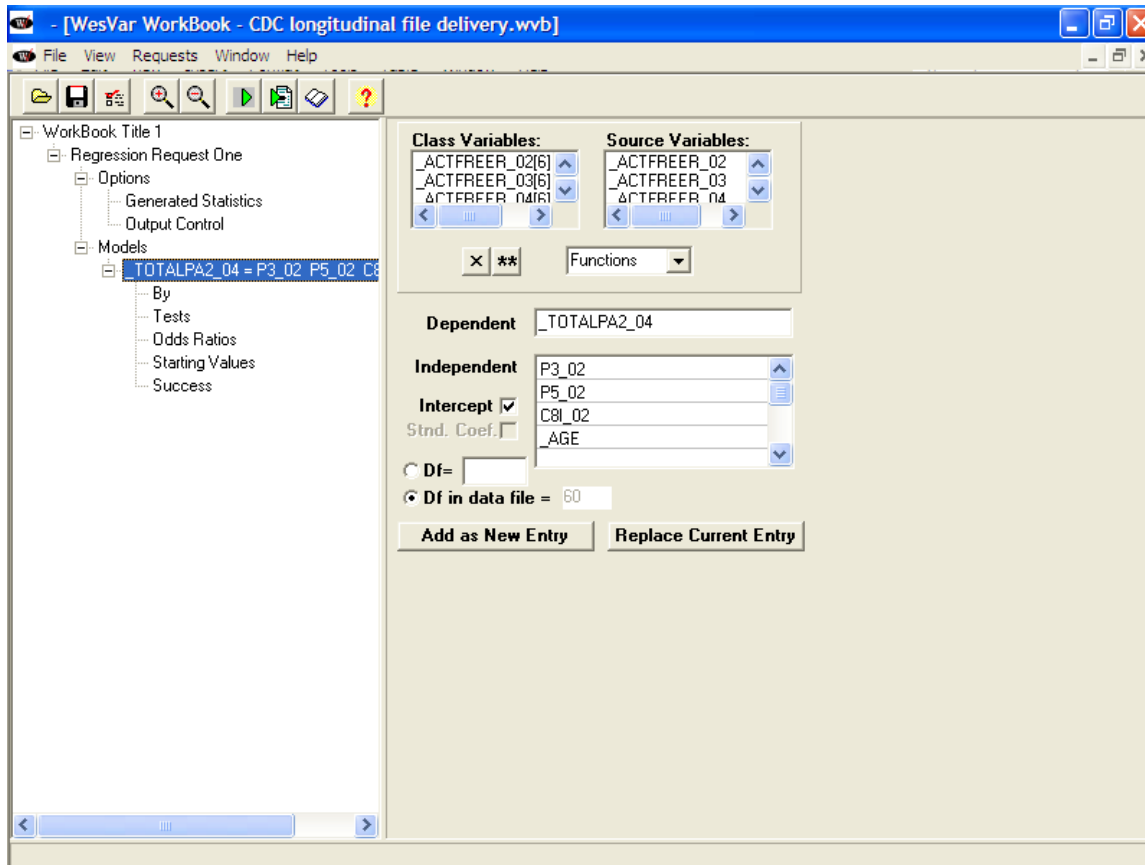


Figure 6-2. Screen shot for creating a regression in WesVar



**Figure 6-3. Screen shot for viewing a regression in WesVar**

WesVar Output File for Workbook Title 1

File View Help

WorkBook Title 1

- Regression Request One
  - Models
    - \_TOTALPA2\_04 = P3\_02 F
      - Measures of Fit
      - Estimated Coefficients**
      - Correlation Matrix
      - Tests
      - Odds Ratios

## Example of Longitudinal Analysis Using SUDAAN

Below is a sample SAS program that uses the SUDAAN MULTLOG procedure to relate the overall physical activity level of 11- to 15- year-olds in 2004 to certain attitudes toward physical activity reported by their parents and by them selves in 2002, and their age. The program starts with the YMCLS longitudinal file as input.

Attachment 6-2 shows the output produced by SUDAAN. In this instance, the parameter estimates, standard errors, and F-test statistics are the same in the WesVar and SUDAAN.

```
libname ymc '\\YMC\LGFile\SASData';

OPTIONS FMTSEARCH=(ymc)
        NOCENTER NOFMterr LINESIZE=105 PAGESIZE=80 VALIDVARNAME=UPCASE;

PROC REGRESS DATA=ymc.ymcls_long_020304 FILETYPE=SAS DESIGN=JACKKNIFE;
    WEIGHT p1_lg_wgt0;
    JACKWGTS p1_lg_wgt1-p1_lg_wgt60 / ADJJACK=1;
    CLASS _totalpa2_04 / DIR=DESCENDING;
    MODEL _totalpa2_04 = p3_02 p5_02 c8i_02 _age / CUMLOGIT;
    PRINT BETA SEBETA T_BETA P_BETA ADJWALDF DF ADJWALDP;
RUN;
```

## Attachment 6-1. Output from WesVar Example

56

### Summary Information of Regression Request One

WESVAR VERSION NUMBER : 4.2  
TIME THE JOB EXECUTED : 10:54:56 03/23/2006  
INPUT DATASET NAME : \\YMC\LGFile\WesVarData\ymcls\_long\_020304.var  
TIME THE INPUT DATASET CREATED : 10:50:19 03/23/2006  
FULL SAMPLE WEIGHT : P1\_LG\_WGT0  
REPLICATE WEIGHTS : P1\_LG\_WGT1...P1\_LG\_WGT60  
VARIANCE ESTIMATION METHOD : JK2

TYPE OF ANALYSIS : LOGISTIC  
CONVERGENCE CRITERION : 1e-06  
MAXIMUM NUMBER OF ITERATIONS : 25  
VALUE OF ALPHA (CONFIDENCE LEVEL %) : 0.05000 (95.00000 %)  
OPTION OUTPUT REPLICATE COEFFICIENTS : OFF  
OPTION OUTPUT ITERATION HISTORY : OFF

MODEL(S) : \_TOTALPA2\_04 = P3\_02 P5\_02 C8I\_02 \_AGE

NUMBER OF REPLICATES : 60  
NUMBER OF OBSERVATIONS READ : 2256  
WEIGHTED NUMBER OF OBSERVATIONS READ : 20699186.392

MODEL : \_TOTALPA2\_04 = P3\_02 P5\_02 C8I\_02 \_AGE

OPTIONS : Intercept,  
No Standardized Coefficient,  
Degrees of Freedom = 60  
t VALUE : 2.000  
STARTING VALUES : INTERCEPT : 0.0000  
P3\_02 : 0.0000  
P5\_02 : 0.0000  
C8I\_02 : 0.0000  
\_AGE : 0.0000  
BY : None Specified.

```

MISSING :      0          (UNWEIGHTED)
           0.000000      (WEIGHTED)
NONMISSING : 2256        (UNWEIGHTED)
           20699186.391535 (WEIGHTED)
Success = records with dependent value equal to      0 : 326          (UNWEIGHTED)
           3431704.959178 (WEIGHTED)
Failure = records with dependent value equal to      1 : 1930        (UNWEIGHTED)
           17267481.432357 (WEIGHTED)
ITERATIONS REQUIRED FOR FULL SAMPLE :      5
MAXIMUM ITERATIONS FOR REPLICATE SAMPLE :      5
-2 LOG LIKELIHOOD FOR FULL SAMPLE :      18038648.46897
-2 LOG LIKELIHOOD FOR MODEL CONTAINING INTERCEPT ONLY :      18593933.03881
  
```

MEASURES OF FIT

```

Negative log-likelihood:      0.030
Likelihood ratio(Cox-Snell):      0.026 Maximum possible value:      0.593
Likelihood ratio(Estrella):      0.027
  
```

ESTIMATED FULL SAMPLE REGRESSION COEFFICIENTS

PARAMETER	ESTIMATE	STANDARD ERROR OF ESTIMATE	TEST FOR H0: PARAMETER=0	PROB> T	COMMENT
INTERCEPT	-3.10	0.556	-5.581	0.000	
P3_02	0.21	0.083	2.531	0.014	
P5_02	0.29	0.097	2.953	0.004	
C8I_02	-0.27	0.076	-3.579	0.001	
_AGE	0.09	0.048	1.822	0.073	

## CORRELATIONS OF ESTIMATES

	INTERCEPT	P3_02	P5_02	C8I_02	_AGE
INTERCEPT	1.000	-0.140	-0.223	-0.119	-0.868
P3_02	-0.140	1.000	-0.348	-0.129	-0.081
P5_02	-0.223	-0.348	1.000	0.263	-0.043
C8I_02	-0.119	-0.129	0.263	1.000	-0.180
_AGE	-0.868	-0.081	-0.043	-0.180	1.000

## HYPOTHESIS TESTING RESULTS

TEST	F VALUE	NUM. DF	DENOM. DF	PROB>F	NOTE
OVERALL FIT	10.623		4	57	0.000
P3_02	6.404	1	60		0.014
P5_02	8.721	1	60		0.004
C8I_02	12.810	1	60		0.001
_AGE	3.318	1	60		0.073

## ODDS RATIO RESULTS

PARAMETER	ESTIMATE	LOWER 95%	UPPER 95%	NOTE
P3_02	1.24	1.045	1.459	
P5_02	1.33	1.097	1.617	
C8I_02	0.76	0.656	0.888	
_AGE	1.09	0.991	1.201	



## Attachment 6-2. Output from SUDAAN Example

The SAS System

18:46 Wednesday, March 22, 2006 18

59

S U D A A N  
Software for the Statistical Analysis of Correlated Data  
Copyright        Research Triangle Institute    February 2005  
                  Release 9.0.1

Independence parameters have converged in 5 iterations

Number of observations read	:	2256	Weighted count:	20699186
Observations used in the analysis	:	2256	Weighted count:	20699186
Denominator degrees of freedom	:	60		

Maximum number of estimable parameters for the model is 5

Sample and Population Counts for Response Variable \_TOTALPA2\_04

1:	Sample Count	1930	Population Count	17267481
0:	Sample Count	326	Population Count	3431705

Date: 03-23-2006  
Time: 11:00:11

Research Triangle Institute  
The MULTLOG Procedure

Page : 1  
Table : 1

Frequencies and Values for CLASS Variables  
by: \_DAYSTOTR IN 2 CATEGORIES.

```
-----  
_DAYSTOTR IN  
  2  
  CATEGORIES      Frequency      Value  
-----  
Ordered  
  Position:  
  1                1930          1  
Ordered  
  Position:  
  2                 326          0  
-----
```

Date: 03-23-2006  
 Time: 11:00:11

Research Triangle Institute  
 The MULTILOG Procedure

Page : 2  
 Table : 1

Variance Estimation Method: Replicate Weight Jackknife  
 Working Correlations: Independent  
 Link Function: Cumulative Logit  
 Response variable \_TOTALPA2\_04: \_DAYSTOTR IN 2 CATEGORIES  
 by: \_TOTALPA2\_04 (cum-logit), Independent Variables and Effects.

-----				
_TOTALPA2_04 (cum- logit), Independent Variables and Effects	Beta Coeff.	SE Beta	T-Test B=0	P-value T-Test B=0
-----				
_TOTALPA2_04 (cum- logit) Intercept 1	3.10	0.56	5.58	0.0000
IMPORTANT CHILD PARTIC IN ORGANIZED PA	-0.21	0.08	-2.53	0.0140
IMPORTANT CHILD PARTIC IN FREE TIME PA	-0.29	0.10	-2.95	0.0045
I SHOULD DO MORE PA THAN I DO	0.27	0.08	3.58	0.0007
CHILD'S AGE IN 2002	-0.09	0.05	-1.82	0.0735
-----				

Date: 03-23-2006  
Time: 11:00:11

Research Triangle Institute  
The MULTLOG Procedure

Page : 3  
Table : 1

Variance Estimation Method: Replicate Weight Jackknife  
Working Correlations: Independent  
Link Function: Cumulative Logit  
Response variable \_TOTALPA2\_04: \_DAYSTOTR IN 2 CATEGORIES  
by: Contrast.

Contrast	Adj Wald F	Degrees of Freedom	P-value Adj Wald F
OVERALL MODEL	104.56	5	0.0000
MODEL MINUS			
INTERCEPT	10.62	4	0.0000
P3_02	6.40	1	0.0140
P5_02	8.72	1	0.0045
C8I_02	12.81	1	0.0007
_AGE	3.32	1	0.0735

## References

- Brick, J.M., and Broene, P. (1997). Unit and item response rates, weighting, and imputation procedures in the 1995 National Household Education Survey. Working Paper 97-06. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- Casady, R., and Lepkowski, J. (1993). Stratified telephone survey designs. Survey Methodology, 19, 103-113.
- Deming, W.E., and Stephan, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Annals of Mathematical Statistics, 11, 427-444.
- Judkins, D. R., Chu, A., DiGaetano, R., & Shapiro, G. (1999). Coverage in screening surveys at Westat. Proceedings of the Section on Survey Research Methods of the American Statistical Association, pp. 581-586.
- Kalton, G., and Kasprzyk, D. (1986). The treatment of missing survey data. Survey Methodology, 12, 1-16.
- Korn, E.L., and Graubard, B.I. (2003). Estimating variance components by using survey data. Journal of the Royal Statistical Society, Series B, 175-190.
- Rao, J.N.K., and Scott, A. J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. Journal of the American Statistical Association, 76, 221-230.
- Rao, J.N.K., and Scott, A.J. (1984). On chi-square tests for multiway contingency tables with cell proportions estimated from survey data. The Annals of Statistics, 12, 46-60.
- Särndal C.E., Swensson B, Wretman J. (1992). Model assisted survey sampling. Springer-Verlag: New York.
- Shapiro, G., Battaglia, M., Camburn, D., Massey, J., and Tompkins, L. (1995). Calling local telephone company business offices to determine the residential status of a wide class of unresolved telephoned numbers in random-digit dialing sample. Unpublished long version of paper in Proceedings of the Survey Research Methods Section, American Statistical Association, 713-718.
- SPSS (1993). SPSS for windows CHAID, release 6.0. Chicago: SPSS, Inc.
- Westat (2000). A user's guide to WesVarPC, version 4.0. Rockville, MD: Westat



**APPENDIX A**

**YMCLS MASTER QUESTIONNAIRE**





**YOUTH MEDIA CAMPAIGN LONGITUDINAL SURVEY  
MASTER QUESTIONNAIRE  
SCREENER**

S1. Hello, this is (INTERVIEWER) and I'm calling for the Centers for Disease Control and Prevention, the CDC, about a research study. Are you a member of this household and at least 18 years old?

- YES ..... 1 (GO TO S4)
- NO ..... 2 (GO TO S2)
- PROBABLE BUSINESS ..... 3 (GO TO S4)

S2. May I please speak with a household member who is at least 18 years old?

- AVAILABLE ..... 1 (GO TO S1)
- NOT AVAILABLE ..... 2 (GO TO RESULT, CALLBACK APPT.)
- THERE ARE NONE ..... 3 (GO TO THANK3)

S4. Is this phone number used for...

- Home use, ..... 1 (CONTINUE)
- Home and business use, or ..... 2 (CONTINUE)
- Business use only? ..... 3 (GO TO THANK1)

S5. The CDC is conducting a voluntary and confidential research study about children's participation in physical and social activities. Are any of the people who normally live in your household age 9 to 13?

- YES ..... 1 (GO TO S6)
- NO ..... 2 (GO TO S20)
- REFUSAL ..... -7 (GO TO THANK3)
- DON'T KNOW ..... -8 (GO TO THANK3)

S6. Please tell me just the first names and ages of children age 9 to13 living in your household. Let's start with the oldest.

What is his or her first name/the name of the next child?	Is this child male or female?	How old is [he/she]?

S7. Have we missed anyone else age 9 to13 who usually lives here but is temporarily away from home? For example, away on vacation, or away at school?

- MATRIX CORRECT ..... 1
- RETURN TO MATRIX ..... 2

- CATI randomly selects 2 children age 9-13 if more than 2.
- Refer to Gary Shapiro's memo dated 1/15/2002

S8. We'd like to ask some questions about the physical and social activities that [CHILD1]/[CHILD2] participates in. Are you a parent of [CHILD1]?

- YES ..... 1 (GO TO S9)
- NO ..... 2 (GO TO S9)
- YES, A PARENT, BUT CANNOT ANSWER.. 3 (GO TO S9)
- REFUSED ..... -7 (GO TO THANK3)
- DON'T KNOW ..... -8 (GO TO THANK3)

**[IF NECESSARY GIVE DEFINITION OF ACTIVITIES: We are interested in physically active things like team sports or outdoor play, group activities like youth groups or volunteer programs, and family activities, both in and outside of the home.]**

BOX 1

If Child1 and S8 is YES, go to S9 and display "your first name?".  
 If Child1 and S8 is NO, go to S9 and display "the first name of the parent or guardian in this household who could tell us about [CHILD1]'s activities?"

S9. What is [your first name?/ the first name of the parent or guardian in this household who could tell us about (CHILD1)'s activities?]

- X IF SCREENER RESPONDENT
- NAME \_\_\_\_\_ [ ]
- REFUSED ..... -7
  - DON'T KNOW ..... -8

S10. What is [your/his/her] relationship to [CHILD1]?

- MOTHER (ADOPTED, BIRTH, STEP, FOSTER) 1 (GO TO BOX 2)
- FATHER (ADOPTED, BIRTH, STEP, FOSTER) 2 (GO TO BOX 2)
- GRANDMOTHER ..... 3 (GO TO BOX 2)
- GRANDFATHER ..... 4 (GO TO BOX 2)
- AUNT ..... 5 (GO TO BOX 2)
- UNCLE ..... 6 (GO TO BOX 2)
- OTHER (SPECIFY)\_\_\_\_\_ 91 (GO TO BOX 2)
- REFUSED ..... -7 (GO TO BOX 2)
- DON'T KNOW ..... -8 (GO TO BOX 2)

BOX 2

If both respondent name (S9) and respondent relationship to CHILD1 (S10) are missing, go to THANK3.  
 If CHILD1 only, go to BOX4. If CHILD2 also, continue with S11.

S11. Are you a parent or guardian of [CHILD2]?

- YES ..... 1 (GO TO BOX3)
- NO ..... 2 (GO TO BOX3)
- YES, A PARENT, BUT CANNOT ANSWER.. 3 (GO TO BOX 3)
- REFUSED ..... -7 (GO TO THANK3)
- DON'T KNOW ..... -8 (GO TO THANK3)

BOX 3

If Child2 and S11 is YES, go to S12 and display "VERIFY PARENT'S NAME".  
If Child2 and S11 is NO or YES BUT CANNOT ANSWER, go to S12 and display "What is the first name of a parent or guardian in this household who could tell us about [CHILD2]'s activities?"

S12. [VERIFY PARENT'S NAME./What is the first name of the parent or guardian in this household who could tell us about(CHILD2)'s activities?]

- (DISPLAY NAME FROM S9)..... 1
- X IF SCREENER RESPONDENT
- SOMEONE ELSE, SPECIFY\_\_\_\_\_ 2 [ ]
- REFUSED ..... -7
- DON'T KNOW ..... -8

S13. What is [your/his/her] relationship to [CHILD2]?

- MOTHER (ADOPTED, BIRTH, STEP, FOSTER) 1 (GO TO BOX4)
- FATHER (ADOPTED, BIRTH, STEP, FOSTER) 2 (GO TO BOX4)
- GRANDMOTHER ..... 3 (GO TO BOX4)
- GRANDFATHER ..... 4 (GO TO BOX4)
- AUNT ..... 5 (GO TO BOX4)
- UNCLE ..... 6 (GO TO BOX4)
- OTHER (SPECIFY)\_\_\_\_\_ 91 (GO TO BOX4)
- REFUSED ..... -7 (GO TO BOX4)
- DON'T KNOW ..... -8 (GO TO BOX4)

BOX 4

If both respondent name (S11) and respondent relationship to CHILD2 (S12) are missing, go to THANK3.  
If any children are selected, go to parent interview.

S20. Because we are conducting this study by the phone, I have some questions about the telephone numbers in your household. Besides the number I called, do you have other telephone numbers in your household, not including cell phones?

- YES..... 1 (GO TO S21)
- NO..... 2 (GO TO S24)
- REFUSED..... -7 (GO TO S24)
- DON'T KNOW..... -8 (GO TO S24)

S21. Including computer and fax phone numbers, how many of these additional phone numbers are for home use?

[IF NEEDED: Do not include cell phones.]

- NUMBER..... \_\_ \_\_ (GO TO BOX 5)
- REFUSED .....-7
- DON'T KNOW .....-8

BOX 5

If S21= 0 (other telephone numbers for home use), go to S29  
If S21=1, go to S22.  
If S21= 2-19 (hard range), 2-7 (soft range) go to S23.  
If Refused/Don't Know, go to S29.

S22. Is this additional phone number used for a computer or fax machine?

- YES ..... 1 GO TO S26ov
- NO..... 2 GO TO S 29
- REFUSED .....-7 GO TO S 29
- DON'T KNOW .....-8 GO TO S29

S23. Of these (DISPLAY NUMBER) additional home use phone numbers, how many are used for a computer or fax machine?

- NUMBER..... \_\_\_\_ \_\_\_\_ (GO TO BOX 6)
- REFUSED .....-7 (GO TO THANK3)
- DON'T KNOW .....-8 (GO TO THANK3)

BOX 6

If S23 = 0, go to S29  
If S23 = 1, go to S26ov  
If S23 = 2 – 19, go to S25  
If Refused/Don't Know, go to S29

S24. Do you have any additional phone numbers for computers or fax machines?

YES .....1 (GO TO S25)  
 NO.....2 (GO TO S29)  
 REFUSED .....-7 (GO TO S29)  
 DON'T KNOW .....-8 (GO TO S29)

S25. Some households have telephone numbers that are used both for talking and for computer or fax.  
 How many of these (DISPLAY NUMBER) computer or fax numbers are ever answered for talking?

NUMBER \_\_\_\_ .....(GO TO BOX 7)  
 REFUSED .....-7 (GO TO S29)  
 DON'T KNOW .....-8 (GO TO S29)

BOX 7

If S25 = 0, go to S29.  
 If S25 = 1, go to S27.  
 If S25 = 2-19, go to S28.  
 If Refused/Don't Know go to S29.

S26ov. Some households have telephone numbers that are used both for talking and for computer or fax.  
 Is this number ever answered for talking?

YES .....1 (GO TO S27)  
 NO.....2 (GO TO S29)  
 REFUSED .....-7 (GO TO S29)  
 DON'T KNOW .....-8 (GO TO S29)

S27. Is this computer or fax number answered for ...

Personal calls, .....1 (GO TO S29)  
 Business calls, or .....2 (GO TO S29)  
 Both?.....3 (GO TO S29)  
 REFUSED .....-7 (GO TO S29)  
 DON'T KNOW .....-8 (GO TO S29)

S28. Of these (DISPLAY NUMBER) phone numbers that are answered, how many are answered for non-business related calls?

NUMBER..... \_\_\_\_\_  
 REFUSED .....-7  
 DON'T KNOW .....-8

S29. During the past 12 months, has your household ever been without telephone service for more than 24 hours?

[IF NEEDED: Do not include temporary loss of service due to storms, damaged wires, or phone company maintenance.]

- YES..... 1 (GO TO S29ov)
- NO..... 2 (GO TO HHSELECT)
- REFUSED.....-7 (GO TO HHSELECT)
- DON'T KNOW.....-8 (GO TO HHSELECT)

S29ov. What was the total amount of time your household was without telephone service for more than 24 hours?

- NUMBER: \_\_\_\_\_ UNIT: ( ) (GO TO HHSELECT)
- 1. DAYS
  - 2. WEEKS
  - 3. MONTHS

THANK 1. Thank you, but we are only interviewing in residences.

THANK2. Thank you so much for answering these questions. We are only interviewing in households with children age 9 to 13.

THANK3. These are all the questions I have at this time. Thank you very much for your time.

**YOUTH MEDIA CAMPAIGN LONGITUDINAL SURVEY**  
**MASTER QUESTIONNAIRE**  
**PARENT INTERVIEW**

This questionnaire contains all questions asked in the YMCLS.

1 = Asked in 2002 only	6 = Asked in 2004 only	11 = Asked in 2006 only
2 = Asked in 2003-2006	7 = Asked in 2004-2006	12 = Asked in 2002-2003 & 2006
3 = Asked in 2005-2006	8 = Asked in 2002-2004	13 = Asked in 2004 & 2005
4 = Asked in 2002-2003	9 = Asked in 2003-2004	
5 = Asked in 2003 only	10 = Asked in 2005 only	

**Panel One, 2002 Introduction**

EX\_PAR

[Hello, this is (INTERVIEWER) and I'm calling for the Centers for Disease Control and Prevention, the CDC, about a voluntary and confidential research study about children's participation in physical and social activities. You can skip any question you don't want to answer or stop at any time.]

(Now) I'd like to ask you some questions related to activities (CHILD1/CHILD2) may have participated in like sports, clubs, and family activities. (It will take about 10 minutes.) Then I would like to ask [CHILD1 and CHILD2] similar questions. If you're ready, let's begin (with questions about [CHILD1]).

If s6. > 1, complete interview about CHILD1, then repeat child-specific questions about CHILD2.

**Panel Two, 2004 & Panel Three, 2006 Introduction**

EX\_PAR CROSS SEC

[Hello, this is (INTERVIEWER) and I'm calling for the Centers for Disease Control and Prevention, the CDC, about a voluntary and confidential research study about children's participation in physical and social activities.]

(Now) I'd like to ask you some questions related to activities (CHILD1/CHILD2) may have participated in like sports, clubs, and family activities. You can skip any questions you don't want to answer or stop at any time. Your answers will be handled in a confidential manner and will help us better understand children's health. The interview will take about 15 minutes, and then I would like to ask [CHILD1/CHILD1 and CHILD2] similar questions. If you're ready, let's begin (with questions about [CHILD1]).]

**Introduction for all follow up years**

EX\_PAR FOLLOW-UP

[FROM CONTACT SPECS: Hello, this is (INTERVIEWER) and I'm calling for the Centers for Disease Control and Prevention, the CDC. Last year we spoke with (YOU/CHILD'S MOTHER) about children's participation in physical activities].

This is a voluntary research study. You can skip any questions you don't want to answer or stop at any time. Your answers will be handled in a confidential manner and will help us better understand children's health. The interview will take about 15 minutes, and then I would like to ask [CHILD1/CHILD1 and CHILD2] similar questions. If you're ready, let's begin (with questions about [CHILD1]).]

EX\_PAR2

Now I'd like to ask about [CHILD2].

P1.<sup>13</sup> First, I'd like to confirm [CHILD1/CHILD2]'s age. In what month and year was (he/she) born?

MONTH \_\_\_\_\_ YEAR \_\_\_\_\_

1	January	7	July
2	February	8	August
3	March	9	September
4	April	10	October
5	May	11	November
6	June	12	December

Calculate AGE2002 = child's age. Calculate current age for display in 1. If current age does not match screener age or birth month is current month, go to P2. Else, go to box after P2a.

P\_1.<sup>14</sup> First, I'd like to confirm [CHILD1/CHILD2]'s age. Is [he/she] (INSERT AGE) years old?

YES ..... 1 (GO TO BOX AFTER P2a)  
NO ..... 2 (GO TO P2a)

P2.<sup>1</sup> That would mean that [CHILD1/CHILD2] [is (AGE/turns AGE) this month]. Is that right?

YES ..... 1 (GO TO BOX AFTER P2a)  
NO ..... 2 (GO TO P2a)

P2a. How old is [CHILD1/CHILD2]?

AGE \_\_\_\_\_  
(IF CHILD IS OUTSIDE OF TARGET AGE AT FIRST INTERVIEW, GO TO THANK5)

If CHILD1, go to PINTRO1. Else, if CHILD2, go to PINTRO2.

#### PINTRO1

Now, I'd like to explain the types of physical activities I will be asking about. The first questions are about organized groups like sports teams and physical activity classes. The next questions are about unstructured physical activities (CHILD1/CHILD2) may do in [his/her] free time for fun.

For all of the questions, please do not include the time spent in activities during the school day like PE, gym class, or recess.

#### PINTRO2

First, I'll ask about physical activities in an organized group, and then I'll ask about unstructured physical activities.

For all of the questions, please do not include the time spent in activities during the school day like PE, gym class, or recess.

<sup>13</sup> Asked in 2002 only.

<sup>14</sup> Asked in 2003-2006.



P3. Now, think about the sports teams and physical activity classes (CHILD1/CHILD2) may have participated in during the past school year like basketball, martial arts, or swimming lessons.

How important is it to you that [CHILD1/CHILD2] participates in sports teams or physical activity classes? Would you say....

- Extremely important, ..... 1 (GO TO P4)
- Very important, ..... 2 (GO TO P4)
- Somewhat important, or..... 3 (GO TO P4)
- Not very important ..... 4 (GO TO P4)
- R VOLUNTEERS CHILD HAS A DISABILITY (NOT A SHORT-TERM ILLNESS OR INJURY) ..... 5 (GO TO P\_D1)

P\_D1 and P\_D2 asked only if parent volunteers that child has a disability.

P\_D1. Does (CHILD1/CHILD2)'s disability or condition prevent (him/her) from participating in sports or other unstructured physical activities (he/she) may do for fun?

- YES ..... 1 (GO TO P\_D2)
- NO ..... 2 (GO TO P\_D2)

P\_D2. Does (CHILD1/CHILD2)'s disability or condition prevent (him/her) from participating in school clubs and community groups?

- YES ..... 1 (GO TO BOX AFTER P\_D2)
- NO ..... 2 (GO TO BOX AFTER P\_D2)

If P\_D1= 1 and P\_D2 =1, go to THANK4.  
Else, if P\_D1= 2 and P\_D2 =1, go to P3.  
Else if P\_D1= 1 and P\_D2 =2, go to P8INTRO (in 2002) or P10INTRO (in other years).

P4. How confident are you that you can influence [his/her] involvement in these activities? Are you...

[IF NEEDED: By "influence involvement" we mean getting (CHILD1/CHILD2) to begin, maintain, or do more of these activities].

- Extremely confident, ..... 1
- Very confident,..... 2
- Somewhat confident, or ..... 3
- Not very confident ..... 4

P5INTRO

Now think about the physical activities [CHILD1/CHILD2] may have done in [his/her] free time during the past school year, things like riding bikes, playing actively outside with friends, or shooting hoops.

P5. How important is it to you that [he/she] do these activities? Would you say...

- Extremely important, ..... 1 (GO TO P6)
- Very important, ..... 2 (GO TO P6)
- Somewhat important, or..... 3 (GO TO P6)
- Not very important ..... 4 (GO TO P6)
- R VOLUNTEERS CHILD HAS A DISABILITY  
(NOT A SHORT-TERM ILLNESS  
OR INJURY) ..... 5 (GO TO P\_D1)

P6. How confident are you that you can influence [CHILD1/CHILD2]'s involvement in these activities? Are you...

[IF NEEDED: By "influence involvement" we mean getting (CHILD1/CHILD2) to begin, maintain, or do more of these activities].

- Extremely confident, ..... 1
- Very confident,..... 2
- Somewhat confident, or ..... 3
- Not very confident..... 4

P8INTRO

The next questions are about [CHILD1/CHILD2]'s involvement in other kinds of group activities both during and after school.

P8.<sup>1</sup> For these questions, think about [CHILD1/CHILD2]'s involvement in school groups like band, drama club, or student government and community groups like Scouts or 4-H.

How important is it to you that [he/she] be involved in school or community groups? Would you say...

- Extremely important, ..... 1
- Very important, ..... 2
- Somewhat important, or..... 3
- Not very important ..... 4

P9.<sup>15</sup> How confident are you that you can influence [CHILD1/CHILD2]'s involvement in these activities? Are you...

<sup>1</sup> Asked in 2002 only.

<sup>15</sup> Asked in 2002 only.

[IF NEEDED: By "influence involvement" we mean getting (CHILD1/CHILD2) to begin, maintain, or do more of these activities].

- Extremely confident, ..... 1
- Very confident,..... 2
- Somewhat confident, or ..... 3
- Not very confident..... 4

P10INTRO

The next few questions are about family activities you or another adult in your household may have done with [CHILD1/CHILD2] in the past 7 days, since last [DAY].

P10. In the past 7 days, how many days did you or another adult in your household do any physical activities with [CHILD1/CHILD2] including things like active games, sports or other physical activities, and so forth? Please include only activities where both you and [CHILD1/CHILD2] were active.

[IF NEEDED: This also includes activities where both (CHILD1/CHILD2) and another adult in your household were active.]

NUMBER OF DAYS \_\_\_\_\_  
(ENTER # 0 to 7)

P\_10.<sup>2</sup> Now, thinking about the next month, how likely are you to do physical activities with [CHILD1/CHILD2]? Are you...

- Extremely likely..... 1
- Very likely ..... 2
- Somewhat likely, or..... 3
- Not very likely? ..... 4

P11. During the past 7 days, how many days did you or another adult in your household have the evening meal with [CHILD1/CHILD2] sitting around a table?

NUMBER OF DAYS \_\_\_\_\_  
(ENTER # 0 to 7)

P\_11a.<sup>2</sup> Now, the next question is about [CHILD1/CHILD2]'s school. How many miles does [he/she] live from school?

[IF NEEDED: Please make your best guess.]

NUMBER OF MILES \_\_\_\_\_

---

<sup>2</sup> Asked in 2003-2006.

P\_11b.<sup>2</sup> (IF ONE MILE OR LESS) In a usual week, how many days does [CHILD1/CHILD2] walk or bike to or from school?

NUMBER OF DAYS \_\_\_\_\_  
(ENTER # 0 to 7)

REFUSED ..... -7  
DON'T KNOW ..... -8

P\_11c.<sup>2</sup>

The next few questions are about TV and video games.

How often do you set limits on the amount of time [CHILD1/CHILD2] watches television? Would you say...

Always, ..... 1  
Very often, ..... 2  
Sometimes, ..... 3  
Rarely, or ..... 4  
Never? ..... 5

P\_11d.<sup>2</sup> How often do you set limits on the amount of time [CHILD1/CHILD2] plays video games? Would you say...

Always, ..... 1  
Very often, ..... 2  
Sometimes, ..... 3  
Rarely, or ..... 4  
Never? ..... 5

The following questions are asked once regardless of number of children interviewed.

P\_11e.<sup>2</sup> Now please just answer yes or no. Do you have cable television in your household?

YES ..... 1  
NO ..... 2

---

<sup>2</sup> Asked in 2003-2006.

P12INTRO

These next questions are about how parents might be involved in their children's activities after school and on the weekend.

P12. During the past school year, did you or another adult in your household\* coach a team or lead a group [CHILD1 participates in / CHILD1 or CHILD2 participate in]?

- YES ..... 1
- NO ..... 2

P13. During the past school year, how often did you or another adult in your household\*.... Would you say....

	ALWAYS	VERY OFTEN	SOME-TIMES	RARELY	NEVER
a. Attend the events related to [CHILD1/ CHILD1 or CHILD2]'s sports, clubs, or other activities? .....	1	2	3	4	5
b. Take [CHILD1/ CHILD1 or CHILD2] to and from activities? .....	1	2	3	4	5

P\_13c.<sup>2</sup> In the next year, how likely are you or another adult in your household to coach a team or lead a group that [CHILD1/ CHILD1 or CHILD2] participates in? Are you...

- Extremely likely..... 1
- Very likely ..... 2
- Somewhat likely, or..... 3
- Not very likely? ..... 4

P\_13d.<sup>2</sup> In the next month, how likely are you or another adult in your household to attend the events related to [CHILD1/ CHILD1 or CHILD2]'s sports, clubs, or other activities? Are you...

- Extremely likely..... 1
- Very likely ..... 2
- Somewhat likely, or..... 3
- Not very likely? ..... 4

P\_13e.<sup>2</sup> For this question, please think about your physical activities in the past 7 days. On how many days, since last [DAY], did you do physical activities or exercise such as brisk walking, jogging, or biking for at least 10 minutes at a time?

NUMBER OF DAYS \_\_\_\_\_  
(ENTER # 0 to 7)

\* "or another adult in your household" added after 2002.

<sup>2</sup> Asked in 2003-2006.

P7. For these next questions, think about the physical activities and sports that kids in general may do and tell me whether you strongly agree, agree, disagree, or strongly disagree.

	STRONGLY AGREE	AGREE	DISAGREE	STRONGLY DISAGREE
P7a. Kids who do regular physical activities have more self-confidence. Do you strongly agree, agree, disagree, or strongly disagree? .....	1	2	3	4
P7b. Kids who do regular physical activities are healthy. ....	1	2	3	4
P7c. <sup>1</sup> Kids who do physical activities have problems in school.....	1	2	3	4
P7d. Kids who do regular physical activities will be healthier adults.....	1	2	3	4
P_7e. <sup>2</sup> I think it's a parent's responsibility to help their children find physical activities to do. ....	1	2	3	4
P_7f. <sup>2</sup> I think all kids should be physically active every day.. ....	1	2	3	4
P_7g. <sup>2</sup> Parents play an important role in whether their kids are physically active when they grow up.....	1	2	3	4

P14. Thinking about all of the activities that [CHILD1 does/ CHILD1 and CHILD2 do], please tell me whether you strongly agree, agree, disagree, or strongly disagree with each of the following statements.

[INSERT STATEMENTS A THROUGH E. READ SCALE IF NECESSARY]

	STRONGLY AGREE	AGREE	DISAGREE	STRONGLY DISAGREE
a. Transportation problems prevent my [child/children] from participating in activities [he/she/they] would like to do. Do you strongly agree, agree, disagree, or strongly disagree? .....	1	2	3	4
b. There are plenty of opportunities for my [child/children] to participate in activities where I live.....	1	2	3	4
c. My [child is/children are] not able to participate in some activities because they are too expensive. ....	1	2	3	4
d. I have enough time to keep my [child/children] involved in activities.....	1	2	3	4
e. I feel comfortable letting my [child/children] play outside near our home. ....	1	2	3	4

P\_14a.<sup>2</sup>

For these next few questions, we would like you to just make your best guess.

Experts recommend that children watch no more than how many hours of television per day?

NUMBER OF HOURS \_\_\_\_\_  
(ENTER # 1 to 10)

<sup>1</sup> Asked in 2002 only.

<sup>2</sup> Asked in 2003-2006.

P\_14b.<sup>2</sup> Experts recommend that children participate in physical activity how many days per week?

NUMBER OF DAYS \_\_\_\_\_  
(ENTER # 0 to 7)

P\_14c.<sup>2</sup> Experts recommend that children do how many minutes per day when they participate in physical activity?

NUMBER OF MINUTES \_\_\_\_\_  
(ENTER # 0 to 240)

P\_14d.<sup>2</sup>

Next we have a few questions about messages and advertising you may have seen.

Have you seen, read, or heard about any messages or advertising for kids getting active?

YES ..... 1 (GO TO P\_14e)  
NO ..... 2 (GO TO P15)

P\_14e.<sup>2</sup> What is the name of the message or advertising?

[PROBE: If something other than VERB mentioned: Are there any others for kids getting active?  
PROBE AGAIN: If second message other than VERB mentioned: Are there any others? Record final answer.] Do not probe Don't Know.

VERB ..... 1  
OTHER (SPECIFY) ..... 91

In 2003-2006, if P\_14e = 1, do not ask P15d.

P15. There are many ads on television, radio, and in newspapers and magazines with slogans you may or may not remember. Have you heard, read, or seen any ads with the slogan...[KEEP A FIRST; ROTATE B THROUGH E.]

	YES	NO
a. "Just do it"? .....	1	2
b. How about "Parents: The Anti-Drug"? .....	1	2
c. <sup>1</sup> How about "Come on Home"? .....	1	2
d. How about "Verb. It's What You Do*."? .....	1	2
e. How about "Akimbo"? .....	1	2

If P\_14e = 1 or 2 or P15d = 1, go to P16 in 2003-2004/ P\_15x in 2005-2006.  
Else, go to P16.

P\_15x.<sup>3</sup> Please tell me all of the places you have seen, read or heard about VERB.

\_\_\_\_\_

<sup>2</sup> Asked in 2003-2006.

<sup>1</sup> Asked in 2002 only.

\* "It's What You Do" added after 2002

<sup>3</sup> Asked in 2005-2006.

[DO NOT READ LIST. CODE ALL RESPONSES.] [RECORD ANSWERS BELOW UNDER P\_15x.]

P_15XTV	On television .....	1
P_15XRAD	On the radio .....	2
P_15XINT	On the internet .....	3
P_15XFRN	From friends .....	4
P_15XCHL	From your children .....	5
P_15XSCH	At school/from teacher. ....	6
P_15XMAG	In magazines .....	7
P_15XSIG	On signs, billboards, or posters.....	8
P_15XEVN	At an event, concert, or the mall.....	9
P_15XCLB	At an after-school or community club, like the YMCA.....	10
P_15XNEW	In newspapers.....	11
P_15XOTH	Anywhere else? Other: Specify_____	91

P16 though P18 were asked of the parent respondent the first time he or she was interviewed.

P16INTRO

Now I'd like to ask you a few questions about you and your household.

P16. Are you currently...

Married, .....	1
Widowed, .....	2
Divorced, .....	3
Separated, or .....	4
Never Married .....	5



P17. What is the highest grade or level of school you completed? [DO NOT READ LIST. ACCEPT ONLY ONE RESPONSE.]

- NONE, OR GRADES 1-4 ..... 1 (GO TO P18)
- GRADES 5,6,7 ..... 2 (GO TO P18)
- GRADE 8 ..... 3 (GO TO P18)
- SOME HIGH SCHOOL (GRADES 9-11)..... 4 (GO TO P18)
- GRADE 12, HIGH SCHOOL DIPLOMA,  
OR GED..... 5 (GO TO P18)
- TRADE/TECHNICAL/VOCATIONAL  
TRAINING BEYOND H.S..... 6 (GO TO P18)
- SOME COLLEGE OR UNIVERSITY  
BUT NO DEGREE ..... 7 (GO TO BOX AFTER  
P18)
- ASSOCIATE'S DEGREE (AA) ..... 8 (GO TO BOX AFTER  
P18)
- BACHELOR'S DEGREE (B.A., B.S.) ..... 9 (GO TO BOX AFTER  
P18)
- SOME GRADUATE OR PROFESSIONAL  
SCHOOL BUT NO DEGREE ..... 10 (GO TO BOX AFTER  
P18)
- GRADUATE/PROFESSIONAL DEGREE  
(MA,MS,PHD,MD,DDS,JD)..... 11 (GO TO BOX AFTER  
P18)

P18. Do you have a high school diploma or its equivalent, a GED?

- YES ..... 1
- NO ..... 2

Race/ethnicity was asked during the first interview about the child. P19 was administered in 2002 only. Subsequent interviews used P19a and P19b to conform to OMB guidelines.

P19. Do you describe [CHILD1/CHILD2]'s race as... [READ 1-6. CODE ALL THAT APPLY.]

- P19WHITE White, ..... 1
- P19BLACK Black or African-American, ..... 2
- P19HISP Hispanic or Latino ..... 3
- P19ASIAN Asian..... 4
- P19HAWAI Native Hawaiian or Other  
Pacific Islander, or ..... 5
- P19AMIND American Indian or Alaska Native..... 6
- P19OTHR OTHER (SPECIFY)\_\_\_\_\_ 91

P19a. Is (CHILD1/CHILD2) of Spanish, Hispanic, or Latino origin?

- P19HISP YES ..... 1
- NO ..... 2

P19b. What is (his/her) race? You may choose more than one. [CODE ALL THAT APPLY.]

- P19WHITE White, ..... 1

P19BLACK	Black or African-American, .....	2
P19AMIND	American Indian or Alaska Native, .....	3
P19ASIAN	Asian, or .....	4
P19HAWAI	Native Hawaiian or Pacific Islander? .....	5
P19OTHR	OTHER (SPECIFY)_____	91

P20. How many adults age 18 or over are currently living in your household including yourself? [ACCEPT ONLY ONE RESPONSE.]

NUMBER OF ADULTS \_\_\_\_\_  
(ENTER # 1 to 20)

P21. [Is (CHILD1/CHILD2)'s other parent currently living/Does either of (CHILD1/CHILD2)'s parents currently live] in the household with (CHILD1/CHILD2)?

YES ..... 1  
NO ..... 2

P22. How many children under age 18 are currently living in your household? [ACCEPT ONLY ONE RESPONSE.]

NUMBER OF CHILDREN \_\_\_\_\_  
(ENTER # 1 to 20)

P23 was asked of the parent respondent the first time he or she was interviewed.

P23. What was the total income of all persons in your household over the past year, including salaries or other earnings, interest, retirement, and so on for all household members? Was it...

\$25,000 or less, or ..... 1 (GO TO SET 1)  
More than \$25,000..... 2 (GO TO P23A)

[P23a.]  
Was it...  
\$50,000 or less, or ..... 1 (GO TO SET 2)  
More than \$50,000..... 2 (GO TO SET 3)

[P23SET 1]  
\$5,000 or less ..... 1  
\$5,001 to \$10,000..... 2  
\$10,001 TO \$15,000..... 3  
\$15,001 TO \$20,000, or..... 4  
\$20,001 TO \$25,000..... 5

[P23SET 2]  
\$25,001 TO \$30,000..... 6  
\$30,001 TO \$35,000..... 7  
\$35,001 TO \$40,000..... 8  
\$40,001 TO \$45,000, or..... 9

\$45,001 TO \$50,000.....	10
[P23SET 3]	
\$50,001 TO \$60,000.....	11
\$60,001 TO \$75,000.....	12
\$75,001 TO \$100,000, or.....	13
Over \$100,000.....	14

### **CLOSING STATEMENT**

[IF PANEL ONE, 2002, PANEL TWO, 2004, OR PANEL THREE, 2006, ADMINISTER QUESTIONS S20 THROUGH S29<sub>ov</sub> FROM SCREENER HERE.]

Those are all the questions I have for you. Next I'd like to talk to (CHILD 1) and (CHILD2).

THANK4. Those are all the questions I have for you. Thank you very much for your time. Goodbye.

THANK5. (We are only interviewing in households with children age 9 to 13.) (Those are all the questions I have about (CHILD/CHILD1) (but I do want to ask you about (CHILD2)/but I do have a few more questions.) Thank you very much for your time. Goodbye.

**YOUTH MEDIA CAMPAIGN LONGITUDINAL SURVEY  
MASTER QUESTIONNAIRE  
CHILD INTERVIEW**

This questionnaire contains all questions asked in the YMCLS.

1 = Asked in 2002 only	6 = Asked in 2004 only	11 = Asked in 2006 only
2 = Asked in 2003-2006	7 = Asked in 2004-2006	12 = Asked in 2002-2003 & 2006
3 = Asked in 2005-2006	8 = Asked in 2002-2004	13 = Asked in 2004 & 2005
4 = Asked in 2002-2003	9 = Asked in 2003-2004	
5 = Asked in 2003 only	10 = Asked in 2005 only	

**EX\_CHILD**

Hi, my name is (INTERVIEWER) and I'm calling for the Centers for Disease Control and Prevention, the CDC. We are asking kids across the country to volunteer to participate in a study about their after-school and weekend activities. It takes about 10 minutes, and your answers will be kept private. I'll try to make my questions very clear, but if you don't understand something, I'll explain again. If there is something you don't want to answer, just say so, and if you want to stop, just tell me. [We have already talked with your (RELATION) and you can check with (him/her) if you have questions about this.] If you're ready, let's begin.

C1. Do you go to school right now, or are you out of school for (a/summer) break?

- IN SCHOOL..... 1
- NOT IN SCHOOL ..... 2
- OUT FOR SUMMER..... 3
- OUT FOR ANOTHER SCHOOL BREAK..... 4
- OTHER (SPECIFY) ..... 91

If P\_D1=1 and P\_D2=2, go to C\_8A. If P\_D1= 2 and P\_D2 =1, go to C2INTRO.  
Else, go to C2INTRO.

**C2INTRO**

For the next few questions, think about the physical activities you may have done in the past 7 days, things that got your body moving like sports, physical activity lessons, or playing actively with your friends. (Do not include things you did during the school day like PE, gym class, or recess.)

C2. In the past 7 days, since last [DAY], did you do any physical activities (after school or on the weekend)?

- YES ..... 1 (GO TO C3)
- NO ..... 2 (GO TO C6INTRO)

- C3. What physical activities did you do in the past 7 days? (Remember not to include things you did during the school day like PE, gym class, or recess.)  
(PROBE: Did you do any other physical activities in the past 7 days?)

C3AEROBC	AEROBICS/WEIGHT TRAINING/GYM/EXERCISE .....	1
	BASEBALL/SOFTBALL/	
C3BASBL1 & 2	CATCH/PITCHING .....	2
C3BSKTBL	BASKETBALL .....	3
C3BIKRD1 & 2	BIKE RIDING/DIRT BIKING/MOUNTAIN BIKING .....	4
C3CHEERL	CHEERLEADING .....	5
C3DANCE1, 2 & 3	DANCE.....	6
C3FIELDH	FIELD HOCKEY/STREET HOCKEY/ROLLER HOCKEY .....	7
C3FOOTBL	FOOTBALL .....	8
C3GOLF	GOLF .....	9
C3GYMNAS	GYMNASTICS/TUMBLING .....	10
C3HIKING	HIKING.....	11
C3ICEHOC	ICE HOCKEY .....	12
C3ICESKA	ICE SKATING .....	13
C3JUMPRP	JUMPING ROPE .....	14
C3LACROS	LACROSSE.....	15
C3MRART1 & 2	*MARTIAL ARTS (KARATE/TAE KWON DO/JUDO, ETC.) <sup>2</sup> .....	16
C3PLYGM1, 2, & 3	PLAYING GAMES (PROBE WERE YOU PHYSICALLY ACTIVE? IF NO, DON'T COUNT) .....	17
C3ROLLER	ROLLER BLADING/ROLLER SKATING .....	18
C3RUNNIN	RUNNING/JOGGING .....	19
C3SCOOTR	SCOOTER RIDING (PROBE: DOES IT HAVE A MOTOR? IF YES, DON'T COUNT) .....	20
C3SKATEB	SKATEBOARDING .....	21
C3SOCCER	SOCCER.....	22
C3SWIM	SWIMMING .....	23
C3TENNIS	TENNIS.....	24
C3TRACKF	TRACK & FIELD .....	25
C3VOLLEY	VOLLEYBALL .....	26
C3WALK	WALKING.....	27
C3WRESTL	WRESTLING .....	28
	OTHER 4 (SPECIFY).....	91

<sup>2</sup> Asked in 2003-2006.

\* Martial Arts (asked in 2003-2006) replaced Karate (asked in 2002 only) after 2002.

C4. Was [ACTIVITY] with an organized group that has a coach, instructor, or leader, or was this an activity you did in your free time?

AC4Z1-AC4Z5	ORGANIZED .....	1
	FREE TIME .....	2
	BOTH ORGANIZED AND FREE TIME .....	3
	OTHER (SPECIFY) _____	91

C5. [IF BOTH ORGANIZED AND FREE TIME, repeat twice: Thinking of just the days you [VERB ACTIVITY] (as an organized group...../ as an activity in your free time....)]

On how many of the past 7 days, since last [DAY], did you participate in [ACTIVITY] (after school or on the weekend)? [REPEAT FOR 5 ACTIVITIES.]

AC5Z1-AC5Z5 (Organized)	NUMBER OF DAYS _____
AC5A1-AC5A5 (Free time)	(ENTER # 0 to 7)

C5c.<sup>6</sup> [ASK FOR UP TO THREE FREE-TIME ACTIVITIES IN ORDER OF LISTING:] About how long do you spend doing (ACTIVITY) when you do it in your free time?

	NONE .....	1	(GO TO C6INTRO)
AC5CHR1-AC5CHR5 (Hours)	HOURS.....	2	(GO TO C5aOV)
AC5CMN1-AC5CMN5 (Minutes)	MINUTES .....	3	(GO TO C5aOV)
AC5C1-AC5C5 (Time)	HOURS AND MINUTES .....	4	(GO TO C5aOV)

C5aOV.<sup>6</sup> NUMBER OF (HOURS/MINUTES/HOURS AND MINUTES) \_\_\_\_\_

In 2002, children with C3=2 (no physical activity in the past 7 days) were not asked C6 (physical activity yesterday). In 2003, a subsample of respondents with C3=2 were asked C6. As a result of that experiment, in 2004-2006 all children were asked C6, regardless of their response at C2.

C6INTRO

For the next few questions, think about the sports, lessons, or physical activities you may have done yesterday. (Please do not include things you did during the school day like PE, gym class, or recess.)

C6. Did you do any physical activities yesterday, on [DAY]?

YES .....	1	(GO TO C7)
NO .....	2	(GO TO C8 in 2002/ C_7a in 2003-2006)

<sup>6</sup> Asked in 2004 only. (Replaced by C\_7x in 2005.)

C7. What physical activities did you do yesterday? [PROBE: Did you do any other physical activities yesterday?]

C7AEROBC	AEROBICS/WEIGHT TRAINING/GYM/EXERCISE .....	1
C7BASBL1 & 2	BASEBALL/SOFTBALL/ CATCH/PITCHING .....	2
C7BSKTBL	BASKETBALL .....	3
C7BIKRD1 & 2	BIKE RIDING/DIRT BIKING/MOUNTAIN BIKING .....	4
C7CHEERL	CHEERLEADING .....	5
C7DANCE1, 2 & 3	DANCE.....	6
C7FIELDH	FIELD HOCKEY/STREET HOCKEY/ROLLER HOCKEY .....	7
C7FOOTBL	FOOTBALL .....	8
C7GOLF	GOLF .....	9
C7GYMNAS	GYMNASTICS/TUMBLING .....	10
C7HIKING	HIKING.....	11
C7ICEHOC	ICE HOCKEY .....	12
C7ICESKA	ICE SKATING .....	13
C7JUMPRP	JUMPING ROPE .....	14
C7LACROS	LACROSSE.....	15
C7MRART1 & 2	*MARTIAL ARTS (KARATE/TAE KWON DO/JUDO, ETC.) <sup>2</sup> .....	16
C7PLYGM1, 2, & 3	PLAYING GAMES (PROBE WERE YOU PHYSICALLY ACTIVE? IF NO, DON'T COUNT).....	17
C7ROLLER	ROLLER BLADING/ROLLER SKATING .....	18
C7RUNNIN	RUNNING/JOGGING .....	19
C7SCOOTR	SCOOTER RIDING (PROBE: DOES IT HAVE A MOTOR? IF YES, DON'T COUNT) .....	20
C7SKATEB	SKATEBOARDING .....	21
C7SOCCER	SOCCER.....	22
C7SWIM	SWIMMING .....	23
C7TENNIS	TENNIS.....	24
C7TRACKF	TRACK & FIELD .....	25
C7VOLLEY	VOLLEYBALL .....	26
C7WALK	WALKING.....	27
C7WRESTL	WRESTLING .....	28
	OTHER 4 (SPECIFY).....	91

<sup>2</sup> Asked in 2003-2006.

\* Martial Arts (asked in 2003-2006) replaced Karate (asked in 2002 only) after 2002.





C8. For the next questions, think about all the ways kids can be physically active, including all of the sports teams, lessons, and physical activities you do in your free time.

For each statement, please tell me if you really agree, sort-of agree, sort-of disagree, or really disagree. The first one is...

[INSERT STATEMENT.] [READ SCALE FIRST TIME, THEN IF NECESSARY.] [IF AGREE, PROBE THE FIRST TIME:] Would you say you really agree or sort of agree?, IF DISAGREE, PROBE the first time: Would you say you really disagree or sort of disagree?)

(ROTATE START.)

SET 1

	REALLY AGREE	SORT OF AGREE	SORT OF DISAGREE	REALLY DISAGREE
C8b. I'd rather watch TV or play video games than do physical activities. ....	1	2	3	4
C8c. <sup>12</sup> I feel safe doing physical activities outside, in my neighborhood. ....	1	2	3	4
C8d. <sup>4</sup> Usually there is no one around for me to do physical activities with. ....	1	2	3	4
C8e. There are lots of places in my neighborhood where I can do physical activities. ....	1	2	3	4
C8f. My family thinks I should do physical activities. ....	1	2	3	4
C8g. My parents show or tell me they really like it when I do physical activities. ....	1	2	3	4
C8h. If I asked my parents to do physical activities with me, they probably would. ....	1	2	3	4
C8i. I should probably do more physical activities than I do. .	1	2	3	4
C8j. I'm too busy to do more physical activities than I do. ....	1	2	3	4
C8k. I could find a physical activity to do that I enjoy. ....	1	2	3	4
C8l. <sup>5</sup> My parents don't really care if I do physical activities or not. ....	1	2	3	4
C8m. <sup>2</sup> My friends think that doing physical activities is fun. ....	1	2	3	4
C8n. <sup>2</sup> Kids my age think that doing physical activities is fun. ....	1	2	3	4
C8o. <sup>2</sup> My friends think that doing physical activities is important. ....	1	2	3	4
C8p. <sup>2</sup> Kids my age think that doing physical activities is important. ....	1	2	3	4
C8q. <sup>5</sup> My friends think that going to after-school programs is fun. ....	1	2	3	4
C8r. <sup>7</sup> I'm not athletic enough to do physical activities. ....	1	2	3	4
C8s. <sup>7</sup> There aren't many ways to play sports without signing up or being on a team. ....	1	2	3	4

<sup>12</sup> Asked in 2002-2003 and 2006.

<sup>4</sup> Asked in 2002-2003.

<sup>5</sup> Asked in 2003 only.

<sup>2</sup> Asked in 2003-2006.

<sup>7</sup> Asked in 2004-2006.

[ROTATE START.]

SET 2

Now tell me if you really agree, sort-of-agree, sort-of-disagree, or really disagree with these statements. If I did physical activities on most days...

	REALLY AGREE	SORT OF AGREE	SORT OF DISAGREE	REALLY DISAGREE
C8aa. <sup>1</sup> It would keep me in shape .....	1	2	3	4
C8ab. It would be boring .....	1	2	3	4
C8ac. It would be fun .....	1	2	3	4
C8ad. It would help me make new friends.....	1	2	3	4
C8ae. It would help me spend more time with my friends ....	1	2	3	4
C8af. It would make me feel good about myself.....	1	2	3	4
C8ag. It would keep me from doing other things I like better	1	2	3	4

Now just a few more questions. Do you really agree, sort-of agree, sort-of disagree, or really disagree with these statements...

[ROTATE START.]

	REALLY AGREE	SORT OF AGREE	SORT OF DISAGREE	REALLY DISAGREE
C8a. I think I can ask my parents to sign me up for a sport or other physical activity. ....	1	2	3	4
C8ah. <sup>7</sup> I think I can be physically active no matter how busy my day is. ....	1	2	3	4
C8ai. <sup>7</sup> I think I can be physically active no matter how tired I may feel .....	1	2	3	4
C8aj. <sup>7</sup> I think I can be physically active even if it is hot or cold outside .....	1	2	3	4
C8ak. <sup>7</sup> I think I have what it takes to be physically active. ....	1	2	3	4

C\_8x.<sup>7</sup> Please tell me which sentence best describes you. During my free time on most days...

- I am sure I **will not** be physically active ..... 1
- I **probably will not** be physically active. .... 2
- I **may or may not** be physically active ..... 3
- I **probably will** be physically active..... 4
- I am sure I **will** be physically active ..... 5

---

<sup>1</sup> Asked in 2002 only.

<sup>7</sup> Asked in 2004-2006

<sup>7</sup> Asked in 2004-2006.

C\_8a.<sup>2</sup> Now I'd like you to describe yourself. Please tell me if each word describes you never, sometimes, or usually. Are you...

	NEVER	SOMETIMES	USUALLY
C_8aa. Confident? .....	1	2	3
C_8ab. Lazy? .....	1	2	3
C_8ac. A leader? .....	1	2	3
C_8ad. Athletic (physically)? .....	1	2	3
C_8ae. Popular? .....	1	2	3

C\_8b.<sup>2</sup> Here are just a few other statements. Please tell me if you really agree, sort-of-agree, sort-of-disagree, or really disagree with these statements.

	REALLY AGREE	SORT OF AGREE	SORT OF DISAGREE	REALLY DISAGREE
C_8ba. In general, I am satisfied with myself. ....	1	2	3	4
C_8bb. My parents don't really understand me. ....	1	2	3	4
C_8bc. My parents have rules about how much time I can spend watching TV .....	1	2	3	4
C_8bd. My parents have rules about how much time I can spend playing video games .....	1	2	3	4

If P\_D1= 1 and P\_D2 =1, go to C13 in 2002-2003/ C14 in 2004-2006. Else, if P\_D1= 2 and P\_D2 =1 go to C\_8c. Else if P\_D1= 1 and P\_D2 =2, go to C10INTRO.

C\_8c.<sup>2</sup> How many kids your age do physical activities every day? Would you say...

- None, ..... 1
- Some, ..... 2
- Most, or ..... 3
- All?..... 4

C\_8d.<sup>2</sup> How many of your friends do physical activities every day? Would you say...

- None, ..... 1
- Some, ..... 2
- Most, or ..... 3
- All?..... 4

C9.<sup>8</sup> Now just answer yes or no. In the last two months, have you tried a new physical activity that you've never done before?

- YES ..... 1
- NO ..... 2

<sup>2</sup> Asked in 2003-2006.

<sup>8</sup> Asked in 2002-2004.

If P\_D1=1 and P\_D2=2, go to C10INTRO. If P\_D1= 2 and P\_D2 =1, go to C13 in 2002-2003/ C14 in 2004-2006.

C10INTRO

These next questions are about some other activities that kids might be involved in.

C10. Do you currently belong to community groups like Scouts or 4-H, or any other special groups for kids?

YES ..... 1  
NO ..... 2

C11. Do you currently belong to school groups like band, drama club, newspaper, or student government?

YES ..... 1  
NO ..... 2

C12.<sup>1</sup> For the next questions, think about all of the group activities kids can be involved in like clubs, organizations, or other activities, and tell me if you really agree, sort-of agree, sort-of disagree, or really disagree with each of the following statements.

[READ SCALE FIRST TIME; REPEAT IF NECESSARY.] (IF AGREE, PROBE THE FIRST TIME: Would you say you really agree or sort of agree? IF DISAGREE, PROBE the first time: Would you say you really disagree or sort of disagree?)

	REALLY AGREE	SORT OF AGREE	SORT OF DISAGREE	REALLY DISAGREE
C12a. If I wanted to, I could find a new club or group activity to join.....	1	2	3	4
C12b. I'd rather watch TV or play video games than be involved in clubs or group activities .....	1	2	3	4
C12c. My parents show or tell me they really like it when I do clubs or group activities .....	1	2	3	4

C13.<sup>4</sup> Now just answer yes or no. Do you currently take any music lessons like piano or guitar?

YES ..... 1  
NO ..... 2

<sup>1</sup> Asked in 2002 only.

<sup>4</sup> Asked in 2002-2003.

C14. How many hours did you watch TV, play video games, or play computer games yesterday, [DAY]? [DOES NOT INCLUDE HOMEWORK ON COMPUTER.]

- NONE ..... 1 (GO TO C15)
- HOURS..... 2 (GO TO C14OV)
- MINUTES ..... 3 (GO TO C14OV)
- HOURS AND MINUTES ..... 4 (GO TO C14OV)

C14OV. NUMBER OF (HOURS/MINUTES/HOURS AND MINUTES)\_\_\_\_\_

C15.<sup>1</sup> There are many ads on television, radio, and in newspapers and magazines with slogans you may or may not remember. Have you heard, read, or seen any ads with the slogan...[KEEP A FIRST; ROTATE B THROUGH E.]

	YES	NO
a. "Just do it"?	1	2
b. How about "My Anti-Drug"?	1	2
c. How about "Stay True to the Game"?	1	2
d. How about "Verb "?	1	2
e. How about "Akimbo "?	1	2

C16.<sup>1</sup> Have you been to any special events lately that encourage kids to do physical activities?

- YES ..... 1
- NO ..... 2

C\_15INTRO.

Next we have a few questions about messages and advertising you may have seen.

C\_15.<sup>2</sup> Have you seen, read, or heard about any messages or advertising for kids getting active?

- YES ..... 1 (GO TO C\_16)
- NO ..... 2 (GO TO C\_17)

C\_16.<sup>2</sup> What is the name of the message or advertising?

[PROBE: If something other than VERB mentioned: Are there any others for kids getting active?  
 PROBE AGAIN If second message other than VERB mentioned: Are there any others?]

- VERB..... 1 (GO TO C\_18)
- OTHER (SPECIFY) ..... 91 (GO TO C\_17)

C\_17.<sup>2</sup> Have you seen, read, or heard any messages or advertising about VERB?

- YES ..... 1
- NO ..... 2

---

<sup>1</sup> Asked in 2002 only.

<sup>2</sup> Asked in 2003-2006.

If C\_16 = 1 or C\_17=1, go to C\_18. Else, go to C\_24.

C\_18.<sup>2</sup> Please tell me in your own words what VERB is all about? [PROBE ONE OR TWO TIMES WITH ANY OF THE FOLLOWING: Can you tell me more? What does that mean? Anything else?]

---

---

C\_19.<sup>2</sup> What ideas did VERB give you? [PROBE ONE OR TWO TIMES WITH ANY OF THE FOLLOWING: Can you tell me more? What does that mean? Anything else?]

---

---

**THE FOLLOWING CATEGORIES WERE USED TO CODE ALL OPEN-ENDED TEXT RESPONSES TO C\_18 AND C\_19:**

- 1 It's what you do/ what I/you can do/what kids do (slogan and its variation)
- 2 Action word
- \*3 Action; get/be/stay active; get moving; get up and out
- \*4 Sports; playing/involved in sports; athletic; specific sports mentioned (e.g. baseball football soccer etc.)
- \*5 General physical activities; running around; specific activities; playing outside
- \*6 Have fun; enjoy; be entertained; what you like to do (in general)
- \*7 Staying healthy/fit; getting in shape; exercising; endurance
- 8 Language/vocabulary/parts of speech; "action word" (when clearly referring to a part of speech)
- \*9 Avoid sitting around/watching TV/getting fat; avoid getting a disease
- 10 Avoid drugs; stay out of trouble
- \*11 Be/believe yourself; have confidence; believe you can do it; work/try hard/harder; put effort into it
- \*12 Start/try/do new things; try different things (e.g. sports activities)
- 13 Inactive pursuits mentioned (hobbies drawing video games cooking talking etc)
- 14 Clubs; service activities; working to help others at school or in the community
- 15 Shows verbs; people/kids with verbs/words/colors; kids with verbs playing sports/find your verb choose your verb pick your verb (describes the ad or repeats ad copy)
- \*16 Play (not associated with specific activities)
- 17 Other (specify)
- 97 Refused (used only when refusal is clear)
- 98 Don't know (nothing/none/no ideas/not sure/no)
- 99 Not ascertain

---

<sup>2</sup> Asked in 2003-2006.

\* Denotes an understanding of VERB messages.

C\_20.<sup>2</sup> Please tell me all of the places you have seen, read or heard about VERB.  
 [DO NOT READ LIST. CIRCLE ALL RESPONSES.] [RECORD ANSWERS BELOW UNDER C\_20.]

C\_21.<sup>2</sup> Have you seen, read or heard about VERB ... [READ ITEMS NOT CIRCLED UNDER C\_20.  
 RECORD ALL THAT APPLY UNDER C\_21.]

		C_20	C_21	
			YES	NO
C_20TV	On television .....	1	C_21A 1	2
C_20RAD	On the radio .....	2	C_21B 1	2
C_20INT	On the internet .....	3	C_21C 1	2
C_20FRND	From friends .....	4	C_21D 1	2
C_20PRNT	From your parents .....	5	C_21E 1	2
C_20SCHL	At school/(from teacher) <sup>3</sup> .....	6	C_21F 1	2
C_20MAGA	In magazines .....	7	C_21G 1	2
C_20MOVI <sup>9</sup>	Before a movie .....	7	C_21H <sup>9</sup> 1	2
C_20SIGN	On signs, billboards, or posters .....	8	C_21I 1	2
C_20EVNT <sup>3</sup>	At an event, concert, or the mall .....	9	C_21K <sup>3</sup> 1	2
C_20CLUB <sup>3</sup>	At an after-school or community club, like the YMCA .....	10	C_21L <sup>3</sup> 1	2
C_20NEWS <sup>3</sup>	In newspapers .....	11	C_21M <sup>11</sup> 1	2
C_20OTHR	Anywhere else? Other: Specify _____	91	C_21J 1	2

If C\_20=1 or C\_21a = 1, and/or if C\_20=2 or C\_21b = 1, ask C\_21x. Else go to C\_24 in 2003/ C\_25INTRO in 2004-2006.

C\_21x.<sup>7</sup> How often do you usually (see/hear/see or hear) a message or advertisement about VERB on (television/radio/television or the radio). Would you say...

- About every day,..... 1
- Several times a week,..... 2
- About once a week, or..... 3
- Less than once a week..... 4

If C\_20=6 or C\_21f = 1, ask C\_21y. Else go to C\_24 in 2003/ C\_25INTRO in 2004-2006.

<sup>2</sup> Asked in 2003-2006.

<sup>3</sup> Asked in 2005-2006.

<sup>9</sup> Asked in 2003-2004.

<sup>11</sup> Asked in 2006 only.

<sup>7</sup> Asked in 2004-2006.

C\_21y.<sup>7</sup> Please tell me all the places during the school day you have seen, heard, or read about VERB. [Code all that apply.]

C_21YCH1	CHANNEL 1 .....	1
C_21YTV	TV OR VIDEOS .....	2
C_21YPOS	POSTERS OR SIGNS .....	3
C_21YSUP	SCHOOL SUPPLIES, E.G., BOOK COVERS OR PENCILS .....	4
C_21YFRD	FROM FRIENDS AT SCHOOL.....	5
C_21YTCH	FROM TEACHERS.....	6
C_21YACT	ANYTIME ACTION CARDS.....	7
C_21YFLY	EVENT FLYERS.....	8
C_21YGYM	GYM, PE RECESS, HEALTH CLASS .....	9
C_21YSCHO	SCHOLASTIC, MAGAZINES, NEWSLETTERS.....	10
C_21YCLS	CLASS OR CLASSROOM, LIBRARY .....	11
C_21YCMP	COMPUTER, INTERNET .....	12
C_21YENG	ENGLISH OR LANGUAGE ARTS CLASS .....	13
C_21YOTH	OTHER ( <i>specify</i> ) .....	91

C\_24.<sup>5</sup> Have you ever attended a Nickelodeon “Wild and Crazy Kids Show” in your area?

YES .....	1
NO .....	2

C25INTRO Now I'd like to ask you about some television commercials that you may or may not have seen. [ROTATE START.]

	YES	NO
C_25a. <sup>6</sup> Have you seen a commercial on television that shows girls and boys on the beach hitting a ball over a net, and then the ball hits a lifeguard in the head? .....	1	2
C_25b. <sup>6</sup> How about a television commercial that shows a boy playing horse on a basketball court who dribbles the ball between his legs and bounces it off a wall but misses his shot? .....	1	2
C_25c. <sup>6</sup> Have you seen a commercial on television that shows a group of kids racing their bikes through a neighborhood while an announcer describes the race? .....	1	2
C_25d. <sup>6</sup> How about a commercial on television that shows a group of boys and girls who keep playing their football game even after it begins to rain and they get all wet?.....	1	2
C_25e. <sup>13</sup> Have you seen a commercial on television that shows tennis star Venus Williams playing with some kids who made up their own rules? .....	1	2

<sup>7</sup> Asked in 2004-2006.

<sup>5</sup> Asked in 2003 only.

<sup>6</sup> Asked in 2004 only.

<sup>13</sup> Asked in 2004 and 2005.



C_25g. <sup>10</sup>	Have you seen a commercial on television that shows football player Donovan McNabb playing touch football in a park with kids who are calling plays that Donovan never heard of? .....	1	2
C_25h. <sup>10</sup>	How about a commercial on television that shows basketball player Allen Iverson taking turns jumping on a trampoline with a bunch of kids? .....	1	2
C_25i. <sup>10</sup>	Have you seen a commercial on television that shows soccer star Landon Donovan playing with some kids who made up their own rules? .....	1	2
C_25j. <sup>10</sup>	How about a television commercial where Ned Bigby, the star of a show on Nickelodeon, shows different ways to have fun while skating on the ice? .....	1	2
C_25k. <sup>10</sup>	Have you seen a commercial where a boy is relaxing on the couch and he gets a call on his cell phone? When he reaches to answer it, the phone slides off the table and goes out through a doggy door? .....	1	2
C_25l. <sup>11</sup>	Have you seen a commercial on television that shows kids who are playing with a <u>glowing</u> ball that looks like the sun and playing different games with it? .....	1	2
C_25m. <sup>11</sup>	Have you seen a commercial on television that shows Emma Roberts, the star of the show "Unfabulous" on Nickelodeon, playing volleyball with her friends on the beach? .....	1	2
C_25n. <sup>11</sup>	Have you seen a commercial on television that shows characters from the show "Juniper Lee" playing games with a ball that looks like the sun (or playing with the sun)? .....	1	2
C_25o. <sup>11</sup>	Have you seen or heard anything about kids playing with and passing around yellow balls? .....	1	2

C\_26.<sup>7</sup> Would you describe yourself as...

- Very underweight,..... 1
- Slightly underweight,..... 2
- About the right weight,..... 3
- Slightly overweight, or..... 4
- Very overweight..... 5

C\_22.<sup>2</sup> Have you ever visited the website *[that had information about VERB]* for "VERB"?

- YES ..... 1
- NO ..... 2

<sup>10</sup> Asked in 2005 only.

<sup>10</sup> Asked in 2005 only

<sup>11</sup> Asked in 2006 only.

<sup>7</sup> Asked in 2004-2006.

<sup>2</sup> Asked in 2003-2006. (Item moved to the end of the questionnaire in 2004-2005 and changed to include "a website that had information about VERB.")

<p>If C_20=1 or C_21a = 1, and/or if C_20=2 or C_21b = 1, ask C_27. Else go to C_23INTRO.</p>
---

C\_27.<sup>11</sup> Thinking of the last time you (saw/heard/saw or heard) a message or advertisement for VERB, would you say it was...

- In the past week..... 1
- More than 1 week, but less than 1 month ago..... 2
- 1 to 2 months ago..... 3
- More than 2 months ago, but within the past year..... 4
- More than 1 year ago..... 5

C\_23INTRO. Now, one last question—you've really done a good job...

C\_23.<sup>12</sup> Have you attended any events, concerts or shows that had messages or advertising about VERB?

- YES ..... 1 (GO TO C\_23ov)
- NO ..... 2 (GO TO EXCTHNK)

C\_23ov.<sup>11</sup> Was this in the past year?

- YES ..... 1
- NO ..... 2

**CLOSING STATEMENT**

EXTDTHNK

Those are all of the questions I have for you. You've helped us understand more about what kids like you think about the activities they do.

(If more than one child) May I speak with [CHILD2] now?

THANK2. Thank you so much for answering these questions. We are only interviewing in households with children age 9 to 13.

THANK4. Those are all the questions I have for you. Thank you very much for your time.

---

<sup>11</sup> Asked in 2006 only.

<sup>12</sup> Asked in 2002-2003 and 2006.