

Contract Nos.: 43-3198-4-0060; AG-3198-D-05-0071
Contract Amounts: \$3,533,831; \$1,458,038
MPR Reference Nos.: 6096; 6196

MATHEMATICA
Policy Research, Inc.

**School Nutrition Dietary
Assessment Study-III:
Volume III: Sampling
and Data Collection**

Final Report

November 2007

*Anne Gordon
John Hall
Eric Zeidman
Mary Kay Crepinsek
Melissa Clark
Elizabeth Condon*

Submitted to:

Office of Research, Nutrition, and Analysis
USDA Food and Nutrition Service
3101 Park Center Drive
Alexandria, VA 22302

Project Officer:
Patricia McKinney

Submitted by:

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director:
Anne Gordon

CONTENTS

Chapter	Page
I INTRODUCTION	1
II SAMPLE DESIGN AND SELECTION	3
A. SELECTING PSUs	4
B. SELECTING SCHOOLS AND STUDENTS	7
III DATA COLLECTION	11
A. RECRUITING DISTRICTS AND SCHOOLS	11
B. DATA COLLECTION	14
1. SFA-Level Data: Initial Contact Survey and SFA Director Survey	15
2. School-Level Surveys (Excluding Menu Survey)	16
3. Menu Survey	18
4. Student and Parent Interviews	20
5. Dietary Recalls	25
6. Height and Weight Measurement	29
C. TRAINING OF DATA COLLECTORS	30
D. INCENTIVES FOR SURVEY RESPONDENTS	32
E. SURVEY RESULTS	32
IV PROCESSING OF MENU AND DIETARY RECALL DATA	37
A. STAFFING AND TRAINING	38
B. PROCESSING OF THE MENU SURVEY DATA	39
1. Review and Editing of Menu Data	39
2. Entering Data into Survey Net	44
3. Recipes.	45
4. Self-Serve Food Bars.....	47
5. Pre-Prepared School Foods	48
6. Quality Control Procedures	50

CONTENTS *(continued)*

Chapter	Page
IV <i>(continued)</i>	
C. PROCESSING OF THE DIETARY RECALL DATA	52
1. Coding Foods and Amounts	52
2. Quality Control Procedures	58
3. Linking of Menu and Recall Data	60
4. Final Data Files.....	62
V WEIGHTING SNDA-III DATA.....	63
A. SFA-LEVEL WEIGHTS	63
B. SCHOOL-LEVEL WEIGHTS.....	67
C. STUDENT-LEVEL WEIGHTS	70
D. PARENT WEIGHTS	72
REFERENCES.....	75
APPENDIX A: TECHNICAL DOCUMENTATION OF SAMPLING AND WEIGHTING OF THE NSLP SAMPLE FRAME/ PRELIMINARY SURVEY	A.1

TABLES

Table		Page
II.1	DISTRIBUTION OF INITIAL SFA SAMPLE.....	6
II.2	COMPLETED SAMPLE SIZES	9
III.1	SNDA-III DATABASE: INSTRUMENTS AND SAMPLE SIZES.....	12
III.2	SNDA-III RESPONSE RATES.....	34
IV.1	RECIPE MODIFICATION GUIDELINES FOR SNDA-III MENU SURVEYS.....	54
V.1	VALUES OF $SWF2B$ FOR NONCERTAINTY SFAs	66
V.2	SFA NONRESPONSE ADJUSTMENT FACTOR	67
V.3	SCHOOL-LEVEL WEIGHTS FORMULAS FOR $W2a_{ijk}$, $W2b_{ijk}$, AND $w2_{ijk}$	69

FIGURES

Figure	Page
III.1	TECHNICAL ASSISTANCE SCHEDULE FOR MENU SURVEY 21

I. INTRODUCTION

The U.S. Department of Agriculture (USDA) sponsors the school nutrition programs—the National School Lunch Program (NSLP) and the School Breakfast Program (SBP)—to promote children’s health and well-being through providing nutritious meals to children in school. As part of its ongoing assessment of program performance, the Food and Nutrition Service (FNS) of USDA sponsored the third School Nutrition Dietary Assessment study (SNDA-III) to provide up-to-date information on the school meal programs, the school environments that affect the food programs, the nutrient content of school meals, and the contributions of school meals to student’s diets. Mathematica Policy Research, Inc. (MPR) was awarded contracts by FNS to collect the study data and to analyze the data and produce reports.

This report is the third of three volumes; it describes the data collection and sampling procedures for SNDA-III. Chapter II describes the multistage sample design for the study and. Chapter III describes the data collection procedures. Chapter IV describes the procedures for food and nutrient coding of the information on school menus and on students’ diets. Finally, Chapter V describes the construction of sample weights necessary to obtain nationally representative estimates from the study sample. This report provides context for the SNDA-III data analysis results, which are presented in two companion reports:

School Nutrition Dietary Assessment Study-III: Volume I: School Foodservice, School Food Environment, and Meals Offered and Served.

School Nutrition Dietary Assessment Study-III: Volume II: Student Participation and Dietary Intakes.

II. SAMPLE DESIGN AND SELECTION

This chapter describes the sample design for the third School Nutrition Dietary Assessment study (SNDA-III). The study collected data from public school food authorities (SFAs; school districts or occasionally small groups of districts), schools, and students and their parents. The SFAs, schools, and students in the study were sampled using a multistage design. Parents were not sampled separately but were interviewed because their children had been sampled. Substantive data for the study were obtained from the entities at each of these levels of sampling.

The primary sampling unit (PSU) in the multistage design was the SFA. In the first step of sampling, 270 PSU equivalents (some PSUs selected with certainty were large enough to count as 2 PSU equivalents) were subselected from the NSLP Sample Frame, a sample of 2,500 public SFAs in school year 2003–2004 that had been selected by Mathematica Policy Research, Inc. (MPR) under a separate contract with the Food and Nutrition Service (FNS).¹ These SFAs had been screened to determine eligibility and to obtain information about their participation in the National School Lunch Program (NSLP) and School Breakfast Program (SBP). This initial sample of 270 sampled PSU equivalents was divided randomly into 135 main selections and 135 alternate selections. If a main selection chose not to participate in the study, the alternate selection was contacted.

¹ Under that contract, the National School Lunch Program Sample Frame Construction Project, MPR selected a sample of approximately 2,500 public school districts from the Common Core of Data (CCD), a comprehensive database maintained by the U.S. Department of Education on school districts and schools. These districts were sampled with probabilities proportional to size (PPS), with the measure of size (MOS) being the square root of enrollment. The sampled districts were then screened for their SFA status (with the help of state child nutrition agencies) in order to compile a sample frame of SFAs. In more than 90 percent of instances, the school district and the SFA were the same. However, in less than 10 percent of instances they were not, because the same SFA served several districts, because the district did not participate in the NSLP, or for other reasons. See Appendix A for technical documentation for the sampling and weighting procedures used in developing the frame.

The sample has two overall components. The first component—the “basic” sample—includes each of the three levels of sampling summarized above. The planned sample sizes for this part of the sample were 100 SFAs and three schools from each SFA, for a total of 300 schools. An average of slightly more than eight children per school were to be sampled, for a total of 2,420 students and their parents. When possible, the three schools selected from each SFA were stratified to include one elementary school, one middle school, and one high school.²

The second component—the “supplemental” sample—was designed to increase the available sample sizes for analyses of the SFAs and schools, without incurring the extensive costs associated with the fieldwork needed for the in-person student and parent interviews. This supplemental sample included 35 additional SFAs and 105 additional schools served by them, but students within these schools were not sampled. The alternate sample selections were similarly designated as part of the basic or supplemental sample.

A. SELECTING PSUs

The NSLP Sample Frame served as the frame for selecting the basic and supplemental samples of SFAs for SNDA-III. The NSLP Sample Frame was itself a sample of public SFAs selected using the square root of the number of enrolled students as the measure of size (MOS), in order to balance the precision of estimates for SFAs and for students. However, this MOS was not optimal for SNDA-III, since SFA-level estimates were not a main focus of the study, while student-level estimates were. Therefore, the probabilities of selection of SFAs from the sample frame into the SNDA-III sample were set such that when schools were selected within SFAs using probability proportional-to-size (PPS) selection (with the MOS being total school

² As discussed in Section B, all three kinds of schools were not always available in an SFA; indeed, in some cases, an SFA served fewer than three schools. Minor modifications in the sampling plans were made to accommodate these special circumstances.

enrollment), and students were selected with equal probability within schools, the overall probabilities of selection of students would be approximately equal across SFAs. Effectively, SFAs were selected so that the sample of SFAs had the properties of a PPS sample with the MOS being the number of students enrolled in schools served by the SFA, rather than the square root of enrollment.

In selecting the SFAs, the sample was explicitly stratified by whether SFAs were large enough to be selected with certainty. The stratum of noncertainty SFAs was implicitly stratified (sorted) on FNS region, enrollment, and method of menu planning (food-based or nutrient standard menu planning), as the menu-planning method was an important analysis variable.

Two of the SFAs were so large that they were allocated double the number of schools (and students) to be selected into the sample. Because these two received double allocations, they are counted as the equivalent of two SFAs each. Because of the “double hits,” the initial sample, before division into main and alternate subsamples, contained 268 unique SFAs and 270 PSU equivalents; after division, the main sample included 133 unique SFAs but 135 PSU-equivalents.

Table II.1 presents the distribution of the SNDA-III sample of SFAs. Eighteen SFAs (20 PSU equivalents) were large enough to be selected with certainty into the initial sample (which included both the main and alternate samples). In turn, six of these SFAs were large enough to be certainty selections for the main sample.³ Of the 12 SFAs large enough to be selected with certainty for the initial sample but not the main sample, 2 were selected with PPS for the main sample, and the other 10 were assigned to the alternate sample. Pairs were formed

³ Because the main sample is only half as large as the initial sample, the threshold MOS for certainty selection is larger; hence, only 6 of the 18 were retained with certainty for the main sample.

TABLE II.1
DISTRIBUTION OF INITIAL SFA SAMPLE
(PSU Equivalents in Parentheses)

	Initial Sample	Main Sample		Alternate Sample
		Basic	Supplemental	
A. Initial Certainty	18 (20)			
1. Main Certainty		4 (6)	2 (2)	0 (0)
2. Others		2 (2)	0 (0)	10 (10)
B. Initial Other	250 (250)	92 (92)	33 (33)	125 (125)
Total	268 (270)	98 (100)	35 (35)	135 (135)

for the 250 initial noncertainty selections, and one of each pair was randomly assigned to be part of the main sample and one to be part of the alternate sample.⁴

Because some SFAs did not want to participate in the study, 35 of the alternate selections were released, of which 28 participated in the study. In some cases, both SFAs in a pair declined to participate. To ensure that the final main sample contained close to 135 PSU equivalents, an additional pool of replacement SFAs was formed, and a random sample was chosen from that pool to serve as potential replacements if either a certainty SFA or both SFAs in a noncertainty pair declined to participate. The sample from this pool is referred to here as “wildcard” SFAs.

After selecting the 133 sample SFAs, a subsample of 100 SFA equivalents (98 unique SFAs) was selected for the main data collection. The rest (35 SFAs) constituted the “supplemental

⁴ The noncertainty SFAs were selected using SAS PROC SURVEYSELECT and Chromy’s method. SFAs were sorted by the implicit stratification variables, and then the program formed “zones” in the sorted frame and selected one SFA per zone. Pairs were formed using adjacent zones, since SFAs selected from adjacent zones would be similar on the characteristics used to sort the frame. One member of each pair was then assigned to the initial sample and the other member was used as a replacement, if needed.

sample,” in which no on-site data collection occurred. In the end, 94 SFAs from the main sample agreed to participate, and 36 SFAs were recruited for the supplemental sample.

B. SELECTING SCHOOLS AND STUDENTS

In each of the SFAs that was selected for the study, an initial sample of up to six schools was selected. Schools were first divided into elementary, middle, and high school strata, and two schools were selected from each, if possible. If there were fewer than two schools in each of these strata, sampling procedures were modified as follows:

- If all three strata were represented and one of the strata (say, high schools) had only one school, then the school in the stratum with only one was selected, and two schools in each of the other strata were selected.
- If two strata were represented (for example, an SFA with 4 elementary schools and 1 high school, but no middle schools), the sample of schools was allocated so that if a stratum (elementary) had three or more schools, then three would be selected; if a stratum had fewer than three schools (high schools), all would be selected. (In the example, three of the four elementary schools and the high school would be selected.)
- If only one stratum was represented (for example, an elementary-only SFA) the entire sample (six if there were six or more schools, or all of the schools if there were fewer than six), was selected from that stratum.

Schools were selected within SFAs with PPS (size was measured by total school enrollment); after the six (or fewer) schools were selected, three schools (unless there were fewer than three selected) were randomly assigned to the main sample, and the rest were assigned to the alternate sample. If two schools had been selected in each stratum, one in each stratum was assigned to the main sample and the other to the alternate sample. The alternate schools were used if the main selection would not participate or had closed or become ineligible.⁵

⁵ To be eligible, schools had to have at least one grade from 1 to 12 and participate in the NSLP or SBP. (One school in the sample appears to have only offered the SBP.) In addition, schools operated for institutionalized populations (prisoners, patients) were excluded.

The sampling frame for selecting schools within SFAs was the U.S. Department of Education Common Core of Data (CCD) file of public schools. Schools that were apparently ineligible (reported that they served no students, only pre-kindergarten and kindergarten students, or were closed) were not given a chance of selection. Because the CCD was two years old at the time of the survey, SFAs were asked if they had schools that had come into existence since that time. If they reported such schools, these were checked against the CCD to see if they were already on the CCD, and if not, were given a chance to be selected. In the end, only schools in SFAs that agreed to participate were recruited for the study.

Within each sampled school that agreed to participate, students were randomly selected as eligible for completing the dietary recalls; both students and one of their parents were interviewed, if possible, and a subsample of students completed another dietary recall interview about a week later, to help in understanding students' usual intakes. Although the goal was roughly eight student interviews per school, larger numbers were selected to allow for failure to obtain consent from parents for the student interview (in districts where active consent was required) and for parent nonresponse to the parent interview. Only students in grades 1–12 were selected for the study.⁶

The selection of students took place in three steps: first, a sample of students was selected from a list provided by the school; second, this sample was randomly ordered and an initial subsample was released. A final stage in student sampling took place on the day of data collection. Students on the list were called from their classrooms for the interview in the order in which they were listed. If the student was absent or otherwise unavailable, the next student on

⁶ Students in kindergarten, pre-kindergarten, or in self-contained special education classes were considered ineligible. In addition, students who were absent on the day covered by the recall were considered ineligible, as the absence was unlikely to be related to the study.

the list was contacted, until the desired number of interviews had been completed. Any students remaining on the list were not considered part of the released sample.

A parent interview was attempted for all cases in which the student was interviewed. However, some parents did not respond even though their child did.

As discussed further in Chapter III, there was some nonresponse to the interviews at each level. Table II.2 shows the final sample sizes at each level, for the main sample, the supplemental sample (where relevant), and the total sample.

TABLE II.2
COMPLETED SAMPLE SIZES

	Basic Sample	Supplemental Sample	Total Sample
SFAs	94	36 (schools with menu survey) 35 (completed SFA interview)	130 (schools with menu survey) 129 (completed SFA interview)
Schools	287	111	398
Students and Parents	2,718 Dietary Recall 2,330 Parent Interview 2,314 Both Recall and Parent Interview	NA	2,314 Both Recall and Parent Interview

III. DATA COLLECTION

Data collection for the third School Nutrition Dietary Assessment Study (SNDA-III) took place primarily between January and August 2005, under a previous contract between the Food and Nutrition Service (FNS) and Mathematica Policy Research, Inc. (MPR). The study team also collected some preliminary information during the recruiting process, which began in October 2004. In total, data were collected from 130 School Food Authorities (SFAs), 398 schools, 2,718 students, and 2,330 of their parents (Table III.1). This chapter first discusses recruitment of districts and schools into the study sample. It next provides an overview of each of the data collection instruments. It then discusses training of data collectors, incentives offered to survey respondents, and survey response rates.

A. RECRUITING DISTRICTS AND SCHOOLS

Recruiting began with an initial sample of 98 SFAs selected for on-site student-level data collection, and an additional 35 SFAs selected to respond to the menu survey and other SFA- and school-level instruments only. As a first step in the recruiting process, the FNS Project Officer sent the State Child Nutrition Directors a letter describing the study and requesting their support and assistance. Recruiters then contacted the State Child Nutrition Directors and told them about the study, provided them with a list of the school districts in their state selected for participation, and asked them to forward a letter of support to SFA Directors in those school districts.

MPR recruiters began to contact SFA Directors by telephone in October 2004. Initial calls discussed the background and purpose of the study, as well as methods for student sampling and the scheduling of data collection. The recruiters also obtained information on the district's policy on research participation, district characteristics, any recent changes in district

TABLE III.1
SNDA-III DATABASE: INSTRUMENTS AND SAMPLE SIZES

Instrument	Respondents	Completed Sample Size	Mode
Initial Contact Survey Part I	SFA director or designee	130	Telephone interview prior to visit or data collection (mailed upon request).
Initial Contact Survey Part II	School staff in visited schools	280	Telephone interview prior to visit (visited schools only)
Survey of SFA Directors	SFA director	129	Telephone interview after visit or data collection (mailed upon request).
Menu Survey	School food service manager	398	Mail with intensive telephone training, technical assistance, and followup; in-person followup in 287 visited schools.
1. Daily Meal Counts Form		(any meal)	
2. Reimbursable Foods Form: Breakfast		397	
3. Reimbursable Foods Form: Lunch		(lunch)	
4. Recipe Form		331	
5. Self-Serve/Made-to-Order Bar Form		(breakfast)	
Point-of-Sale Form	n.a.	287	Completed by interviewer during visit to 287 schools
School Food Service Manager Survey	School food service manager	398	Telephone (mailed upon request) in 111 schools; in-person interview in 287 visited schools
Principal Survey	Principal	395	Telephone (mailed upon request) in 108 schools; in-person interview in 287 visited schools
Alternative Food Source Checklist	n.a.	284	Completed by interviewer during visit to 287 schools
A La Carte Checklist	n.a.	287	Completed by interviewer during visit to 287 schools
Vending Machine Checklist	n.a.	282	Completed by interviewer during visit to 287 schools
Student Dietary Recall and Interview	Day 1 Recall (plus parent-assisted recall for elementary school students)	2,718	In-person interview
	Day 2 Recall (plus parent-assisted recall for elementary school students)	778	
	Student Interview	2,709	
Parent Interviews	Parent	2,330	In-person interview for parent of elementary student/ telephone interview for parent of secondary student
Weight and Standing Height Measurement	Student	2,689	In-person observation

Note: Completed sample sizes for specific instruments are usually smaller than the analysis sample described in Chapter II, as the analysis files at each level allow for some nonresponse to particular instruments.

n.a. = not applicable.

configuration that were not reflected in data originally used for sampling. Some districts had specific research requirements, such as submission of a research application, a review of survey instruments, or security checks of site visitors; the study team fulfilled these requirements where relevant.

Several strategies were attempted to persuade reluctant school districts to participate in the study. These included a letter from the Director of the Child Nutrition Division of the USDA, a phone call from the Survey Director, intervention by the FNS Project Officer, and soliciting the encouragement of the State Child Nutrition Director. These strategies met with mixed success. Reasons cited by school districts refusing to participate included a negative attitude towards research in general, lack of resources, concerns about security and confidentiality, and concerns about intrusion on instructional time. When initially sampled districts refused to participate, recruiters contacted sampled replacement school districts.

Once a district had agreed to participate, recruiters made follow-up calls to sampled schools within that district. These calls included discussion of the school's willingness to participate in the study, characteristics of the student population, daily school and food service schedules, and the method to be used for obtaining parental consent. Any schools that refused to participate or were found to be ineligible were replaced with a sampled replacement school. Schools deemed to be ineligible were those that did not participate in the NSLP or that served only kindergarten aged children or younger, an exclusively adult population, or an exclusively special needs population. Non-traditional schools, such as charter, magnet, or alternative schools (for example, dropout prevention programs or vocational schools) were considered eligible and remained in the sample.¹

¹ These specific types of eligible schools were not tracked separately.

The schools were asked to choose between two options for the sampling of students. They could either provide a roster of students from which MPR would select a sample, or they could request a set of random numbers generated by MPR that the school would then match to a numbered list of students to select a sample. Under both options, the schools were free to exclude student names to protect student privacy, and only needed to provide ID numbers until consent was obtained from students' parents. Once the student sample was selected, consent forms were sent home with students. Each school was asked to designate a liaison for the study who would coordinate sending out consent forms and assist interviewers on the day of the visit. MPR followed up with the school liaisons to learn of any non-consenting students according to either the passive or active consent procedures being employed in a particular school.

The MPR survey team arranged the data collection schedule with each district after it agreed to participate, selecting a target week for the data collection that was mutually convenient. Within each district, MPR staff attempted to schedule student dietary recalls and interviews across the days of the week in each school type (elementary, middle, or high), so that the sample would be representative of students' dietary intakes on a typical school day. However, this scheduling was not done according to a strict formula and exceptions were made, as needed, in order to accommodate schools' preferences.

B. DATA COLLECTION

This section describes the types of data collected at the SFA-, school-, and student/parent-levels of data collection, the modes, data collectors, and respondents for the various surveys and forms, and, as relevant, special details regarding the processing of the data and the timing of the data collection.

1. SFA-Level Data: Initial Contact Survey and SFA Director Survey

The Initial Contact Survey was a hardcopy instrument split into two parts, with questions in the first half asked of the district and questions in the second half collected from the individual schools. A major purpose of the district portion of the survey was to collect information required for sampling decisions and to facilitate the scheduling of further data collection, and the information collected from the districts was used to determine if any re-sampling of schools was necessary. The survey also collected important background information concerning the three schools in the primary sample. One hundred and thirty SFA directors (or their designees) completed the first part of the survey. The study recruiters administered the first part of the survey by telephone during the course of the initial recruiting calls to the SFA or faxed the survey to the SFA for self-administration upon request. For analytic purposes, only the school-specific data were used, so the data were converted into a school-level file.

The original purpose of the school portion of the survey was to collect any necessary or helpful logistical information in advance of the school visits. However, it also provided useful information on the timing of school meals and the length of meal periods. This portion of the Initial Contact Survey was completed by staff in 280 schools scheduled for on-site data collection.² Study recruiters administered this portion of the survey to school staff by phone during the course of the recruiting and scheduling calls to the schools or, upon request, faxed the survey to the schools for self-administration.

The purpose of the SFA Director Survey was to collect additional information on the school meal programs in the SFA. This survey was completed by 129 SFA Directors. The survey was a hardcopy instrument consisting of 30 questions in the following topic areas: menu planning and

² This form is missing for seven of the visited schools.

computer systems, food purchasing, food safety and sanitation, nutrition promotion/education, meal pricing, and SFA director background and experience. The SFA directors were initially contacted by telephone and given the option of completing the survey by telephone or mail. The majority of surveys were completed by telephone. The survey was administered after the completion of school- and student-level data collection in each district, from May through September 2005, because data collection staff needed to focus on the other levels of data collection while school was still in session.

All completed Initial Contact Surveys and SFA Director Surveys underwent quality control review and were data-entered for the creation of an electronic data file.

2. School-Level Surveys (Excluding Menu Survey)

The school-level surveys (other than the Menu Survey, which is discussed next) included the Principal Survey, the Food Service Manager Survey, and a set of observation forms: the Vending Machine Checklist, the Alternative Food Source Checklist, and the A La Carte Checklist. All were hardcopy instruments. For those schools at which an on-site visit took place, all instruments were completed during the visit. A member of the data collection team offered to interview both the principal and food service manager directly. If either was not available, the appropriate survey was left for self-administration and retrieved later during the site visit. For those schools not visited, the Principal and Food Service Manager Surveys were conducted by telephone from May through August 2005 (with a small number of outstanding surveys completed later—for example, if the respondent was away for the summer). For schools with no on-site data collection, these surveys always took place after the Menu Survey was completed.

The Principal Survey was completed by 395 principals. This survey consisted of 35 questions in the following topic areas: meal-time policies, eating locations and environment,

vending machines, school stores, after-school programs, and nutrition education and promotion. Slightly more food service managers completed the Food Service Manager Survey (399), which consisted of 30 questions in the following topic areas: kitchen characteristics, meal prices, meal counting, meal periods, services for students with special dietary needs, and nutrition promotion/education.

The data collection team completed the three observation forms on-site with assistance from the school liaison or food service manager. All were hardcopy instruments. The Vending Machine Checklist consisted of a detailed inventory of on-site snack and beverage vending machines and their locations. The Alternative Food Source Checklist consisted of a detailed inventory of foods available at the school through sources other than the cafeteria, such as school stores, snack bars, or snack carts, as well as the hours of operation of these alternative food sources. The A la Carte Checklist consisted of a detailed inventory of foods sold a la carte in the school cafeteria at breakfast (if applicable) and at lunch. These forms were only completed for schools that were visited for student data collection.

The on-site data collection team also completed a Point-of-Sale Form and Milk Form, both with the help of the school food service manager, to aid in the classification of data collected during the dietary recalls with individual students. The Point-of-Sale Form listed each location where students could obtain food in the school, for later coding and matching with student reports on food sources. The Milk Form listed the types of milk available and the color carton, for use by the interviewers during the dietary recalls. All of the school-level surveys and forms, with the exception of the Milk Form, underwent quality control review and were data-entered for the creation of electronic data files. (Milk information was also captured in the Menu Survey, so the Milk Form was not needed for analysis.)

3. Menu Survey

School food service managers in a total of 398 schools completed a mail menu survey, with help by telephone from specially-trained MPR staff. The main purpose of the Menu Survey was to collect sufficiently detailed information on the foods offered to and selected by students in USDA reimbursable meals in order to complete a nutrient analysis. Menu data were collected over a typical five-day period (the target week); however, in some schools, data were available for only three or four days due to holidays or other school closings.³ School menu data collection took place between mid-January and late June 2005.

To capture all of the required information, the menu survey consisted of six different types of forms:

- The **Daily Meal Counts Form** collected counts of reimbursable meals for each day of the target week, separately for breakfast and lunch, by whether the meals were free, reduced price, or full price; in addition, it collected the dollar amount of a la carte sales each day.
- The **Reimbursable Foods Forms** (one each for breakfast and lunch) identified and described foods and beverages offered to students in reimbursable meals. To reduce burden on the respondent, common items were pre-printed on the forms and check boxes were included to describe such characteristics as the form of the food (fresh, frozen, canned); method of preparation (baked, fried); and whether modified for fat content or vitamin/mineral fortified. Respondents added foods not already listed; manufacturer, brand, and product code (if applicable); and portion sizes.

For each food item, the forms also collected the number of portions served, excluding a la carte servings. Because these data were not always readily available, respondents could report the total amount of each food item prepared or available, the amounts of each left over, and whether any portions were sold on an a la carte basis. Reimbursable servings were later calculated by MPR staff using estimates of the *percentage* sold a la carte (described below). One Reimbursable Foods Form was completed for each lunch and breakfast (if offered) meal during the target week.

³ A total of 397 schools provided menu data for school lunches: 329 (83 percent) provided data for five days, 66 for four days, and 2 for three days. School breakfast data were available for 331 schools: 284 (86 percent) provided five days of menus, 44 provided four days, and 3 provided three days.

- The Recipe **Forms** supplemented the Reimbursable Foods Forms by collecting recipe information for all menu items made by combining two or more foods or ingredients. Ingredients were listed and fully described, including the amount used to prepare the recipe; the Recipe Form also collected total yield (number of servings) and size of one serving. Respondents could substitute copies of their printed recipes instead of completing the Recipe Forms, adding additional information as needed.
- The **Self-Serve/Made-to-Order Bar Forms** described all food items included in various self-serve and made-to-order bars—for example, salad bars, deli bars, and condiment bars, if available. One form was completed for each type of self-serve bar offered during the target week. In a few instances where the items varied from day to day, additional bar forms were completed for each variation. Respondents recorded the total number of students that selected all or part of a reimbursable meal from each bar, each day of the target week, on the **Daily Counts for Self-Serve/Made to Order Bars** form.

In addition, technical assistants used the Proportion A La Carte Form to ask school food service managers who were unable to provide counts of reimbursable servings to estimate, for each food item listed on the Reimbursable Foods Forms and/or each self-serve food bar, the percentage of total sales that were a la carte. This form was administered as a telephone interview by the MPR technical assistants, once they had received the completed menu survey. The information was used in coding the menu survey but was not data-entered.

Each menu survey form was printed in a different color. To further assist respondents in locating the relevant forms, a complete set was provided for each day of the target week (except for the Proportion A La Carte Form). The five sets of daily forms were then packaged together, along with (1) an instruction booklet that included detailed guidelines for completing each form as well as sample completed forms, (2) extra copies of the Recipe Form, and (3) a pre-addressed overnight mail envelope for returning the survey materials to MPR.

Preparation for the menu survey began about six weeks in advance of the target week. The target week for menu data collection was scheduled with the SFA director and was generally the same for all sampled schools in the SFA. Every attempt was made to schedule the menu survey for the same target week that included the student data collection in the sampled schools, both to

contain the disruption of data collection to a single week and to allow school menu data to be used in analyzing students' dietary intakes.

Although the menu survey was self-administered, a small number of MPR staff were selected and trained by senior project staff to assist school food service managers in completing the survey. These staff, called technical assistants, served as the main point of communication between the school food service staff and MPR's project and survey directors. Technical assistants' primary responsibilities were (1) reviewing the menu survey instructions and materials with respondents, (2) assisting with any questions that arose while the forms were being completed, (3) reviewing, editing, and data retrieval after the forms were returned, and (4) collecting the necessary data to complete the Proportion a la Carte Form. The typical schedule of technical assistance contacts is summarized in Figure III.1.

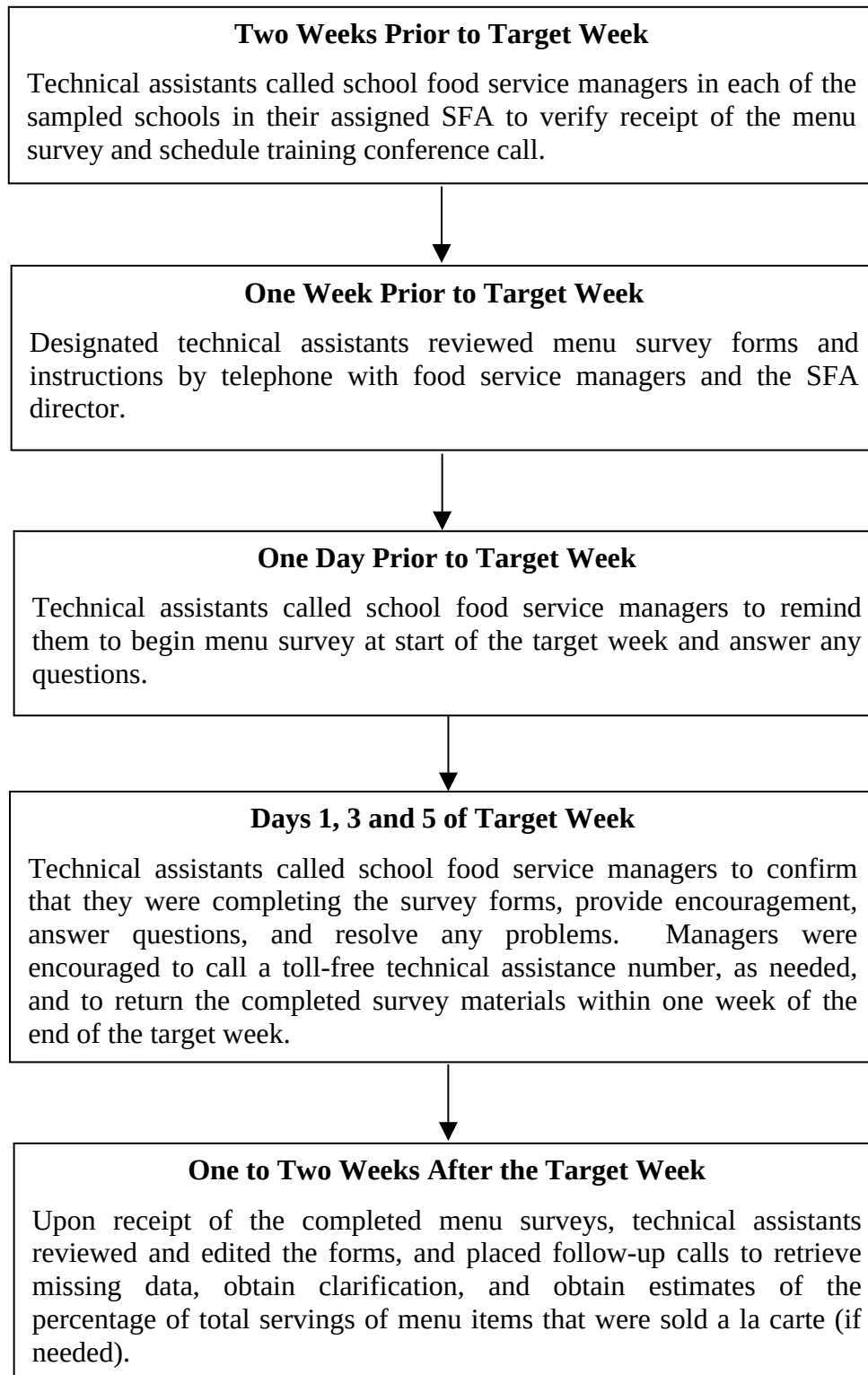
If the menu survey was not received within the expected time, additional follow-up calls were made to assess progress and encourage completion. In some cases, a new target week was assigned. Where managers were particularly reluctant to complete a menu survey, usually due to competing demands for their time, the SFA directors were asked to help obtain their cooperation. In addition, a small number of managers agreed to send printed menus, production records, and printed recipes in lieu of completing the standard forms. The technical assistants then used this information, making additional follow-up calls as needed, to complete a menu survey for those schools.

4. Student and Parent Interviews

The purpose of the Student and Parent Interviews was to collect information about students' school meal program participation, dietary intakes, and personal and family characteristics. 2,709 students completed the student interview, and 2,330 of their parents completed the parent interview.

FIGURE III.1

TECHNICAL ASSISTANCE SCHEDULE FOR MENU SURVEY



Parental consent was required for all students participating in the study. A consent form was sent home from school with each student selected into the sample, and a brochure describing the study in more detail and addressing frequently asked questions accompanied this form. A toll-free telephone number and a website were also available to provide more information about the study and address any questions or concerns. FNS's preference was to use a passive consent process, under which parents only needed to return the consent form if they were unwilling to grant consent. If the parent did not return the form, it was assumed that consent was granted. However, at particular schools' or school districts' request, the study used an active consent process, and a signed consent form was required for each student participating in the study. The study used an active consent process in approximately 23 percent of the schools, and this often led to lower consent (and participation) rates among students at those locations.

The data collection team conducted student interviews at all schools with an on-site visit. The original target number of interviews was 9 students at each elementary school, and 11 students at each secondary school; approximately 8 students at each school were expected to both complete the student interview and have their parent or guardian complete the parent interview. Sampled student lists were about triple the size of these targets, in anticipation of students being absent, having transferred out of school, refusing to participate, lacking parental consent, or being otherwise unavailable due to scheduling conflicts or other restrictions. As it became clear in the early stages of on-site data collection that the targets would not be met at some schools due to high rates of non-consent or highly restrictive limitations placed on the data collection (for instance, only allowing students to be interviewed after school, or during study hall) data collectors were instructed to complete additional interviews beyond the targets in schools where additional data collection was possible.

The data collection teams attempted to interview students in the order in which they were sampled. If certain students were not available, the data collector would go as far down the list of sampled students as necessary to meet the target or complete as many interviews as possible within the time available. All students, whether interviewed or not, within the portion of the list for which interviews were attempted became the final universe of sampled students (unless they were absent on the target day). Most, but not all, student interviews in a given school were conducted on a single day. Student interviews were generally conducted together with the dietary recalls and measurement of height and weight. The main exceptions were for secondary school students whose dietary recall interviews were sometimes conducted on the weekend, in order to collect data about Friday's meals (since the recalls asked secondary school students about dietary intakes over the preceding day). These students were usually interviewed and weighed and measured in school earlier in the week. As mentioned above, an effort was made to schedule interviews and dietary recalls in each school type uniformly across the different days of the week. However, since the dietary recalls for secondary school students collected data about the preceding day and the days covered by the recall were limited to school days, no secondary school student dietary recalls were conducted on Mondays, because the previous day was not a school day; as a result, student interviews in secondary schools were also typically not conducted on Mondays.

The student interview was a Computer Assisted Personal Interviewing (CAPI) instrument that typically took about 10 minutes to administer and was conducted immediately after the dietary recall if the two were being conducted together. The student interview consisted of up to 57 questions pertaining to eating habits, attitudes toward school meals, school meal participation, and activity levels.

The parent interview consisted of up to 84 questions pertaining to the child's health, eating habits and activity levels, parent's attitudes toward school meals, and household demographics and food security. For the parents of elementary students, the interview was a CAPI instrument and normally took place in the school, after class time, or in the parent's home in the evening, in the day or two following the student interview in school. The parent assisted in the completion of the student's dietary recall during the same meeting. The scheduling of the interviews with parents of elementary students was particularly challenging, since the data collectors could not schedule the interviews before the student interview was completed, as it was not clear which students would ultimately be participating in the study until the data collection actually took place. Once the student interview was completed, the protocol for the dietary recall required that the recall be completed with the parent's assistance within 48 hours. This left little time to contact the parent and successfully schedule and conduct an interview. In a few cases when an in-person meeting with the parent was not possible, the data collectors conducted the parent interview (and obtained the parent's assistance in completing the dietary recall) over the telephone. If the dietary recall was not completed with the parent's assistance within the prescribed time period, no attempts were made to complete the interview at a later time.

Parent interviews for secondary students did not face the same restrictions. Since the dietary recalls of secondary students were completed entirely by the student, the parent interview consisted of a Computer Assisted Telephone Interviewing (CATI) instrument conducted by telephone at a time convenient for the parent. There was no specific timeframe within which both the student and parent interviews needed to take place. Parent interviews for secondary students were completed between March and July 2005.

On-site data collection typically took place at three to four schools in a district over a two-week period. During the first week, a team of two data collectors, including a team leader,

would spend an entire day in each school, weighing and measuring students, conducting interviews and dietary recalls, and completing observation forms. The team would follow up with elementary students' parents in the afternoons and evenings. During the second week in each district, a single data collector would conduct second-day dietary recalls with a subsample of students and would attempt to complete any incomplete instruments.

5. Dietary Recalls

A single, in-person, 24-hour dietary recall interview was conducted with a total of 2,718 students in the 287 schools in the basic sample.⁴ The 24-hour recall for Day 1 (also referred to as the “target day”) collected the necessary dietary data to provide an estimate of students' intakes of food energy and nutrients on a typical school day. In general, the 24-hour recalls were conducted during the same target week as other components of the data collection in each school, and at the same time as the Student Interview and height and weight measurement. The second (Day 2) 24-hour recall was completed by a randomly selected sub-sample of 778 students (29 percent), about a week later.⁵ These second 24-hour recalls were used to estimate the percentage of students that had inadequate or excessive usual daily intakes of specific nutrients (see Volume II, Chapter V).

Twenty-four hour recall data were collected by highly-trained field interviewers using a modified version of the Automated Multiple-Pass Method (AMPM) interview. The AMPM instrument was developed by the Food Surveys Research Group at USDA and is primarily used in large-scale, national surveys of the dietary intakes of both adults and children (Raper et al.

⁴ The actual analysis sample was limited to students for whom both the 24-hour recall and parent interview had been completed (n = 2,314).

⁵ The interval between first and second dietary recalls for any particular student was no less than 3 days and rarely more than two weeks.

2004). The five steps (or passes) of the AMPM were designed to maximize the respondent's ability to recall and report foods and beverages consumed:

1. **Quick List.** Collects an uninterrupted listing of all foods and beverages consumed in a 24-hour period, the day before the interview.
2. **Forgotten Foods.** Probes for any foods that may have been forgotten during the Quick List, including beverages, sweets, snacks, fruit, vegetables, cheese, and breads.
3. **Time and Occasion.** Collects the time the respondent began to consume each reported food item and what he or she would call the eating occasion for the food (such as breakfast, lunch, dinner, or snack).
4. **Detail Cycle.** Through a set of standardized questions, probes for descriptions of each item, including the quantity consumed, where the food was obtained, whether it was eaten at home, and any additions to the food. Also probes for other items that may have been eaten between the reported eating occasions.
5. **Final Probe.** Collects information for any additional foods not previously mentioned.

The AMPM dietary recall does not collect intakes of dietary supplements.⁶

Several modifications were made to the basic AMPM interview to tailor the instrument to school children and the unique needs of SNDA-III. The main issues that led to protocol and software modifications were: (1) the time of the interview, (2) the targeted 24-hour period, (3) parental assistance for younger children, (4) the need for detailed information on the sources of foods and beverages obtained at school, and (5) an interest in capturing students' consumption of bottled water. In addition, since comparisons of the diet quality of school meal participants and nonparticipants were a main focus, all 24-hour recalls covered students' intakes only on school days (excluding weekends and holidays).

Children in middle and high schools were interviewed in the morning and reported the previous day's intake (from midnight to midnight). However, because young children tend to

⁶ Descriptive data about students' use of dietary supplements were collected as part of the Student Interview (age 12 and up) and the Parent Interview (all ages).

have difficulty recalling their intake, children in elementary schools were interviewed in two parts and with parental assistance.⁷ The first part of the dietary recall took place in the afternoon, after lunch, and children were asked to report everything they had consumed that day since awakening. The second part of the interview was conducted with the child and his or her parent (or other adult knowledgeable about the child's food intake outside of school), one to two days later. The child was the primary respondent and parents helped children recall and describe details about the foods and beverages consumed during the rest of the 24-hour period. The second part of the 24-hour recall took place at either the school or the child's home, or, occasionally, at a public library or other mutually convenient place.

Interviewers used several memory aids and other prompts during the Detail Cycle of the 24-hour recall interview to help improve the accuracy of the data. These included (1) a copy of the school menu, (2) a completed Milk Form, (3) portion size estimation tools, and (4) a completed Point-of-Sale Form. The school menu (for the day corresponding to the 24-hour recall) was consulted if a child reported taking a school meal but was having difficulty recalling the specific foods consumed. A separate Milk Form was completed for each school, which listed the percent fat, flavor, cap or carton color, and quantity of each type of milk served with reimbursable meals or sold a la carte. Interviewers used this form to assist the child in naming the type of milk he or she obtained from the school cafeteria or the size of the container.

To help report the quantity of food eaten, children were shown how to use the USDA two-dimensional food model booklet, measuring cups and spoons, and two-dimensional drawings of common sizes of bottled beverages and juice boxes. The food model booklet contained drawings of various measuring aids, such as glasses, mugs, bowls, mounds, circles, rectangles, and wedge

⁷ Some of the screens in AMPM had to be modified for young children to reflect the different target time frame and two-part interview format.

shapes of various sizes, thickness measures, and pictures that showed how to describe pieces of chicken. In all cases, students were asked to describe the amount actually drunk or eaten—not the amount served.⁸

In addition to information on the types and quantities of food and beverages consumed, the time each item was consumed, and the reported eating occasion (breakfast, brunch, lunch, supper, dinner, snack), the 24-hour recalls collected information on where each item was obtained (source of the food), and whether or not the food was eaten at school.⁹ For items obtained at school, students were asked to identify a specific location or “point of sale” in the school (for example, reimbursable cafeteria line, vending machine, snack window or cart, canteen, school store.) The various sources of food at each school were recorded on a Point-of-Sale Form and then used by interviewers during the 24-hour recall when probing for where the food was obtained. Information on the source of the food, and the extent to which a student could obtain a reimbursable meal from that source, was also used to help identify school meal participants. Additionally, this information is important for the analysis of the contribution of competitive foods to children’s diets.

Finally, the need to obtain information on student’s use of bottled water posed some complexities for the 24-hour recall interviews. The version of AMPM available at the time was not designed to collect detailed information about water consumption, except in the case of products that have calories. So, for example, the dietary system would not have captured students’ consumption of bottled water products, such as spring water, mineral water, flavored

⁸ While collecting the foods and amounts served to or selected by students would have helped identify NSLP and SBP participants, the AMPM software did not have this capability. The programming resources required to add this feature were much greater than those needed for the other AMPM modifications implemented for the study.

⁹ The AMPM food source screens were modified to allow the entry of additional codes corresponding to the locations on the Point-of-Sale Form when the food was reported as obtained at school.

water that was artificially sweetened, or water fortified with vitamins and/or minerals but containing no calories. Special procedures were therefore developed for using the AMPM “unknown food” feature to collect the required data on bottled water.

6. Height and Weight Measurement

A height and weight measurement was taken on 2,689 students selected for the dietary recall. In the analysis sample of 2,314 (with both completed recalls and parent interviews), 2,233 have valid height and weight data. To obtain the most accurate information, the measurements were taken in-person by trained interviewers using standardized procedures and a common set of equipment across all schools. The main reason for measuring height and weight was to obtain the data needed to assess students’ weight status on the basis of their Body Mass Index (BMI).¹⁰ Student height was also used as a covariate in regression models of mean dietary intakes of various nutrients.

Standing height was measured with a portable stadiometer or height measuring board (Seca model 214) using a slightly modified version of a procedure developed for the National Health and Nutrition Examination Survey (NHANES) and other national and international surveys (Shorr 1998). The child was asked to remove his or her shoes, hats, hair ornaments, or other items that might have affected the accuracy of the height measurement. Interviewers instructed the child to stand in the proper position against the stadiometer, but were not to touch the child. After lowering the headpiece to the top of the child’s head, the height measurement was read and recorded to the nearest half inch. A second height measurement was taken to reduce the effects

¹⁰ BMI is defined as the ratio of weight (in kilograms) to the square of height (in meters). The percentiles of the distribution of BMI are used to categorize whether a child is underweight, at risk for overweight, or overweight or obese based on national age and gender norms.

of any measurement or recording errors. (Height was measured a third time if the first and second readings differed by more than an inch.)

An electronic, digital-display floor scale (Seca model 840) was used to measure weight. In addition to removing shoes or boots, children were asked to remove any heavy outer clothing (such as a jacket, sweatshirt, heavy sweater), heavy jewelry, or anything else that might have interfered with the weight measurement. The child was instructed to step onto the scale and, when in the proper position and not moving, the interviewer read the digital display and recorded the measurement in pounds. (The scale weighed to the nearest 0.2 pound.) As for height, the weight measurement was repeated, and, if the two measurements differed by more than 3 pounds, a third measurement was obtained.

Both the height and weight measurements were recorded by the interviewers on a hard copy Weight and Standing Height Measurement Form. The form also required that interviewers select a code to describe the clothing worn by the student and any concerns about the accuracy of the measurements (for example, child refused to remove shoes/heavy clothing, cast on arm or leg, unable to place stadiometer headpiece on top of head/interfering hair). Factors that sometimes prevented interviewers from completing a measurement included an uncooperative child, a child in a wheelchair, or interruptions (for example, when a child was called back to class).

C. TRAINING OF DATA COLLECTORS

Field data collectors conducted the dietary intake interviews with students, student and parent interviews, height and weight measurements and also interviewed principals and school food service managers, as well as completing three checklists. Regional teams of two data collectors included a designated team leader and traveled to the different SFAs in their areas to collect data. Many of the staff were experienced field interviewers, and recruiting focused on

identifying staff with prior experience interviewing children in both school and home settings, or experience collecting dietary intake data.

The data collectors underwent extensive training over an 11-day period at a centralized location in Princeton, New Jersey. The training was led and overseen by senior project staff, nutritionists, and representatives of FNS and the Food Surveys Research Group at USDA. The training included supervised on-site data collection in several school districts, as well as in-depth coverage of the methodology and administration of each survey instrument. Training procedures included large group instruction and demonstration, role-playing exercises, and closely supervised small group practice sessions. A training manual was provided and included detailed procedures for each component of the data collection and question-by-question specifications for each instrument. Approximately 40 hours were spent training data collectors to use the Automated Multiple Pass Method (AMPM) Dietary Recall software package; staff from the Food Surveys Research Group at USDA led the first three days of this training. All data collectors were tested for proficiency in administering dietary recalls before being cleared to work with actual respondents. Additional topics covered in the training included gaining respondents' cooperation, respondents' rights, administrative reporting, and quality control and monitoring.

The data collectors included a team of English-Spanish bilingual interviewers. (All student- and parent-level survey instruments were available for administration in both Spanish and English.) Interviews in additional languages were rarely conducted, and only if a third-party translator was available to assist.

A separate team of technical assistants working out of MPR's Survey Operations Center in New Jersey was trained to assist by phone with the self-administration of menu surveys by school food service managers. Two half-day training sessions for technical assistants were

conducted in-person by a senior nutritionist and covered reviewing the menu survey forms, follow-up contacts with respondents, reviewing completed surveys and retrieval of missing data, and the Proportion A La Carte interview. Telephone interviewers at the same location were trained by senior survey staff to conduct follow-up interviews with the parents of secondary students (and also interviews with principals and food service managers in menu-only schools).

D. INCENTIVES FOR SURVEY RESPONDENTS

The study offered incentives to some groups of survey respondents. Food Service Managers received a \$40 gift card to a national bookstore chain; this card was included in their Menu Survey. Elementary students were given a \$5 bookstore gift card following the student interview and dietary recall, and their parents were given a \$20 bookstore gift card upon completing the parent interview and assisting their child with the dietary recall. Secondary students were given \$10 bookstore gift card following the student interview and dietary recall, and their parents were sent a \$10 check following the parent interview, which was conducted by telephone. School Liaisons, who coordinated on-site school- and student-level data collection, were given a \$40 bookstore gift card for their assistance.

E. SURVEY RESULTS

Recruiting SFAs to participate in SNDA-III was challenging, for several reasons. School districts face many requests for information and requirements to complete forms related to various funding sources; they also have security and confidentiality concerns. In addition, participation in the SNDA-III study was time-consuming for districts and schools. As described in Section B, MPR and FNS staff made extensive efforts to recruit SFAs.

Recruiting efforts led to an 83 percent weighted response rate among SFAs in the full menu sample and a 79 percent rate among SFAs selected for student data collection (Table III.2).¹¹ This rate is based on all SFAs ever released for recruitment efforts, including replacements for those that refused. Essentially all nonresponse at the SFA level was due to refusals; only one SFA agreed to participate (and provided school-level data) but did not complete the SFA Director Survey.

After the SFA agreed to participate, schools in the SFA generally agreed as well. About 95 percent of schools in SFAs that agreed to participate completed the menu survey, the criterion for considering a school a completed sample case; 93 percent of schools selected for both school- and student-level data collection participated.

Obtaining interviews with students was another challenging stage of data collection. The weighted response rate among students selected for data collection is 63 percent. There were several reasons for this, including consent and sampling in the schools. SFAs or schools required that parents either actively consent to their child's participation by returning a signed form (active consent), or that they be offered the chance to opt out of participation for their child by returning a form (passive consent). School staff assisted in circulating consent forms and reminding students and parents to return them. The response rate includes in the denominator families who did not return consent forms in sites requiring active consent. In addition, on the target day, some students were in school but not available to be interviewed; for example, if they were taking a test. Secondary school students, who were interviewed the day after their target day, were considered nonresponders if they were in school on the target day but not on the day after.

¹¹ These response rates were weighted using raw sampling weights—the inverse of the probability of selection, before nonresponse adjustment. They thus reflect the proportion of SFAs or schools nationally represented in the sample.

TABLE III.2
SNDA-III RESPONSE RATES

	Response Rate (Percentage)	Completed Sample Size
SFAs (Menu Sample)	83	130
SFAs (Student Sample)	79	94
Schools (Menu Sample)	95	398
Schools (Student Sample)	93	287
Students (Recall Sample)	63	2,709
Parents (Parent Interview Sample)	89	2,330

Source: School Nutrition Dietary Assessment-III.

Note: Response rates for schools reflect the percentage of eligible sample members participating, given agreement to participate at the SFA level. The response rate for students reflects the percentage of eligible students participating, given the school participated, and the response rate for parents reflects the percentage of parents participating, given their child completed a dietary recall. Response rates are weighted using raw sampling weights—that is, weights that correct for unequal probability of selection, before any nonresponse adjustments.

Almost all of the nonresponse to the recall and student interview resulted from failure to obtain active consent or from students being unavailable at the time of data collection. These two sources of nonresponse seem unlikely to be correlated with students' participation in school meals or their dietary outcomes. Nonetheless, it is possible that parents who had an interest in healthy eating, or, alternatively, concerns about their child's weight, would be more likely to return an active consent form. Further analysis of these issues may be warranted.

Even when students were interviewed in school, not all parents completed the parent interview. As noted in earlier sections, it was challenging to schedule the in-person interviews

within a short timeframe. Interviews with the parents of secondary students were by telephone, and an 80 percent response rate is the usual goal for a telephone interview. The weighted response rate for the parent interview, given their child completed the recall, was 89 percent.

IV. PROCESSING OF MENU AND DIETARY RECALL DATA

This chapter describes the methods used to review and prepare the school menu and dietary intake data for analysis. To answer many of the key research questions posed for the School Nutrition Dietary Assessment Study-III, raw data collected in the menu surveys and 24-hour recalls had to be converted into electronic data files that provided nutrient amounts and food group designations for individual food items. This involved reviewing, coding, and in the case of the menu survey, entering the data. Although the same computer-assisted food coding and nutrient analysis system was used for both the menu and recall data—Survey Net—the data preparation steps were quite different. Survey Net is based on the USDA National Nutrient Database for Standard Reference 16-1 (USDA Food Surveys Research Group 2004).

Processing of the 24-hour recall data was relatively straightforward in that the procedures employed were essentially the same as those used by USDA's Agricultural Research Service for the 1994–1996, 1998 Continuing Survey of Food Intake of Individuals (CSFII) and, beginning in 2002, the National Health and Nutrition Examination Survey (NHANES). However, since Survey Net was developed for the analysis of dietary intake data, it presented some challenges for the processing of school menu data. Most of these were met through the creative use of existing data fields¹ and training; thus, no software modifications were required for the menu data.

Section A briefly describes staffing and training for the menu and recall data processing effort. Section B covers the steps involved in processing the menu data, including review and

¹ Some of data fields in Survey Net that applied to dietary intake data, such as time of day, eating occasion, and where the food was obtained, were not needed for analysis of school menu data. Therefore, these fields were used for the entry of other information essential to the menu analysis, including daily meal counts and the number of reimbursable portions of each menu item served.

editing of the various forms, and entry of the data. Methods for obtaining nutrition information for commercially prepared school foods and quality control procedures are also described. Data processing for the 24-hour recalls is described in Section C and includes the procedures for using information from the corresponding school menus to code the recalls. A description of the food grouping system developed for SNDA-III was provided in the Volume I (Appendix B) and Volume II (Appendix D) reports and is not repeated here.

A. STAFFING AND TRAINING

The SNDA-III menu and dietary intake data were processed out of MPR's Cambridge, Massachusetts office under the direction of a senior nutritionist. A team of 12 nutrition coders and 2 coding supervisors was recruited and hired locally. Supervisors had advanced nutrition degrees, previous research experience, and had worked with several computerized nutrient analysis systems in the past. Coders had at least an undergraduate degree in nutrition or previous experience in food service, as well as a range of computer skills.

All nutrition coders and coding supervisors were trained by the senior nutritionist to use USDA's Survey Net food coding system and on the specific procedures developed for processing the SNDA-III school menu and dietary intake data. Three 8-to-10-hour training sessions were conducted to cover each of the main components of the data processing task. Training procedures included group instruction and demonstration, supervised hands-on practice, and exercises to be completed and checked by the supervisors before beginning work with "live" data. Detailed training and reference manuals were provided.

The first nutrition coder training session covered computer-assisted coding of the dietary recall data. This allowed coders to become familiar with working with Survey Net as it was intended (as opposed to coding of school menus). Training focused on locating foods in the

database, entering food amounts, coding multi-component foods (sandwiches, salads, some Mexican foods), recipe modification, and coding rules.

The second and third training sessions covered the review and editing, and data entry of the menu surveys. Prior to familiarizing the coders with the various menu survey forms, some background information was provided, such as the concepts of reimbursable versus a la carte menu items, meal patterns/components, menu planning systems, and quantity recipes. Coders were then trained to review and prepare the menu surveys for data entry and, subsequently, to add the menu items, portion sizes, recipe modifications, and meal and food count data into Survey Net.

B. PROCESSING OF THE MENU SURVEY DATA

Completed menu surveys were forwarded to MPR's Cambridge office by the technical assistants, after they had completed data retrieval and final editing. The surveys were logged into an Excel database as they were received, and tracked through each step of data processing: review and editing, supervisor review, data entry, and problem resolution. Coding supervisors assigned all surveys from a particular School Food Authority (SFA) to the same coder because of the potential for similarities in the menus, recipes, and purchased products used across schools.

1. Review and Editing of Menu Data

Each menu survey was reviewed in a systematic manner to identify occurrences of missing information, inconsistencies within and across the various forms, and instances where the number of reimbursable portions was not directly reported but could be calculated from the data provided. During the initial review, coders also identified unambiguous linkages between food items (for example, syrup served with pancakes) and commonly offered pre-prepared foods (pizza, chicken nuggets, burritos, etc.). Two coders were responsible for the specialized coding

of self-serve salad bars and other food bars. A checklist was used to promote consistency across coders and to ensure all review and editing steps were completed.

Missing Data. Attempts were made to reconcile missing data problems by cross-referencing with other menu forms in the survey and with surveys completed by other schools within the same district (or SFA).² For example, if a food description or the portion size of a food was vague or incomplete, coders checked if the same or a similar food was served on other days of the week and filled in the information accordingly. When it appeared that condiments had been omitted from a Reimbursable Foods Form, coders checked the forms completed for other menu days to determine if the school usually offered condiments when they served certain items and added them, if appropriate. The same procedure was used for salad dressings served with salads. Incomplete or missing manufacturer or brand information was obtained from forms for other days on which the food was served or from menu surveys completed for other schools in the district that offered the same items.

When portion size information could not be obtained from other survey forms, and in cases where the students served themselves (self-serve), coders assigned a standard default portion size. The default portion sizes used for lunch and breakfast menus were based on those used in the previous School Nutrition Dietary Assessment (SNDA) studies, except for salad dressing. The default of three-fourths of a tablespoon of salad dressing, originally defined for SNDA-I and maintained in SNDA-II, was considerably smaller than the average amount consumed by school-age children (according to the 1994–1996, 1998 CSFII). To ensure that the SNDA-III menu

² Nutrition coders did not directly contact school food service staff to inquire about menu information that was missing or needed clarification. Supervisors did however contact the technical assistants who were often able to answer the coders' questions.

analysis would not underestimate the fat content of school lunches that included salad dressings, the default portion size for salad dressing was increased to two tablespoons.

Reimbursable Servings. Data on the number of portions of each item served in reimbursable meals were required for the nutrient analysis of meals served to students (weighted analysis). The menu survey allowed respondents to either directly report the number of reimbursable portions of each food served *or* provide production data that could be used to approximate total portions served. This required two basic steps³:

- Determining the amount of food served by subtracting the amount of food left over from the amount prepared/available, and
- Converting the amount served to the number of portions served by dividing by the portion size.

For example, if the form indicated that 30 pounds of raw carrots were used, 2 pounds were left over, and the portion size was $\frac{1}{4}$ cup cooked carrots, the coder calculated the number of $\frac{1}{4}$ cup servings of cooked carrots that would yield from 28 pounds of raw carrots. The USDA Food Buying Guide for Child Nutrition Programs, and measurement equivalents and conversion charts were provided for reference and to minimize errors.

Another type of calculation was required when schools were unable to separate out the number of portions of a menu item sold in reimbursable meals versus a la carte. As described in Chapter III, the Proportion a la Carte Form documented the school food service managers' estimates of the typical percentage of total sales that were a la carte for these food items. Coders applied these percentages to the total number of servings provided (or calculated) to determine

³ Note that this is a simplified representation of the actual calculations required. In many cases, both steps also involved converting the food amounts to an equivalent unit of measure. So, for example, if 4 gallons of soup were prepared and 3 quarts were left over, the coder first converted gallons to quarts. Then, if the portion size was 1 cup, the total number of quarts served was converted to cups.

the number of reimbursable portions served. For example, if typical a la carte sales for 1-percent low-fat milk were reported as 20 percent of total served, and 200 milks were served, the total number of reimbursable milks served was coded as 160 $([100 - 20]/100 * 200)$.

Coders also compared the total numbers of reimbursable servings of entrees and milk, and the number of servings of individual menu items to the total number of meals reported for that day. Large discrepancies were flagged for supervisor review to ensure they were not due to miscalculation of the number of portions served. In addition, codes were assigned to flag entrees and accompaniments for future electronic data checks and analysis purposes.

Linked Menu Items. When a menu item, such as a condiment or topping, was clearly offered along with another food item, they were “linked” for analysis purposes.⁴ During the review and editing step, coders assigned special link codes to identify and categorize linked items. Salad dressings were always linked to salads. Other menu items were linked when the school food service manager reported offering the items together, as opposed to each item being available to all students (for example, spaghetti served with garlic bread, crackers served with salad, and rice served with stir-fried beef and vegetables).

In addition to the types of linkages mentioned above, link codes were assigned to the individual components of pre-plated meals, bag lunches, and multi-component foods to facilitate aggregation for nutrient analysis. A multi-component food was defined as a menu item for which ingredient and portion size information were provided, but which could not be entered into Survey Net as a single item. For example, chili cheese fries were not in the Survey Net database and could not be coded by modifying an existing recipe. Instead, this entree was entered as three separate items—french fries, cheese, and chili—and linked with the same code.

⁴ All condiments that could have been taken with more than one food or were one of several condiments that could all go with the same food were considered “unlinked” and not assigned codes.

Different link codes were assigned based on the types of foods being linked (for example, bread with additions, sandwich ingredients, entrees with accompaniments).

Pre-prepared School Foods. Many commercially prepared foods are formulated specifically for school food service, sometimes with less fat, more vitamins or minerals, or added protein. As a result, the nutrient content of the pre-prepared school foods reported on the menu surveys may not be accurately represented by a similar product in the Survey Net nutrient database. During the review of the menu surveys, coders assembled a list of the most commonly served pre-prepared school foods in order to subsequently obtain nutrient information that was specific to the product being served. The types of foods tracked included:

- French fries and other frozen potato products
- Chicken patties, nuggets, and related products
- Frozen pizza and pizza products
- Burritos
- Beef patties, beef crumbles
- Barbequed pork patties
- Chicken fried steak
- Fish patties, nuggets, sticks
- French toast sticks, waffle sticks, pancake-on-a-stick
- Frozen breakfast sandwiches (egg and cheese, sausage and biscuit)
- Breakfast pastries fortified with vitamins and/or minerals (Super Buns, Super Donuts)

A total of 1,357 unique pre-prepared school foods were logged into a spreadsheet, along with the manufacturer/brand, product code (if available), portion size, and identification numbers for all schools that served the product.

Self-Serve Food Bars. Coding the self-serve salad bars, theme bars (Mexican, Italian, potato), and condiment or fixins' bars was particularly challenging. By definition, students served themselves, there were few pre-portioned items, and the combinations of foods taken were not known. For example, entree salad bars offered the option to take different types of meats, cheeses, eggs, vegetables, and other items. It was unknown what types, combinations, and amounts of different food items each student truly selected from the food bar. Therefore, in order to define an average serving, detailed coding rules were developed for each type of food bar and for each meal component offered on the food bar, using a methodology employed by the previous SNDA studies. This approach assumes that students are offered everything on the bar and assigns default portion sizes to individual items on the bar based on the minimum portions required for each specified meal component in food-based menu planning, or defined default portion sizes for items such as condiments and toppings.

To ensure consistency, two lead coders were specially-trained to review, code, and enter the data from the Self-Serve/Made-to-Order Bar Forms.

2. Entering Data into Survey Net

Once a menu survey was reviewed, edited, and cross-checked by a supervisor or lead coder, it was ready for entry into Survey Net. Coders entered the information using procedures developed specifically for this study. A separate file was created for each school, with records for each lunch and breakfast menu recorded in the menu survey. The date and total number of reimbursable meals served was entered for each menu.

Food items from the Reimbursable Foods Form were matched to the closest food in the database, considering such characteristics as the form of the food (raw, cooked, canned), preparation method (baked or fried), and any modifications that might affect nutrient content—particularly fat. To expedite the food selection process, coders were provided with search terms

and food codes for commonly served foods. Information on portion size (reported or the assigned default) and the total number of reimbursable portions served was also entered for each menu item. In addition, for selected menu items, the link codes and entree and accompaniment flags that were added to the forms during editing, along with any special instructions pertaining to how a food should be treated in the analysis, were entered into Survey Net.

A set of coding guidelines was developed to assist coders and standardize entry of vaguely reported foods. The menu coding guidelines were designed to reflect the most common school food practices, which did not always correspond to the Survey Net “not further specified” option typically used when complete information has not been provided. For example, if a school reported serving a baked potato but did not specify whether fat was added in cooking, the options for entering the baked potato in Survey Net included fat added, no fat added, and not further specified (NFS), which assumes fat was added. The menu coding guideline for this scenario was to select baked potato, fat not added.

Special procedures were developed for entering school recipes, self-serve food bars, and pre-prepared school foods (discussed in the next three subsections). For self-serve food bars and pre-prepared school foods, “placeholder” food codes were entered in the Survey Net menu files to flag the items for subsequent replacement of nutrient data.

3. Recipes

Survey Net was not designed to allow users other than USDA to add recipes to the database. However, existing recipes in the database can often be modified to more closely match the foods reported eaten, or in this case, the recipes provided by the schools. Coders followed specific guidelines to decide if recipe modification was appropriate. Table IV.1 shows the types of allowed modifications, which were modeled after recipe modification guidelines provided by USDA’s Food Survey Research Group.

The decision to modify a recipe was based primarily on the importance of the modification to the overall fat content of the food. For example, if the school provided a recipe for a ham and cheese sandwich that was comprised of turkey ham and reduced-fat cheese, a recipe modification was created to account for the lower-fat foods reported in the recipe. Another consideration was the amount of the meat/meat alternate in school-prepared sandwiches, entree salads, and some Mexican foods, compared with the standard recipes for these foods in Survey Net. Single serving recipes for sandwiches, Mexican entrees and entree salads were modified when the amount of meat, cheese or bread provided in the school recipe differed from the Survey Net recipe by more than one-half ounce. In order to identify the type of modification, the ingredients and/or amounts that were changed were noted in the name assigned to the new recipe.

There were limits to the feasibility of modifying recipes depending on how the recipe existed in Survey Net—as a quantity recipe or a single-serving recipe. Since the single-serving recipes had a yield of one, both the amounts and types of ingredients could be modified without affecting the yield. However, for quantity recipes, modifications were limited to ingredient substitutions; changes to ingredient amounts could not be made due to the unknown effect on the recipe yield. Complications also arose when changing the type of meat in a quantity recipe. The form of the food (raw versus cooked) to be substituted was not always comparable to what was in the recipe. For example, cooked ground turkey (the only form of ground turkey in Survey Net) could not be substituted for raw ground beef in a recipe due to the effect on fat and moisture losses. In order to calculate the yield of a recipe, Survey Net takes into account the moisture and fat retention of each ingredient after cooking. Substituting a different form of an ingredient and/or altering the ingredient amount in quantity recipes would have required entering retention factor codes for each altered ingredient, which is not a simple or straightforward process. Due to their complexity, these types of recipe modifications were not used.

4. Self-Serve Food Bars

Each unique self-serve food bar was entered separately from the rest of the menu survey, as if it were a “menu” of all of the food items offered on the bar. Default portion sizes were assigned to individual items on the bar based on the minimum portions required for specific meal components in food-based meal patterns: fruits/vegetables, bread/grain products, and meat/meat alternates. (Milk was not usually included on food bars.) For non-meal pattern items, such as condiments, toppings, salad dressing, and desserts, the same default portion sizes were used as for self-serve menu items not on bars.

If more than one option within a meal component group was offered, they were entered so as to give equal weight to each food item. This was accomplished by creating a recipe. The recipe “ingredients” consisted of a full portion of each item from the meal component group available on the food bar, and the recipe yield (number of servings) equaled the total number of items or ingredients. For example, a sandwich bar offered a choice of turkey, ham, or tuna, and a choice of white bread, a hoagie roll, or wheat bread. The recipe created to represent one average serving of meat from the bar would have a yield of three servings (since there are three meat ingredients). The coding rules for a sandwich bar also called for *two* average servings of breads/grains. In these cases, where the coding rules called for more than one serving from a meal component group, the yield of the recipe was equal to the total number of ingredients, divided by the desired number of servings. Thus, in this example, an average serving of breads/grains would have a yield of 3 bread/grain choices divided by 2 servings, or 1.5 servings.

An average serving from the entire self-serve bar was the simple sum of the average nutrients per serving for each of the meal components included in the bar.

5. Pre-Prepared School Foods

Pre-prepared school foods accounted for an unexpectedly large percent of the total food records in the menu data. Overall, there were 5,306 occurrences of these foods in the daily lunch menus, which accounted for 14 percent of the 38,529 total food records. In addition, there were 1,252 occurrences of these foods in the daily school breakfast menus, which was 6 percent of the 19,673 total foods. Pre-prepared school foods were most often served as entrees or meat/meat alternates and comprised 43 percent of all entrees or meat/meat alternates at lunch, and 34 percent of all entrees or meat/meat alternates at breakfast.

Since manufacturer food labels were not collected from individual schools, nutrient and ingredient information for pre-prepared school foods was researched on the Internet and obtained from selected manufacturers. The most frequently logged items from the top 30 manufacturers were selected for additional research on nutrient and ingredient information by contacting manufacturers. Thirty-eight food type groups were created to identify which products needed further research. Food type groups were defined as foods that seemed essentially “the same” based on their food description and any nutrients available, and foods that were offered in 10 or more schools. Products that were not assigned to a food type group could not readily be grouped with other products and were foods offered in fewer than 10 schools.

Several options were investigated for improving the accuracy of the nutrient coding of these commonly served school foods. A sample of pre-prepared foods was used to examine the differences between the manufacturers’ nutrient values and those implied by using the closest match in Survey Net or the USDA National Nutrient Database for Standard Reference. It was determined that using either of these databases to code the foods would lead to systematic errors in estimates of some key nutrients in school menus (namely fat and calories). It would have also

affected the accuracy of the intake data since menu survey data was used for students who obtained meals from the cafeteria (as discussed in Section C of this chapter).

One hundred pre-prepared products from various food type groups were sent for further analysis to the Agricultural Research Service (ARS) to obtain an accurate, expanded set of nutrients for these foods. Selection of these products was based on foods that fit into a food type group, were commonly served in schools, and that had accompanying ingredient information. The nutrient information provided by ARS was used to replace placeholders in the menu data with the nutrients for the exact products researched. Additionally, the nutrient information from ARS was used to impute nutrient values for similar school foods within food type groups that were not specifically researched by ARS—products that did not have manufacturer or product information provided or specific products that were not commonly served. Foods that were not assigned to a food type group were coded in Survey Net with the most comparable item in the database.

Nutrient replacement for school food placeholders used as single menu items (e.g., pizza, french fries, burritos) was the simplest in the placeholder replacement process. Cases where a school served two or more different products (with different nutrients) that were coded with the same placeholder code were manually cross-checked with the hard copy of the menu to ensure each food received the correct replacement nutrients. For example, a sausage pizza and a pepperoni pizza, both coded with the meat pizza placeholder, were served on the same day or on different days in the same school. To assign the correct nutrients to each type of pizza, the hard copy of the menu was reviewed to determine which pizza was served on each day with the correct number of servings. Pre-prepared school foods used as ingredients in modified recipes (beef patty used in a cheeseburger recipe), within aggregated multi-component foods (chili cheese fries), and on self-serve food bars (beef crumbles on a nacho bar) required a more

involved process for replacing placeholders with more accurate nutrients due to Survey Net's coding system. The majority of pre-prepared school foods received nutrients imputed from foods directly researched by ARS. Over 1,000 foods in the menu data received exact nutrients from ARS specific to that product. A relatively small number of pre-prepared foods (less than 400) received nutrients from Survey Net.

6. Quality Control Procedures

During the initial phases of menu data processing, supervisors reviewed each coder's menu survey editing and entry for one district (three to four schools). Coders received detailed feedback and the process was repeated until a level of accuracy greater than 90 percent was achieved. In addition, during the editing phase, each menu survey was cross-checked by a second coder and any discrepancies were resolved by supervisors.

Menu data entry was also carefully reviewed by supervisors to ensure that the appropriate food selections were made from the database, portion sizes were entered correctly, coding rules were applied when necessary, and recipe modification guidelines were followed. Overall, full quality review checks were conducted for 23 percent of all menu surveys. A similar procedure was followed for the quality reviews of coding and entry of self-serve food bars.

In order to maintain standardized procedures, coders attended periodic meetings and received "coder updates" clarifying issues that were identified or changes to procedures. Throughout the editing and entry phases, coders documented issues that arose in a central location, which facilitated supervisor follow-up. Coding guidelines were updated regularly, and coders were required to review updates each day. Supervisors were available at every shift to answer questions and resolve emerging issues. The senior nutritionist met weekly with coding supervisors to discuss coding progress and resolve coding issues they needed help with.

Data Cleaning. Once all of the menu information was entered in Survey Net, a set of detailed data checks were performed to identify potential coding errors. Problem cases were identified, and each was reviewed and corrected by coding supervisors. The cleaning runs included the following types of checks:

- **Basic Data Integrity.** Daily menus were checked for missing data, duplicate entries, and valid values for the following items: school ID numbers, consecutive menu days and dates, minimum number of meal components, and number of meals served. Individual menu items were checked for valid and non-missing portion sizes, number of portions served, appropriate linking codes, and entree and accompaniment identifiers. Problem cases were identified and checked against hard copy menu surveys, and corrections were made as necessary. Self-serve food bars were also checked for valid values and the minimum expected items based on the type of bar.
- **Out-of-Range Menu Items.** Nutrients per serving for calories, total fat, and sodium were reviewed for individual foods to identify possible outliers in the menu data. Foods with nutrient values that were below the 5th or above the 95th percentile were identified and checked against hard copy menu surveys to verify entry. Corrections were made where appropriate. The same procedure was followed for self-serve food bars, with the assumption that the total nutrients for any particular food bar “menu” would be reasonably close to the expected range for one serving from the particular meal component group in which the bar falls. For example, the range of nutrients for entree salad bars, Mexican bars, and sandwich bars should have approximated the nutrients for other “entrees” on the main menus.
- **Recipe Modifications.** Every recipe modification created by coders was individually reviewed by a coding supervisor. Recipes were checked for compliance to guidelines and approved when acceptable. Incorrect or unnecessary recipe modifications were adjusted or deleted.
- **Imputing Missing Serving Data.** Accompaniments were the most common type of menu item missing data on the number of portions served. This was mainly due to the nature of the data being reported as “self-serve.” Missing data for linked items (i.e., salad dressings and linked accompaniments) was imputed based on the servings reported for the menu item to which each was linked. For unlinked accompaniments, serving weights were imputed based on the mean/median number of servings of accompaniments per meal, in schools that provided servings data.
- **Over-Reporting of Portions Served.** Checks were run to identify cases where the number of servings reported for a menu item was greater than the total number of meals served. The number of servings of milks, side salad bars, french fries and tater tots, desserts, entrees, and salad dressings were adjusted to ensure that the weighted analysis would not overestimate the nutrient content of meals served. This was based on the assumption that students generally select no more than one milk, one entree

and one serving of any particular side item per reimbursable meal. In addition, instances where the number of portions served for french fries, tater tots, and salad dressings was greater than the number of meals served were often cases where the manager provided the information as bulk amounts prepared and left over.

C. PROCESSING OF THE DIETARY RECALL DATA

The 24-hour dietary recall data for SNDA-III were collected and processed with USDA's Dietary Intake Data System (Raper et al. 2004). This system includes three main components:

1. The Automated Multiple-Pass Method (AMPM), a computer-assisted interview used to collect dietary intake data (described in Chapter III);
2. The Post-Interview Processing System (PIPS), which extracts the recall information from Blaise data files, automatically assigns food codes to certain foods, and formats the data for further coding and analysis; and
3. Survey Net, the computer-assisted food coding and nutrient analysis system, which includes USDA's Survey Nutrient Database.

The recall data from the AMPM interviews were processed through PIPS by senior systems analysts in MPR's Princeton's office. The data were batched by school and transmitted to the coding supervisors in Cambridge. Each batch received was logged into an Excel database, and tracked through the various data processing and quality review steps, using similar procedures as used with the menu data. Coders worked on a single batch of recall data at a time and were responsible for coding all records within the assigned batch (usually ten or fewer).

1. Coding Foods and Amounts

Food Coding. The food description and quantity data reported in AMPM for each food were automatically imported into Survey Net, appearing in a text box at the top of the first food coding screen. This information was used to select the most appropriate food from the database and to enter the amount consumed. Some commonly reported foods were "auto-coded," that is, during PIPS processing, the foods were automatically linked to the appropriate Survey Net food

code. For these items (for example, whole milk, apple, sugar, catsup), only the amount eaten needed to be entered by the coder. Other data collected in AMPM were pre-filled into various Survey Net fields (for example, student ID, date of birth, date of intake, time of eating occasion, and food source).

When selecting matching items in the Survey Net database, coders were instructed to pay particular attention to the form of the food reported, the preparation method, and any modifications in fat (for example, fat-free mayonnaise, 90 percent extra lean ground beef) to accurately reflect nutrient content. Search terms and food codes were provided to help coders locate difficult-to-find foods in the database. Foods that were reported as having been prepared or obtained at school were coded in Survey Net—even if the item did not have a good match in the database—since the presumably more detailed information from the school menu data was ultimately to be used to code these foods (unless they were items sold through vending machines or a la carte only). If a coder could not find a food in the database or was unsure about the food code selected, a flag was set to signal that supervisor review was needed.

Coders sometimes created recipe modifications to more closely match the food description information reported in the recall. The procedures and guidelines for modifying USDA Survey Database recipes were very similar to those described previously for the school menu data. (See also Table IV.1.) Typically, modifications were created to account for different types and amounts of fat used in cooking or preparing a food. Unlike recipe modifications for school menus, changes to ingredient amounts were only made for non-standard dilutions of soups and beverages from concentrate, and for removing fat from a recipe.

Coding Amounts. After the appropriate food selection was made, coders entered the amount of food consumed. Food amounts could be entered by shape, weight, volume, or dimensions in inches, and the system then calculated and stored the weight in grams. The

TABLE IV.1
 RECIPE MODIFICATION GUIDELINES FOR SNDA-III MENU SURVEYS

	Ingredient to Modify					
	Type of Fat Used (Animal Fat, Butter, Oil, Margarine, Shortening)	*Fat Not Used (in prep)	Type of Milk (1%, 2%, nonfat etc.)	Type of Mayonnaise or Salad Dressing (reduced fat, fat-free, low calorie or diet)	Type of Meat (ground turkey; chicken; turkey deli meat or sausage; extra lean beef; beef or pork crumbles, etc.)	Type of Cheese (reduced fat, low fat, part skim only)
Food On Menu Survey						
Baked desserts (homemade)	X	X				
Cooked cereal (grits, oatmeal, etc.)	X	X	X			
Condensed "cream" soups			X			
Dry beans, peas, other legumes	X	X				
Eggs & omelets	X	X				
Garlic bread	X	X				
Macaroni & cheese, other mixed dishes with cheese	X	X	X			X
Mixed dishes with meat (in spaghetti sauce, meatloaf, chili, etc.)					X	
Pudding (homemade or mix)			X			
Rice/pasta	X	X				
Salads – NOT greens based (e.g., tuna salad, chicken salad, pasta salad, coleslaw)				X		
Sausage gravy (homemade)	X		X		X	
Vegetables	X	X				

*Applies when survey code for 'fat not added' is not added' is not an option in Survey Net.

RECIPE MODIFICATION GUIDELINES (CONTINUED)

	Ingredient To Modify						
	Type & Amount** of Fat Used (butter, oil, margarine)	High Fat Ingredients (omit or add**: margarine, sour cream, cheese, guacamole, nuts)	Amount of Bread/Grain** (only if differs by more than ½ oz)	Type of Mayonnaise or Salad Dressing (reduced fat, fat-free, low calorie or diet)	Type of Meat (beef, pork, turkey, chicken; strips, filet, patty, steak;; breaded; etc.)	Type of Cheese (reduced fat, low fat, part skim only)	Amount of Meat / Meat Substitute** (only if differs by more than ½ oz)
Food on Menu Survey							
Sandwiches (burgers, subs, chicken, deli meat, tuna, pbj, egg & cheese/meat)	X	X	X	X	X	X	X
Mexican entrees (tacos, nachos, burritos, etc.)	X	X	X		X	X	X
Entrée salads (chef's, cobb, asian chicken, etc.)		X		X	X	X	X

****NOTE:** Modifications that involve changes to ingredient amounts or the addition of ingredients (shaded in yellow) apply only to non-commercially-prepared items with recipes that make 1 serving. Please do not add or modify ingredient amounts in quantity recipes, i.e., recipes that make more than 1 serving. In addition, if the closest Survey Net recipe is a quantity recipe or scaled to yield 100 grams, do not add ingredients or modify ingredient amounts. Adapted from Recipe Modification Guidelines, Food Surveys Research Group, Agricultural Research Service.

amount and unit of measure (cup, grams, piece, etc.) were entered separately, generally, exactly as they appeared in the Survey Net text box. During the AMPM interview, as described in Chapter III, a food model booklet of two-dimensional drawings of various sizes of glasses, mounds, bowls, and other measuring aids was used to help respondents estimate the portion size of foods consumed. In these instances, the coder directly entered the code representing the specific model used (for example, B4B represented a particular size bowl and the level to which the bowl had been filled).

If an amount was reported as an imprecise measure, coders used Survey Net's built-in portion size guidelines to code the amount. For example, the system could convert a "handful" of chips or a "swallow" of juice into a gram weight. Survey Net had an additional feature that allowed coders to enter, when necessary, a default amount for selected foods based on the amount consumed of another food. Default amounts were available for cream or milk added to a beverage, milk added to cereal, and sugar added to a beverage. For single food items that had a reported amount of "don't know," coders entered the "quantity not specified" option in Survey Net. This amount is intended to represent the average or typical amount consumed by persons of all ages; it was not available for many foods in the database, and thus was not widely used.

If the amount reported could not be entered using any of the methods described above, coders first referred to established coding guidelines (see below) to determine how to enter specific types of reported amounts. If the coding guidelines did not resolve the issue, coders entered the amount as "unknown" and flagged the food for supervisor review.

Combination Foods. Foods added to another food (for example, margarine added to toast) or a group of foods or ingredients that constituted a single food (such as a sandwich made up of bread, peanut butter, and jelly) were considered combination foods. These foods were usually designated as combinations when they were collected in AMPM, and the associated combination

type code that linked the foods together was pre-filled in Survey Net. However, there were situations in which a combination type code had not been assigned but needed to be added to appropriately link foods together. For example, if cheese and crackers was reported as one food in AMPM but required two food codes in Survey Net, the coder assigned the same combination type code to both the cheese and the crackers.

Coding Guidelines. Coders were provided with a detailed set of coding guidelines (coding rules) for use with the dietary recall data, to help resolve common coding problems and promote consistency across coders. These guidelines were adapted from materials provided by the USDA Food Surveys Research Group and reflected coding situations that had occurred in past food intake surveys. Coders referred to these guidelines when food descriptions or amounts were missing, or the information reported was vague or contradictory. Additional coding guidelines were established throughout the coding period as new issues arose.

Missing Foods/Amounts. If a new-to-market food or an unusual food was not found in the database and could not be coded using existing food codes or modifications, the closest match in Survey Net was selected and the food was flagged for supervisor review. Coding supervisors reviewed the foods flagged for unknown or questionable foods and amounts, and recorded the details in a central location for further research. (The “unknown food” option in Survey Net was not used.) Designated coders searched the Internet and local supermarkets to obtain additional information on the unknown foods and portion sizes of these foods. Once decisions were made by supervisors, coders edited the recalls to reflect the most appropriate food selection and portion size.

Bottled Water. Certain types of bottled water reported in the recalls did not have a corresponding food code in Survey Net. These included noncarbonated sweetened water; noncarbonated unsweetened water; noncarbonated water with artificial sweetener; and

noncarbonated water fortified with vitamins and/or minerals. Supervisors created recipe modifications (using the code for water as an ingredient) for these items to code them in Survey Net. For any of the bottled waters that contained calories, vitamins or minerals, the modification codes served as “placeholders.” Nutrient information was researched for these products and added to the recall records prior to nutrient analysis.⁵

2. Quality Control Procedures

During the initial phase of coding the 24-hour recall data, coding supervisors conducted quality review of each coder’s work. The review ensured that the proper food code was selected, the portion size was entered correctly, recipes were modified according to guidelines, and relevant coding guidelines were applied. Each coder continued to receive review and feedback until performance reached an accuracy level greater than 90 percent. Over the course of dietary recall coding, a minimum of 5 percent of each coder’s work was carefully reviewed by coding supervisors.

A supervisor was available at every shift to answer coders’ questions and resolve issues. Coders also communicated with supervisors through the notepad feature in Survey Net, and any issue that needed supervisor attention was flagged in the “request review” field. Supervisors were required to review and approve every flagged item and corresponding note entered by the coders. Coding guidelines were updated on an ongoing basis to reflect issues that arose during the coding process and their resolution. Before logging into Survey Net, coders received an automatic daily reminder to check for these updates. Meetings led by the supervisors were held periodically to discuss coding issues and changes in procedures. In addition, coding supervisors

⁵ The only other foods “added to the database” in this way were breakfast bars, some of which were highly fortified with nutrients and differed by brand.

met regularly with the senior nutritionist to make decisions about new coding guidelines and coding unusual and new-to-market foods.

Automated edit checks built into the Survey Net system also played a role in quality control. Each time a portion size was entered for a food, Survey Net performed an automatic gram weight check. If the gram weight was above or below a pre-established range for each food, the coder was alerted to check for an entry error. If the weight remained above or below the range, Survey Net flagged the food to prompt for supervisor review. Logic checks were also built into the AMPM system to ensure quality and consistency of the data. These included checks on variables such as eating occasion, time of day, and foods eaten in combination. For example, an “eating occasion” logic check did not allow the reporting of two different eating occasions at the same time.

Data Cleaning. After about half of the batches of 24-hour recall data were coded, they were run through Survey Net’s analysis system which performs systematic checks of data integrity. The resulting error report identifies records with unresolved missing foods, unknown portion sizes, invalid food codes, missing nutrient files, and recipe modifications that have not been approved by a supervisor. Error reports were initially reviewed by the senior nutritionist, and problem cases were edited in Survey Net by the coding supervisors. The analysis program was re-run until all of the batches had been checked and all of the errors were resolved.

Cleaning runs were also developed to check for extreme nutrient values that may have resulted from entry errors. Coding supervisors reviewed the top and bottom 5 percent of cases (24-hour totals) for calories, total fat, sodium, and calcium for two age groups (less than 9 years and 9 years and older). All reported foods in these records were reviewed for potential coding errors. Data were only re-coded when there was an obvious coding error or there was clear evidence that the reported portion was implausible. The latter case usually occurred when the

portion size had been reported in three dimensions (length, width, height) and these were recoded with a default portion size.

3. Linking of Menu and Recall Data

After the dietary recall data were cleaned and nutrient files were created, foods and beverages obtained at school (from sources other than vending machines or strictly a la carte venues) were linked to data on menu items offered in reimbursable meals. The primary goal of linking the menu data with the recall data was to ensure that National School Lunch Program (NSLP) and School Breakfast Program (SBP) foods were represented in the analysis as accurately as possible. For example, rather than hamburgers or cheese pizzas obtained at school being consistently represented by the “default” or average values available in the nutrient database, the nutrient values of the hamburgers and pizzas actually served in each child’s school were used. Thus, if a school purchased extra-lean hamburger patties or pizzas made with less or low-fat cheese, this was reflected in the dietary intake data. A secondary goal was to improve measures of target-day participation.

The process of linking the menu and recall data involved several steps and required extensive manual coding and careful quality control review. The first step was to compare, within a given school, all foods reported by students as having been obtained at school from sources other than vending machines to corresponding menu data. A flag was assigned to recall foods that appeared to be “on menu”—foods described by students that were purchased in the school cafeteria and seemed to be similar to school menu items. (This “on menu” variable was one of the data elements considered in constructing measures of target-day participation in the NSLP and SBP.)

Next, a linking system was developed to connect the recall foods to the corresponding menu items. The approach to linking differed depending on whether the food was (1) a single food,

such as a slice of pizza or a pre-prepared burrito; (2) a multi-component food—an item that included several components, each of which was coded separately, such as a sandwich; or (3) a single ingredient within a multi-component food. Multi-component foods were the most complicated to deal with because these foods were entered into Survey Net differently for menus and recalls. In the recalls, interviewers asked children for details about each component and each component was entered and subsequently coded separately (bread, peanut butter, jelly). For the menus, on the other hand, multi-component items were generally entered as recipes, resulting in a single entry for each multi-component food (peanut butter and jelly sandwich). Thus, the linking system had to link multiple foods in the recalls with one food in the menus.

After recall foods were linked to appropriate menu items, two different methods were used to determine the nutrient values that would be used in the recalls (in place of the original Survey Net values). For foods that could reasonably be considered as “units” (for example, pieces of pizza, pre-prepared burritos, cookies, chicken nuggets), one “recall” unit (or fraction/multiples thereof) was replaced with one “menu” unit. For foods that were reported by volume or by other “non-unit” measures using cups, bowls, and mounds shown in the food model booklet, the reported portion size was used, in combination with the nutrient values for the associated menu item, to estimate the nutrient content of the volume of food consumed. So, for example, a cup of “recall” chili was replaced by a cup of “menu” chili.

In rare cases, nutrient replacement could not be completed. Examples include situations where a student reported a hamburger patty and a bun but the corresponding menu item was a cheeseburger, or a student reported eating the entire hamburger patty, but only half the bun. Because the Survey Net analysis files did not provide ready access to ingredients within recipe items, it was not possible to make replacements at the ingredient level.

4. Final Data Files

Preparation of the 24-hour recall data for analysis involved a combination of manual and automated editing. As noted previously, the Survey Net recall files were first edited manually, based on nutritionists' reviews of automatically-generated error reports and cleaning runs for basic data integrity and extreme nutrient values. Once the editing in Survey Net was completed, the Survey Net analysis program was run a final time to generate 24-hour nutrient totals for each recall. These files were then combined and converted to SAS data files. The linking of the menu and recall data, which involved replacing nutrient values (and, in some cases, portion sizes) for reimbursable school meal items, was done with the SAS files. The final data files were then used for analyses of 24-hour food and nutrient intakes as well as meal-specific intakes (at breakfast and lunch).

V. WEIGHTING SNDA-III DATA

Given the complex, multi-stage sample design for SNDA-III, as described in Chapter II, sample weights were required to account for unequal sample selection probabilities and survey nonresponse at all levels. Weights were constructed for four groups at three levels: (1) SFAs; (2) schools; and (3) within schools, students and their parents. The weights at the three levels are not independent—the final weight for the SFA served as the initial weight at the school level, and the final school weight served as the initial weight for the student and parent data.

Several data collection efforts were conducted at each level. For SFAs and schools, one set of weights (constructed to reflect probabilities of selection and nonresponse as described below) was adequate for all the data collected at those levels. However, because some parents of responding students did not complete the parent interview, the weight for analyses using parent interview data had to be adjusted to reflect parent nonresponse.

A. SFA-LEVEL WEIGHTS

The initial weight at the SFA level is the sampling weight from the NSLP Sample Frame, adjusted for eligibility in the NSLP survey.¹ The SFA weight also incorporates (1) the inverse of each SFA's probability of selection into the initial sample of 270; (2) adjustments reflecting allocation into basic and supplemental samples; (3) adjustments for selection into the main sample, including adjustments for alternates that were used in the final sample of SFAs;

¹ The sampling weight, rather than the final weight from the NSLP Sample Frame, was used because the final weight adjusted for nonresponse to the NSLP Sample Frame survey, but SNDA-III sampled both responders and nonresponders to the survey. For more information on sampling and weighting for the NSLP Sample Frame, see Appendix A.

- (4) adjustments for nonresponse not accounted for by the adjustments for use of alternates, and
 (5) poststratification.

Initial Weights. The initial weight for the k th SFA participating in the study was:

$$SFAWGT1_k = SWF1_k * SWF2_k ,$$

where:

$$SWF1_k = WGTSEL(NSLP)_k * ERATE(NSLP) .$$

$WGTSEL(NSLP)_k$ is the SFA's sampling weight for the NSLP Sample Frame, and $ERATE(NSLP)$ is the eligibility rate for the sample frame (only districts that were SFAs and participated in the NSLP and/or SBP were eligible).

$SWF2_k$ adjusts for probability of selection into the SNDA-III sample (and, in some cases, initial nonresponse) and varies according to how the SFA was selected into the sample. As discussed above, selection into SNDA-III took place in two phases: (1) the selection of 268 unique SFAs that were 270 PSU equivalents, and (2) the selection of 135 SFAs to be initially released for data collection. Those not selected into the main sample of 135 had a chance of being selected as a replacement or wildcard. For defining $SWF2_k$, the SFAs were divided into three groups based on how they were selected into the sample:

1. SFAs selected with certainty in the initial sample and again into the main sample
2. SFAs selected with certainty into the initial sample of SFAs, but not selected with certainty into the main sample of 135
3. SFAs selected with PPS (noncertainty selections) into the initial sample of SFAs

For each of three groups, $SWF2_k$ reflects the probability of selection into the sample of 135 from the NSLP Sample Frame:

$$SWF2_k = SWF2a_k * SWF2b_k ,$$

where $SWF2a_k$ is the inverse of the probability of selection into the initial sample of 270 ($P270$), and $SWF2b_k$ reflects the inverse of the probability of selection into the sample of 135 ($P135$).

These terms are defined as follows for the three groups:

- Group 1. For those selected with certainty into the main sample, $SWF2_k = 1$. For these SFAs, $SWF2a_k$ and $SWF2b_k$ are the inverses of the probability of selection at the two phases. Because selection was with certainty at both phases, $SWF2a_k = 1/P270 = 1$, and $SWF2b_k = 1/P135 = 1$.
- Group 2. For the SFAs selected with certainty into the initial sample but not with certainty into the main sample, $SWF2_k$ reflects both the initial probability of selection and the chance of release as a replacement or wildcard.

For this group, $SWF2a_k = 1/P270 = 1$. $SWF2b_k = 0$ if the SFA was never released as part of the sample. For all other cases, $SWF2b_k = 1/P135_k$.

$P135_k = P135A_k + (1-P135A_k)*P135B$, where $P135A_k = 0.864435105 * MOS_k$, and MOS_k is the MOS for SFA k . $P135A_k$ was calculated using the expected number of hits for those that were directly selected for the main sample.² $P135B$ reflects the chance of selection as a replacement or wildcard. Because 2 of these SFAs were selected at random from the 10 that were not selected into the main sample, $P135B = 0.20$.

- Group 3. For the noncertainty selections, $SWF2a_k$ reflects only the chance of being selected into the initial sample:

$SWF2a_k = 1/P270$, the inverse of the probability of selection into the initial sample.

$SWF2b_k$ is a pair adjustment, reflecting the “selection” of 250 noncertainty SFAs into the main sample, as replacements, or as wildcards. $SWF2b$ takes on the value 0, 1, or 2 and adjusts for selection into the sample and for nonresponse within the pair. The values of $SWF2b$ for Group 3 are presented in Table V.1. The sum of $SWF2b$ for a pair will always equal 2; when only one district in a pair was released, $SWF2b$ reflects subsampling within the pair; if both were released, the weight reflects no subsampling within the pair, but if one of the pair was not completed, $SWF2b$ reflects nonresponse within the pair.

² The expected number of hits is from the SAS PROC SURVEYSELECT output. It corresponds to the probability of selection.

TABLE V.1
VALUES OF *SWF2B* FOR NONCERTAINTY SFAs

Within a Pair		
Released	Completed	<i>SWF2b_k</i>
1	0	2 for the released district (based on $1/p$; $p = 1/2$); 0 for the other
1	1	2 for the released district(based on $1/p$; $p = 1/2$); 0 for the other
2	0	1 for each of the districts
2	1	2 for the completed district ($1/p \times 1/rr$ where $p = 1$ and $rr = 1/2$); 0 for the other
2	2	1 for each of the districts

Nonresponse Adjustment. For all three groups, the next step was to form cells to adjust for nonresponse (not already accounted for). For those selected with certainty into the main sample (group 1 above), there was no nonresponse, but there was for groups 2 and 3. The nonresponse weighting cell was the region, and *SWF3_c* is a nonresponse adjustment factor within cell *c*:

$$SWF3_c = \frac{\sum_{released\ SFA \in c} SFAWGT1_k}{\sum_{completed\ SFA \in c} SFAWGT1_k}.$$

The values of *SWF3_c* are shown in Table V.2. These weight factors are the inverse of the weighted response rate for each region. The SFA weight adjusted for nonresponse is: $SFAWGT_NR_k = SFAWGT1_k * SWF3_c$.

TABLE V.2
SFA NONRESPONSE ADJUSTMENT FACTOR

Region	SFA Nonresponse Adjustment Factor (SWF3_c)
Mountain	1.000000
Mid-Atlantic	1.014200
Midwest	1.083209
Northeast	1.205978
Southeast	1.002110
Southwest	1.005488
West	1.250646

Poststratification. The SFA weights were ratio-adjusted (poststratified) so that the weighted total of the completed sample matched that of the NSLP sample that was used as the SNDA sampling frame. The target total was 14,478. The final SFA weight is:

$$SFAWFT_PS_k = SFAWGT_NR_k * RAF_{SFA},$$

where:

$$RAF_{SFA} \text{ (ratio adjustment factor)} = \frac{14,478}{\sum_{k \in \text{Completed SFAs}} SFAWGT_NR_k}.$$

B. SCHOOL-LEVEL WEIGHTS

Initial Weights. The initial weight for school i in stratum j and SFA k is the variable $SFAWGT_PS$ for the SFA to which the school belongs, adjusted for probability of selection within the SFA. The first adjustment factor, $W1_{ijk}$, is the inverse of the probability of the first phase of selection of the school within its SFA. Because schools were selected with PPS, some were large enough to be selected with certainty. Thus:

$W1_{ijk} = 1$ if school was selected with certainty;

$$= \frac{n'_{jk} MOS_{ijk}}{\sum_{ijk=1}^{N'_{jk}} MOS_{ijk}} \text{ otherwise}$$

where:

n'_{jk} is the number of noncertainty selections made in stratum j , SFA k

N'_{jk} is the number of schools available for noncertainty selection with PPS in j and k

MOS_{ijk} is the measure of size/student enrollment for the i th school in stratum j in SFA k

The next factor, $W2_{ijk}$, accounts for subselection into the main and alternate samples. How $W2_{ijk}$ is calculated depends in part on the method used to select the “main” schools from the initial sample. Three such approaches were used:

1. If there were no schools selected with certainty within a stratum in an SFA, then subselection was made with equal probability.
2. Some schools were subselected with certainty (for example, if a school was the only school in the stratum).
3. Some schools were subselected with PPS.

If the subselection was made with equal probability or certainty, $W2_{ijk}$ was computed directly. Where subselection was PPS, it was computed in one or two steps ($W2_{ijk} = W2a_{ijk} * W2b_{ijk}$), based on the number of schools released for data collection and the number of schools completed. The calculation of $W2_{ijk}$ is presented in Table V.3.

The initial school-level weight, before adjustment for nonresponse is:

$$SCHWGT1_{ijk} = SFAWGT_PS_k * W1_{ijk} * W2_{ijk}.$$

Nonresponse Adjustment. The next step was to adjust for nonresponse not accounted for within SFAs. Initial nonresponse adjustment cells were formed based on region and school level

TABLE V.3

SCHOOL-LEVEL WEIGHTS
FORMULAS FOR $W2a_{ijk}$, $W2b_{ijk}$, AND $W2_{ijk}$

Weighting Factors Accounting for Subselection into Main and Alternate Samples				
Subselection Method Within Stratum	Released	Completed	$W2a_{ijk}^a$	$W2b_{ijk}^b$ $W2_{ijk}^c$
Equal probability or certainty	n.a.	>0	--	-- $\frac{n(sel)_{jk}}{n(rel)_{jk}} \left(\frac{n(rel)_{jk} - n(inel)_{jk}}{n(comp)_{jk}} \right)$
PPS	All	>0	--	-- $\frac{n(sel)_{jk}}{n(rel)_{jk}} \left(\frac{n(rel)_{jk} - n(inel)_{jk}}{n(comp)_{jk}} \right)$
PPS	Some	All or none	$\frac{n(main)_{jk} * MOS_{ijk}}{\sum_{ijk \in initial} MOS_{ijk}}$	1 $W2a_{ijk} * W2b_{ijk}$
PPS	Some	Some	$\frac{n(main)_{jk} * MOS_{ijk}}{\sum_{ijk \in initial} MOS_{ijk}}$	$\frac{\sum_{ijk \in Rel} (W1_{ijk} * W2a_{ijk})}{\sum_{ijk \in comp} (W1_{ijk} * W2a_{ijk})}$ $W2a_{ijk} * W2b_{ijk}$

^a $n(main)_{jk}$ are the number of main selections in stratum j and $SFA\ k$. MOS_{ijk} is the measure of size for school i , stratum j , $SFA\ k$.

^b $W1_{jk}$ is defined in the text.

^c $n(sel)_{jk}$, $n(rel)_{jk}$, $n(inel)_{jk}$, and $n(comp)_{jk}$ refer to the number of schools in stratum j , $SFA\ k$ that were selected, released, found ineligible, and completed, respectively.

n.a. = Not applicable.

(elementary, middle, high). Six cells were too small and were combined with larger cells to form the final cells. Within each cell $[c]$, the nonresponse adjustment $W3_c$ is:

$$W3_c = \frac{\sum_{ijk \in (resp, c)} SCHWGT1_{ijk} + \sum_{ijk \in (nonresp, c)} SCHWGT1_{ijk}}{\sum_{ijk \in (resp, c)} SCHWGT1_{ijk}} .$$

The school-level weight, adjusted for nonresponse, is $SCHWGT_NR_{ijk} = SCHWGT1_{ijk} * W3_c$ for completed schools.

Poststratification. Finally, the school weights were ratio-adjusted so that the sum of weights for completed schools was 92,916, the best estimate of the number of schools in SFAs offering the NSLP or the SBP. Thus,

$$RAF_{school} = \frac{92,916}{\sum_{ijk \in complete} SCHWGT_NR_{ijk}}$$

and

$$SCHWGT_PS_{ijk} = SCHWGT_NR_{ijk} * RAF_{school} .$$

C. STUDENT-LEVEL WEIGHTS

The initial weight for student h in school i in stratum j in SFA k is $SCHWGT_PS_{ijk}$, the poststratified school-level weight for that student's school.

Only 98 of 133 SFAs remained in the basic sample where student data were collected, and a ratio adjustment was applied to the weights of the sampled students in the noncertainty SFAs to reflect this fact:

$$W1_{hijk} = \frac{\sum_{k \text{ in noncertainty sample}} SFAWGT_PS_k}{\sum_{k \text{ in noncertainty, basic sample}} SFAWGT_PS_k} .$$

The next adjustment was the inverse of the within-school probability of selection. The within-school selection probability (P_{hijk}) was defined as follows:

$$P_{hijk} = Pa_{hijk} * Pb_{hijk},$$

where:

$Pa_{hijk} = 1/I_{ijk}$ if a predefined systematic sampling was used or if *ad hoc* selection was used and I_{ijk} was recorded

$$= na_{ijk}/M_{ijk} \text{ otherwise}$$

I_{ijk} is the sampling interval used for systematic selection

na_{ijk} is the size of the initial sample in school ijk

M_{ijk} is the number of students in the school estimated from the list

Because most samples were selected from hard-copy lists, M_{ijk} was usually very accurate for small schools but less accurate for large ones. Thus, I_{ijk} was preferred, because it would yield the exact probability of selection:

$$Pb_{hijk} = \frac{nb_{ijk}}{na_{ijk}},$$

where:

nb_{ijk} is the number released (subsamped) from the initial sample for school ijk

na_{ijk} is the initial sample actually selected in school ijk

Then:

$$W2_{hijk} = 1/P_{hijk}$$

$$STUWGT1_{hijk} = SCHWGT_PS_{ijk} * W1_{hijk} * W2_{hijk}.$$

Nonresponse Adjustment. Finally, the student weights were adjusted to account for nonresponse. Nonresponse adjustment cells were formed by crossing the seven regions by school level (elementary, middle, high).

Within each cell, the following response categories were defined:

- Complete (CO)
- Eligible, not complete (EN)
- Unknown if eligible (UNK)
- Ineligible (IE)
- Not released (NR)

For each response cell c :

$$RR_c = \frac{\sum_{h \in CO} STUWGT1_{hjk}}{\sum_{h \in (CO, EN)} STUWGT1_{hjk} + ER_c \left(\sum_{h \in UNK} STUWGT1_{hjk} \right)}$$

$$ER_c = \frac{\sum_{h \in (CO, EN)} STUWGT1_{hjk}}{\sum_{h \in (CO, EN, IE)} STUWGT1_{hjk}}$$

where RR_c is the response rate for cell c , and ER_c is the eligibility rate for cell c .

The final weight for the student data is:

$$STUWGT_NR_{hijk} = STUWGT1_{hijk} * (1/RR_c) .$$

D. PARENT WEIGHTS

Initial Weights. The basic weight for a parent (g), where the parent was selected because his or her child had been sampled ($g = h$), was the final weight for the student ($STUWGT_NR_{hijk}$).

Nonresponse Adjustment. A nonresponse adjustment was computed using logistic regression to predict the likelihood of a parent responding. The dependent variable in the model was whether the parent responded (given that the student responded).

For independent variables, a set of dummy variables was created representing (1) the seven regions, (2) three levels based on SFA-level estimates of percent in poverty, (3) student gender and grade levels, and (4) metropolitan status of the SFA. The model was estimated using SAS with forward stepwise selection of independent variables. The variables selected for the model were:

- Whether student is in grades 1-3
- Whether student is in grades 4-6
- Whether student is in an SFA with the lowest level of poverty
- Whether located in the Southeast region
- Whether located in the Northeast region
- Whether located in the Midwest region
- Whether located in the West region
- Whether in a central city of a metropolitan statistical area (MSA)
- Whether student is female

SAS output included the predicted likelihood of response (*phat*); the adjustment factor was the inverse of the predicted likelihood. Hence, the final parent weight was:

$$\text{PARWGT}_{gijk} = \text{STUWGT_NR}_{gijk} * (1/\text{phat}_{gijk}).$$

REFERENCES

- Shorr, Irwin J. "How to Weigh and Measure Children." In United Nations Department of Technical Co-operation for Development and Statistical Office. *How to Weigh and Measure Children: Assessing the Nutritional Status of Young Children in Household Surveys*, United Nations, New York, 1986; revised 1998.
- Raper, N., Perloff, B., Ingwersen, L., Steinfeldt, L., Anand, J. An overview of USDA's dietary intake data system. *Journal of Food Composition and Analysis*. 2004;17:(3-4):545-555.
- USDA Food and Nutrient Database for Dietary Studies (FNDDS) 1.0. 2004. Beltsville, MD: Agricultural Research Service, Food Surveys Research Group.

APPENDIX A

TECHNICAL DOCUMENTATION OF SAMPLING AND WEIGHTING OF THE NSLP SAMPLE FRAME/PRELIMINARY SURVEY

MEMORANDUM**TO:** Pat McKinney**FROM:** Anne Gordon**DATE:** 11/24/2004
FRAME-074**SUBJECT:** Overview of Sampling and Weighting Memos for the NSLP
Sample Frame Construction Study

The sampling and weighting procedures used in the NSLP Sample Frame Construction Study are documented in four memos, prepared by John Hall and others. This memo lists them in order, and briefly describes what they cover, to help users of the data to find what they need.

The four memos are:

1. FRAME-10—by John Hall, “NSLP Sample Frame Construction,” 1/15/2004. This memo describes the procedures for selecting an initial sample of 2,500 PSUs including 2,942 districts for the study. (This initial sample was then sent to the state agencies to determine which districts were SFAs.) The memo describes the sample frame (the Common Core of Data), stratification, and sample selection procedures. It also discusses how missing data on key size measures and stratification variables (notably district poverty level) were imputed.
2. FRAME-54—by John Hall, “Final Steps in NSLP Sample Frame Construction,” 7/28/2004. This memo describes how the information from state agencies on which members of the original sample were SFAs was processed and analyzed, leading to an eligible sample of 2,621 SFAs. It then describes how 2,310 SFAs were selected from the 2,621, to be used as the sample for the three parts of the survey (the fax-back, phone survey, and self-administered cost and revenue survey).
3. FRAME-55—by John Hall and Yuhong Zheng, “Weighting the NSLP Sampling Frame,” 8/6/2004. This memo describes the procedures for weighting the survey responses, including construction of the base weight and of nonresponse adjustments, using procedures which were applied for all three datasets.
4. FRAME-73—by Anne Gordon and John Hall, “Final Weight Adjustment and Additional Weights,” 11/24/2004. This memo describes a small, constant technical adjustment to all the weights, which only affects estimates of population totals. It then also describes how weights were constructed for (1) the sample who completed

MEMO TO: Pat McKinney
FROM: Anne Gordon
DATE: 11/24/2004
PAGE: 2

all three parts of the survey, and (2) the sample that completed the fax-back and telephone surveys.

cc: Chris Logan, Joanne Guthrie, John Hall, Rhoda Cohen, Jim Ohls

MEMORANDUM**TO:** Pat McKinney**FROM:** John Hall**DATE:** 1/15/2004
FRAME-010**SUBJECT:** National School Lunch Program (NSLP) Sample Frame
Construction

This memo describes the sampling task of the NSLP Sample Frame Construction project being conducted by Mathematica Policy Research, Inc. (MPR) for the Food and Nutrition Service (FNS). The Frame Construction project involves selecting constructing a frame in two phases for a sample for the Integrated Study of School Meal Costs and Outcomes (or similar but less comprehensive studies). In the first phase we will construct a sampling frame and select a sample of school districts for a preliminary survey of public (local) School Food Authorities (SFAs). In the second phase, the preliminary survey will identify public SFAs that participate in the National School Lunch Program (NSLP). These SFAs will comprise the sample frame for the latter survey(s). MPR's proposal was based on 2,079 completed preliminary survey interviews with SFAs, which would require an initial sample release of 2,310 assuming response of 90 percent. The current plan calls for selecting an initial sample of 2,500 primary sampling units (PSUs) each comprising one or more districts and contacting states to identify SFAs among the sampled districts. The initial release would then be sampled from those identified as SFAs. The proposal also suggested an allocation of the overall sample to each FNS region that would yield the precision required for regional estimates. Below we discuss task objectives, constructing the sampling frame, stratification, sample size and allocation, choosing a method of sample selection, selecting the sample and plans for sample release and management during the preliminary survey.

Objectives. The objective of the sampling task is to select a national sample of public SFAs participating in the NSLP and/or School Breakfast Program [SBP] for the preliminary survey. This sample will also serve as a sampling frame for the Integrated Study of School Meal Costs and Outcomes, or for other future nationally representative surveys if funding for the integrated study is not available. This sampling frame is needed because a listing of all public SFAs in the U.S. that participate in NSLP and/or SBP with information such as enrollment size, poverty level, and menu planning system is not available. However, a sampling frame for the selection of a nationally representative sample of school districts is available. Therefore, the strategy is first to select a large sample of school districts and collect the required information on the SFAs

MEMO TO: Pat McKinney
FROM: John Hall
DATE: 1/15/2004
PAGE: 2

from them. Then it will be possible to construct a sampling frame of SFAs based on the sample of school districts.

Sampling Frame. The sampling frame for the selection of the sample of school districts for the preliminary survey was constructed from the 2001–2002 Common Core of Data (CCD), maintained by the National Center for Education Statistics (NCES) of the U.S. Department of Education. The CCD provides a listing of all public school districts and all public elementary and secondary schools in the United States, with data that are comparable across all states. The CCD has current enrollment data for most districts¹ and also contains other district characteristics that can be used in later stages of sampling or future analyses, such as grade range and number of schools. The CCD also contains information that was used in eliminating ineligible districts.

The project requires that the sample be stratified by FNS region and by level of poverty. Information on the poverty level of students in each school district is not on the CCD, but poverty estimates are found on a file produced by the U.S. Census Bureau for all school districts. These estimates are derived from the U.S. Census Bureau's school district mapping project, using a list of all school districts and their boundaries obtained from each state's Department of Education. The information from the Census Bureau file was merged with information on the CCD to enable the required stratification of school districts.

The sampling frame for the preliminary survey was constructed by modifying the CCD: adding the Census data, eliminating ineligible districts and forming multi-district primary sampling units (PSUs) where warranted. These steps were necessary because: (1) as stated above the CCD does not contain poverty data; (2) there were districts on the frame that were clearly ineligible for the survey; and (3) some districts were linked in a way that indicated that they might be served by a common SFA. In eliminating ineligibles and linking districts we relied on the variable TYPE01 on the CCD. TYPE01 is the NCES code for type of agency. It has 7 values:

- 1 indicates a local school district that is not a component of a supervisory union
- 2 is a local school district component of a supervisory union sharing a superintendent and administrative services with other local school districts. Multiple districts are assigned to each union.

¹Enrollment data are needed because we proposed to consider selecting the sample using probability proportional to size (PPS) methods using as a measure of size (MOS) the number of students or some transformation of that number.

MEMO TO: Pat McKinney
FROM: John Hall
DATE: 1/15/2004
PAGE: 3

- 3 denotes a supervisory union administrative center, or a county superintendent serving the same purpose. There is only one administrative center per union.
- 4 indicates a regional education services agency, or a county superintendent serving the same purpose.
- 5 indicates a state-operated institution charged, at least in part, with providing elementary and/or secondary instruction or services to a special need population. This groups includes schools for deaf and blind, correctional facilities, departments of health and mental health and entities with titles like Math and Science Academy or School for the Arts.
- 6 includes Federally-operated institutions charged, at least in part, with providing elementary and/or secondary instruction or services to a special need population. This group includes schools on military bases in the US, others on Indian Reservations, some schools overseas, a few Job Corps centers.
- 7 includes all districts not classified as 1-6; many if not all charter school districts are in this group.

Before ineligible districts were eliminated, the file from which we constructed the frame included records for 17,321 districts, 17,276 from the CCD and 45 that were found only on the Census file. In consultation with FNS, we specified 5 criteria by which a district would be considered ineligible. The 967 ineligible districts included those:

- Found only on the Census file and not on the Core of Common Data (CCD)
- Located outside the 50 United States plus the District of Columbia
- That are state or federally operated schools (TYPE01 is 5 or 6).
- Where CCD data indicate that the district has ceased to operate
- That report no schools or students and cannot be connected to any other eligible district, to an operating school or to students on the school level CCD file.²

²Under this criterion (connection) a district other than a TYPE01 = 2 or 3, is considered ineligible if: (a) the district level report (on the CCD) does not indicate any schools or K-12 students, and; either (b) the district does not have the same NCES identifier, or Local Education Agency ID (LEAID) as any school in the school level file, or (c) any school having the district's LEAID is closed or has no students. For a TYPE01 2 or 3 district, a district is ineligible if it

MEMO TO: Pat McKinney
FROM: John Hall
DATE: 1/15/2004
PAGE: 4

Many of those classified as ineligible met multiple criteria. Even after eliminating these districts we expect there will be some that are found to be ineligible during the preliminary survey. For example, not every district will be an SFA and some may have ceased to exist by the time they are contacted.

After we eliminated ineligible districts from the frame we examined the added poverty measure from the Census file. Where needed, we imputed poverty level or numbers of enrolled students (used as measure of size -- MOS) where either was missing. Finally, we combined districts that are linked through a supervisory union (TYPE01 is 2 or 3).

After the CCD and Census data were merged, 1,896 of the 16,354 eligible school districts were missing poverty information because the districts were not on the Census list (for example, districts that had been created since 2000). We imputed poverty information using two procedures:

- A model was fit predicting the percent in poverty using data on percentage of students certified for free meals in a school district, degree of urbanicity and TYPE01 (this model provided predicted poverty level for 821 cases)
- For districts missing the both the poverty data and the variables in the model, data from Genesys³ were used based on the telephone number for the district listed on the CCD. Specifically, if the percent of households in the telephone exchange matching that of the district with annual income less than \$25,000 was predicted to be 30 percent or more, we imputed that the district was high poverty. Genesys data were used for 1075 districts.

If the numbers of enrolled students were missing from a district's records, we attempted to directly add the numbers of students in schools linked to the district by a common LEAID. Where this was not possible, we imputed within cells defined by TYPE01 and region provided there were at least 30 districts in the cell. For each cell we computed the average ratio of students per school for districts reporting both. For districts missing the number of students, we multiplied the number of schools by the cell average ratio to obtain the imputed estimate.

meets the criteria for other type districts and in addition does not link to any other district (through its UnionID) where that other district is eligible.

³MPR uses software licensed from Genesys Sampling Systems (Marketing Systems Group) to generate Random Digit Dialing telephone samples. Auxiliary data on predicted household income, ethnicity and other items are provided at the telephone exchange level.

MEMO TO: Pat McKinney
FROM: John Hall
DATE: 1/15/2004
PAGE: 5

The next step before forming strata was to form primary sampling units (PSUs). We decided to form PSUs to deal with the issue of multiple districts in a single supervisory union (TYPE01 is 2 or 3 and all have the same UNIONID). It is unknown if each district in such a union would be serve as its own SFA or if the supervisory district would serve as the SFA for all, or if some other arrangement might be in used. Since the supervisory unions raise the possibility of multiple districts being served by a single SFA, a logical approach is to merge all districts in the same union into a single PSU for sampling purposes. If multiple SFAs are found within such a PSU, we can subsample them to control the sample size for the preliminary survey. All districts not in a supervisory union were assigned to their own PSUs. Tables 1 and 2 show the distribution of the 16,354 eligible districts and the 14,862 PSUs formed.

Stratification and Allocation. As stated above the target sample size for the preliminary survey is 2,079 completed SFA interviews, subject to minimum precision requirements⁴ for each region. Because the proposal assumed gains from a finite population correction within each region the effective sample size for regional estimates would be greater than that of the nominal sample size. The proposal suggested an initial sample of 2,310 would be needed, assuming a response rate of 90 percent and no ineligibility.

The stratification and sample allocation described below deviate from that in the proposal in two ways: the initial sample is larger (although the final sample remains the same size), and the strata are defined in a slightly different way. In addition, the selection was done PPS rather than without regard to district size. (While the proposal discussed using PPS sampling the focus was on non-PPS methods.) The factors influencing these changes are:

- First, we now assume that there may be a non-trivial portion of sampled districts that are ineligible (i.e., ineligible districts would be those not operating or not having/being an SFA).
- Second, we are selecting PSUs rather than districts and some PSUs comprise multiple districts.

⁴The proposal spoke in terms of a region level estimate of + or - \$0.10 around a mean of \$2.00 for cost per meal served, assuming a standard deviation of \$0.90. One way to characterize the minimum precision requirements is to use the term effective sample size, which is defined as actual sample size divided by the design effect. Thus, the effective sample size is the size of a simple random sample that would be required to achieve the same precision as the actual design has. In the example used in MPR's proposal a simple random sample of 311 cases would be required in each stratum. The proposal assumed a design effect of less than 1.0 for many strata, based on the finite population correction factor (fpc).

MEMO TO: Pat McKinney
FROM: John Hall
DATE: 1/15/2004
PAGE: 6

- Third, (as stated in the proposal) we examined the issue of sampling districts using a PPS method; we concluded that there were more advantages to selecting the sample PPS with the measure of the size being the square root of the numbers of students.
- Fourth, the strata boundaries were modified so that there will be a certainty stratum for the PPS selection and so that the districts where TYPE01 = 7 (mostly charter schools) will be proportionately represented.

To examine the trade-offs of using some form of PPS sampling, we selected three samples and compared them on expected precision for estimates about districts (X per cent of districts employ a certain meal planning system) and those about students (Y per cent of students are found in districts that meet a certain dietary standard). Each of these two types of estimates is important for the study and the design that is optimal for one type is unlikely to be optimal for the other.

The three samples included one selected with equal probability (no districts were oversampled), one selected PPS with the MOS being the estimated number of enrolled students and one selected PPS with the MOS being the square root of the numbers of enrolled students. Each sample included 2,500 PSUs and analysis was conducted with the whole data set, although the actual survey will have fewer cases. This will not affect conclusions because what we are looking for is the relative efficiency of the 3 approaches. In the past, other exercises and projects have shown that the PPS with the MOS being the square root of the numbers of enrolled students is a good compromise solution.

Table 3 presents the results of our analysis. Each district was assigned values for 5 variables. The variables were generated randomly rather than taken from some other survey:

- 2 continuous variables, one from the uniform distribution, another from the normal distribution
- 3 binary (0,1) variables, one random with an expected mean of 0.15, another correlated with poverty level with expected mean of 0.15 and a third correlated with size and mean of 0.67.

Analyses were conducted using weighted data. For the district estimates the weight was computed as the inverse of the probability of selection. For the student estimates the weight was the inverse of the probability of selection multiplied by the numbers of enrolled students. The results shown in the table indicate (as expected) that, compared to the no-oversampling strategy, the PPS with MOS being the number of students doubled the standard errors of district estimates and reduced the standard errors for student estimates by a factor of $\frac{1}{4}$ or less. The results where the MOS was the square root show district level standard errors increased by a factor of about

MEMO TO: Pat McKinney
FROM: John Hall
DATE: 1/15/2004
PAGE: 7

1.3 (compared to no oversampling) and the standard errors of student estimates nearly decreased as much as the design using enrolled students as the MOS.

Based on this analysis, we recommend selecting the sample using the square root of enrolled students as the MOS. We have selected a sample of 2,500 PSUs, which will yield more than 2,500 districts, but as discussed below, we will contact only enough presumed SFAs to yield the target of 2,079 completes for the initial survey. Table 4 shows the distribution of the sample by strata. The main stratifying variables are region and poverty level (as defined in the proposal). In addition, employing a PPS selection method requires having a certainty stratum. Finally, we recommend placing the TYPE01=7 districts (mostly charter schools) in separate strata, because they may behave quite differently than other districts. We selected a minimum of 330 PSUs per region. This should allow adequate precision for regional estimates although this precision should greatly exceed the RFP's precision standards for estimates about students but may fall short in terms of estimates about districts. Tables 5 and 6 show the distribution by FNS region of the sampled PSUs and the districts they include.

Sample Release and Management. The sample of 2500 PSUs includes 2,942 districts. The names of the sampled districts will be sent to states that will be asked to review them, confirm whether the sampled districts are in operation and indicate which have or are SFAs. We expect that there will be some instances where multiple SFAs are found within a multi-district PSU. In such cases, we will probably subsample one SFA for potential use in the preliminary survey. The exact strategy, however will depend both on the extent of PSUs with multiple SFAs and the prevalence of SFAs among the sampled districts. For example, we probably would not subsample if there are few PSUs with multiple SFAs, no PSUs with more than two or three SFAs, and the review by the states yields only 2,400 to 2,500 SFAs total.

After the review by the states, we plan to release 2,310 SFAs for the preliminary survey. If the review by the states yields more than 2,310 SFAs, we will select a random subsample for release and hold the rest in reserve in case response to the initial survey is less than expected (due to non-cooperation or unexpected ineligibility).

MEMO TO: Pat McKinney
FROM: John Hall
DATE: 1/15/2004
PAGE: 8

cc: Rhoda Cohen, Barbara Carlson, Laura Kalb, Eric Zeidman

TABLE 1

DISTRIBUITION OF DISTRICTS BY REGION AND POVERTY LEVEL AFTER
INELIGIBLES WERE REMOVED

	High Poverty	Low Poverty	All
Region			
Mid-Atlantic	91	1,516	1,607
Midwest	260	3,546	3,806
Mountain	235	2,766	3,001
Northeast	108	2,208	2,316
Southeast	216	943	1,159
Southwest	380	1,874	2,254
Western	369	1,842	2,211
Total	1,659	14,695	16,354

TABLE 2

DISTRIBTION OF PSUS ON THE SAMPLING FRAME BY REGION AND POVERTY
LEVEL

	High Poverty	Low Poverty	All
Region			
Mid-Atlantic	91	1,516	1,607
Midwest	260	3,207	3,467
Mountain	232	2,498	2,730
Northeast	108	1,338	1,446
Southeast	216	943	1,159
Southwest	380	1,874	2,254
Western	368	1,831	2,199
Total	1,655	13,207	14,862

TABLE 3

STANDARD ERRORS (SE) OF FIVE VARIABLES FOR THREE TYPES OF SAMPLE SELECTION
STRATEGIES WEIGHTED ESTIMATES ABOUT DISTRICTS/SFAs AND ABOUT STUDENTS

Variables	Sample Selection					
	A. No Oversampling		B. PPS with MOS = Square Root of Students		C. PPS with MOS = Number of Students	
	District Weight	Student Weight	District Weight	Student Weight	District Weight	Student Weight
Continuous Variables						
Uniform Distribution (0,1; mean = 0.50)						
Ratio (Percent) of this SE to No Oversampling Design SE	0.005282 100.00 %	0.015081 100.00 %	0.006709 127.02 %	0.004207 27.90 %	0.010868 205.76 %	0.003362 22.29 %
Normal Distribution (0,1; mean = 0.50)						
Ratio (Percent) of this SE to No Oversampling Design SE	0.002592 100.00 %	0.009262 100.00 %	0.003300 127.31 %	0.002123 22.92 %	0.005604 216.20 %	0.001641 17.72 %
Categorical Variables						
Binomial (mean = 0.15)						
Ratio (Percent) of this SE to No Oversampling Design SE	0.6417 100.00 %	2.0988 100.00 %	0.8493 132.35 %	0.5015 23.89 %	1.4502 225.99 %	0.4271 20.35 %
Binomial, correlated with poverty (mean = 0.15)						
Ratio (Percent) of this SE to No Oversampling Design SE	0.6480 100.00 %	1.8843 100.00 %	0.8322 128.43 %	0.5327 28.27 %	1.1427 176.34 %	0.4015 21.31 %
Binomial, correlated with size (mean = 0.67)						
Ratio (Percent) of this SE to No Oversampling Design SE	0.8706 100.00 %	3.6732 100.00 %	1.1171 128.31 %	0.6236 16.98 %	1.7927 205.92 %	0.5504 14.98 %

TABLE 4
PSU'S SELECTED BY STRATUM

Stratum	PSU'S
Certainty	73
Non Certainty Strata	
Mid-Atlantic Low Pov	294
Mid-Atlantic Hi Pov	10
Midwest Low Pov	451
Midwest Hi Pov	11
Mountain Low Pov	307
Mountain Hi Pov	18
Northeast Low Pov	302
Northeast Hi Pov	14
Southeast Low Pov	264
Southeast Hi Pov	32
Southwest Low Pov	260
Southwest Hi Pov	47
Western Low Pov	312
Western Hi Pov	35
Type7 Low Pov	47
Type7 Hi Pov	23
Total Non Certainty	2427
Grand Total	2,500

TABLE 5

PSU'S SELECTED BY REGION (ALL STRATA)

Region	PSU's
Mid-Atlantic	331
Midwest	484
Mountain	330
Northeast	329
Southeast	330
Southwest	331
Western	365
Total	2,500

TABLE 6
DISTRICTS IN SAMPLED PSU'S BY REGION

Region	Districts
Mid-Atlantic	331
Midwest	591
Mountain	399
Northeast	587
Southeast	330
Southwest	331
Western	373
Total	2,942

MEMORANDUM

TO: Pat McKinney

FROM: John Hall

DATE: 7/28/2004
FRAME - 054

SUBJECT: Final Steps in Constructing NSLP Sampling Frame

The purpose of this memo is to describe the final steps of selecting the sample of School Food Authorities (SFA) for the National School Lunch Program (NSLP) Sample Frame. An earlier memo from John Hall to Pat McKinney dated January 15, 2004 (MPR memo # FRAME-010) describes the process of constructing a sampling frame of public school districts and selecting an initial sample of 2,500 primary sampling units (PSUs) that included 2,942 districts (some PSUs included multiple districts).

State officials were contacted about the status of the 2,942 districts and 2,621 were identified as SFAs. The results of these contacts are summarized in Table 1.

From these 2,621 SFAs we selected a sample of 2,310 for the SFA Characteristics survey. The selection procedures varied based on two dimensions: (1) whether the SFA's PSU was selected with certainty and (2) whether the PSU was multi-district. In PSUs where one or more SFAs were identified:

- SFAs were identified in all 71 single-district PSUs in the certainty stratum; all of these were retained
- In each of the two multi-district PSUs that were selected with certainty, more than one SFA was identified; in each of these PSUs we selected one SFA with probability proportional to size (PPS); then, both of these SFAs were retained
- From the 2,215 SFAs identified in the non-certainty strata as part of single-district PSUs, we selected a random sample of 2,151¹

¹ Selection was done in three steps. We had received information about SFAs in all but two states, and needed to proceed. First, we selected a sample of 2,025 from the 2,113 in those states from which we had information. Second, we received information from the other two states, and these states reported 102 SFAs in single district, noncertainty PSUs, from which we selected 99. Third, to reach our target and have equal probabilities of selection we selected an additional 27 from the 88 that had not been selected in the first step.

MEMO TO: ???
FROM: John Hall
DATE: 7/14/2004
PAGE: 2

- SFAs were identified in 89 multi-district PSUs that were selected in non-certainty strata; among these, 55 PSUs had multiple SFAs and 34 had one SFA. We first selected one SFA with PPS from each of the 55 with more than one SFA; we pooled these 55 with the 34 SFAs from single SFA PSUs and from these 89, selected a random sample of 86.

Table 2 shows the distribution of the 2,310 SFAs selected for screening by stratum.

TABLE 1
RESULTS OF CONTACTS WITH STATES REGARDING
DISTRICTS IN INITIAL SAMPLE

Sample Released			SFAs Identified	
PSU Type	PSUs	Districts	PSUs With 1 or More SFAs	Total SFAs
A. Certainty stratum, single district PSU	71	71	71	71
B. Certainty stratum, multi-district PSU	2	51	2	33
C. Other strata single district PSU	2,334	2,334	2,215	2,215
D. Other strata multi-district PSU	93	486	89	302
Total	2,500	2,942	2,377	2,621

TABLE 2

SFAs SELECTED FOR SFA CHARACTERISTICS SURVEY

Stratum	Total
Certainty	73
Non Certainty Strata	
Mid-Atlantic Low Pov	282
Mid-Atlantic Hi Pov	5
Midwest Low Pov	427
Midwest Hi Pov	10
Mountain Low Pov	276
Mountain Hi Pov	12
Northeast Low Pov	282
Northeast Hi Pov	11
Southeast Low Pov	252
Southeast Hi Pov	29
Southwest Low Pov	252
Southwest Hi Pov	41
Western Low Pov	284
Western Hi Pov	34
Type7 Low Pov	25
Type7 Hi Pov	15
Total Non Certainty	2,237
Grand Total	2,310

MEMORANDUM

TO: Pat McKinney

FROM: John Hall and Yuhong Zheng

DATE: 8/6/2004
FRAME - 055

SUBJECT: Weighting the NSLP Sampling Frame

This memo describes the procedures for weighting the National School Lunch Program (NSLP) Sample Frame of School Food Authorities (SFAs) and the SFA Characteristics survey. Sample selection procedures for this project are discussed in two memos: FRAME-010, dated January 15, 2004 and FRAME 054 dated July 15, 2004.

As described in those memos, MPR used the Core of Common Data (CCD) file of public school districts published by the National Center for Education Statistics to construct an initial sampling frame. The frame included groups of districts referred to as primary sampling units (PSUs). Some PSUs included only one, while others had more districts. We then selected an initial sample of 2,500 PSUs that included 2,942 districts. Next, we contacted State officials to determine which of the 2,942 districts were School Food Authorities (SFAs). This process identified 2,621 SFAs from which we selected a sample of 2,310 for the SFA Characteristics Survey.

The weights described in this memo serve two purposes: (1) they can be used for analysis of a preliminary survey, the SFA Characteristics Survey conducted with the sample of presumed SFAs; (2) they can serve as initial weights when this sample is used as a sampling frame for a subsequent surveys. Alternative weights that could be used if alternative sampling frames are desired are described briefly in Attachment I.

The weighted totals of survey responses should reflect the distribution that would be obtained from observing the entire population. The weights adjust for differences in probability of selection associated with the sample design; they also adjust for non-response and ineligibility identified in the SFA Characteristics Survey. As such the weighted total should reflect the population of local public SFAs in the 50 states and the District of Columbia.

The weights are the product of three factors. The first of these factors reflects probabilities of selection for the initial sample of 2,500 PSUs, the second reflects the probabilities of selection for the subsample of 2,310 SFAs for the SFA Characteristics Survey, and the third reflects differences in response rates:

MEMO TO: Pat McKinney
FROM: John Hall and Yuhong Zheng
DATE: 8/6/2004
PAGE: 2

- The first of these factors is the inverse of each Primary Sampling Unit's (PSU's) probability of selection for the sample of 2,500 PSUs described in FRAME 010. For PSUs in the noncertainty strata, this weight was calculated as:

$$W1_{hi} = \frac{1}{p1_{hi}} = \left[\frac{a_h M_{hi}}{\sum_{i=1}^{A_h} M_{hi}} \right]^{-1}$$

Where the subscripts refer to stratum (h), and PSU (i), respectively, the factor in brackets is the probability of PSU selection for the sample in phase 1.

$W1_{hi}$ is the initial weight for the i th PSU in stratum h ,

$p1_{hi}$ is the probability of selection of the i th PSU in stratum h ,

M_{hi} is the measure of size for the i th PSU in stratum h , (the square root of the estimated number of students in the PSU)

a_h is the number of PSUs selected in stratum h .

A_h is the number of all PSUs in stratum h

For cases where the PSU was selected with certainty ($p1_{hi} = 1.0$):

$$W1_{hi} = \frac{1}{p1_{hi}} = 1$$

- The second factor is the inverse of each public local SFA's probability of selection for the SFA Characteristics survey. (See memo FRAME 054.)

For selecting this sample, procedures differed for 4 types of PSUs defined on whether PSUs were selected with certainty and whether PSUs contained more than one district.

MEMO TO: Pat McKinney
FROM: John Hall and Yuhong Zheng
DATE: 8/6/2004
PAGE: 3

1. For those SFAs belonging to single-district SFAs in noncertainty strata, there were three steps in selecting the sample.¹ The second factor for this group was calculated as :

$$W2_j = 1/P2_j$$

For all states other than Oklahoma and North Carolina

$$P2_j = \frac{nsel_1}{navail_1} + \left(1 - \frac{nsel_1}{navail_1}\right) \left(\frac{nsel_2}{navail_2}\right) = \frac{2025}{2113} + \left(1 - \frac{2025}{2113}\right) * \left(\frac{27}{88}\right) \text{ for each SFA } j.$$

where

$nsel_1$ is the number of SFAs sample in the first step

$nsel_2$ is the number of SFAs sample in the second step

$navail_1$ is the number of SFAs from which $nsel_1$ were selected

$navail_2$ is the number of SFAs from which $nsel_2$ were selected

For Oklahoma and North Carolina

$$P2_j = \frac{nsel_1}{navail_1} = \frac{99}{102}$$

2. For those SFAs belonging to single-district PSUs in the certainty stratum, the second factor was calculated as:

$$W2_j = 1/P2_j = 1, \text{ since}$$

¹Selection was done in three steps. We had received information about SFAs in all but two states, and needed to proceed. First, we selected a sample of 2,025 from the 2,113 in those states from which we had information. Second, we received information from the other two states, and these states reported 102 SFAs in single district, noncertainty PSUs, from which we selected 99. Third, to reach our target and have equal probabilities of selection we selected an additional 27 from the 88 that had not been selected in the first step.

MEMO TO: Pat McKinney
FROM: John Hall and Yuhong Zheng
DATE: 8/6/2004
PAGE: 4

$P2_j = 1$ for each SFA $_j$.

3. For those SFAs belonging to multi-district PSUs in noncertainty strata in the second factor was calculated as:

$$W2_j = 1/P2_j$$

$$P2_{ij} = \left(\frac{M_{ij}}{\sum_{j=1}^{N_i} M_{ij}} \right) \cdot \frac{86}{89}$$

where the subscripts refer to PSU (i) and SFAs (j), respectively.

M_{ij} is the measure of size for SFA $_j$ in PSU $_i$,

N_i is the number of SFAs in the PSU

(As explained in FRAME 054, we first selected one SFA from PSUs with multiple SFAs and then randomly selected 86 of 89.)

4. For those SFAs belonging to multi-district in the certainty stratum, the second factor was calculated as:

$$W2_j = 1/P2_{ij}$$

$$P2_{ij} = \left(\frac{M_{ij}}{\sum_{j=1}^{N_i} M_{ij}} \right)$$

where the subscripts refer to PSU (i) and SFAs (j), respectively.

M_{ij} is the measure of size for each SFA $_j$ in PSU $_i$,

MEMO TO: Pat McKinney
FROM: John Hall and Yuhong Zheng
DATE: 8/6/2004
PAGE: 5

- The third component compensated for nonresponse and the ineligibility of the sampled SFAs. The response rate was calculated as:

$$RR_c = \frac{ncmp_c}{nelg_c + \frac{nelg_c}{(nelg_c + ninelg_c)} * nnd_c}$$

$ncmp_c$ is the number of the completed interviews per response adjustment cell²stratum; $nelg_c$ is the number of the eligible SFAs per cell; $ninelg_c$ is the number of the ineligible SFAs per cell; nnd_c is the number of eligibility-not-determined SFAs per cell.

Hence, the final analysis weights were the product of the initial weight, second weight and the inverse of the response rate:

$$WFNL_{hj} = W1_{hi} * W2_j * \frac{1}{RR_c}$$

where the terms W_{hi1} , W_{hj2} and RR_c are defined above.

²We used the sampling stratum as the response adjustment cell, except in strata with fewer than 30 cases. In which case we.....?

ATTACHMENT I

ALTERNATIVE WEIGHTS

Rather than select a sample of SFAs from the 2,150 that responded to the SFA characteristics survey, one might wish to select a sample from the entire sample of 2,310 or include the 2,150 plus those that did not respond, except for those determined to be ineligible. For either of these situations the appropriate weight would be:

$$W_{TALT} = W_{1_{hi}} * W_{2_j}$$

which would adjust for differences in probabilities of selection but not response propensity.

MEMORANDUM**TO:** Pat McKinney**FROM:** Anne Gordon and John Hall**DATE:** 11/24/2004
FRAME-073**SUBJECT:** Final Weight Adjustment and Additional Weights

Since the submission of the previous weighting memo (FRAME-055), we have done additional work on the weights for the data from the NSLP Sample Frame Development study in two areas: (1) we made a small multiplicative adjustment to all of the weights to ensure the sum of the weights adds up to the relevant population to the greatest extent possible; and (2) we created some additional sets of weights for samples that responded to several instruments, which required computing new nonresponse adjustments for these smaller samples. This memo documents both activities—it should be seen as an add-on to memo FRAME-055, rather than a stand-alone piece.

1. Multiplicative Adjustment

This adjustment involves multiplying all of the weights by a constant to ensure that the weights add up to our best estimate of the relevant population—the population of SFAs—and that we fully account for ineligibility in the frame at each stage. K. P. Srinath at Abt Associates raised the issue and we were pleased to respond. Note that this adjustment is the same for all sample members—thus, it will not change any estimates of proportions or means, only estimates of population totals.

First, some background is useful. We initially selected a sample of PSUs (primary sampling units) from the U.S. Department of Education's Common Core of Data (CCD). The PSUs were districts except where the CCD indicated that districts were joined in a supervisory union, in which case all the districts in the union were included in one PSU. Our initial sample comprised 2500 PSUs; these PSUs included 2,942 districts. Some districts on the CCD were excluded from the initial sample because the CCD indicated that they were not operating or served special populations, such as prisoners in correctional facilities or other institutionalized populations.

Next, we sent the list to each state child nutrition agency, to check if the districts or unions selected were School Food Authorities (SFAs), or joined with other districts in an SFA. When we checked with state agencies about the status of the 2,942, we assigned a status code based on the states' responses. These were:

MEMO TO: Pat McKinney
FROM: Anne Gordon and John Hall
DATE: 11/22/2004
PAGE: 2

- 0 District does not offer the school meal programs and is not an SFA or part of one.
- 1 District is its own SFA and the SFA does not include other districts
- 2 District is part of a multi-district SFA, but is not the lead district
- 3 This is the designated supervisory union or lead SFA for a multi-district entity
- 4 This is another administrative unit but not an SFA or supervisory union.
- 5 This is not a functioning school district.

When we selected the fax-back sample, we included only those with codes 1 (n= 2,542) and 3 (n=79) in our frame for the final stage of selection. The schools/students in those districts with code 2 (n=83) will be represented by their lead SFA. Those with code 0 (n=218) are school districts but not SFAs—they do not participate in the NSLP or SBP. Those with code 4 (n=1) are not really school districts but some other type of educational entity. Thus we had a sample of 2,621 eligible SFAs (as best we could tell). However, we only wished to have 2,310 SFAs as the initial sample for the study, so we selected a subsample of 2,310 from the 2,621 (see memo FRAME-054).

Thus, the goals of the adjustment are the following:

- The sum of the weights for the 2,310 should equal the sum of the weights for the $2,542+79 = 2,621$ that comprised the frame for selecting the 2,310.
- The sum of the weights for the sample of 2,500 PSUs (which contained the 2,942 districts) should equal the number of PSUs in the frame from which that sample was selected.

The specification for the adjustment is as follows:

$ADJ1 = \# \text{ of PSUs on the CCD frame} / \text{sum of the weights for the 2,500 PSUs}$
 $ADJ2 = \text{sum of the weights for the } (2,542+79 = 2,621) / \text{sum of the weights for the 2,310}$
 $ADJNEW = ADJ1 * ADJ2$

The revised final weight for the completes for the faxback (n=2,150) is equal to the current final weight*ADJNEW. This should account for ineligibility at each stage. Similarly, the revised final weights for the telephone and cost surveys are also the previous weights multiplied by ADJNEW. The value of ADJNEW was 1.0954213.

2. Additional Weights for Samples of Respondents to Multiple Instruments

Chris Logan at Abt Associates noted that, for his analysis, he will likely use the following two samples: (1) respondents who completed both the fax-back and telephone surveys

MEMO TO: Pat McKinney
FROM: Anne Gordon and John Hall
DATE: 11/22/2004
PAGE: 3

(fax/phone sample, for short); and (2) respondents who completed the fax-back, telephone, and cost surveys (all three survey sample). He asked if we can prepare weights with the appropriate nonresponse adjustments for these samples. We have prepared such weights—the first weight is on the telephone survey file (**fnlwgt_faxtele**), and the second weight is on the cost survey file (**fnlwgt_all**). Construction of these weights involved setting the weight to zero for all SFAs that did not meet the criteria for sample inclusion, and recalculating the nonresponse adjustment for SFAs that do meet the criteria, using the same approach as in preparing previous weights.

Although the method is the same, one concern is that these groups represent smaller and smaller samples. Thus, there is more dispersion in the weights—in particular, some of the weights for the “all 3 surveys” sample are quite large. We considered whether trimming the weights (reducing the size of the largest weights, and slightly increasing the others, using recognized statistical methods) would be appropriate. To gauge how useful this would be, we trimmed the weights on a preliminary basis, but found the reduction in the estimated design effect was small, so we have not included the trimmed weights. We can provide the trimmed weights if others wish. The estimated design effects (due to weighting only) are: (1) for the sample of respondents to all 3 surveys--1.78 not trimmed, 1.64 trimmed; and (2) for the sample of respondents to both the telephone and fax-back, 1.62 not trimmed, 1.59 trimmed.

cc: Rhoda Cohen, Jim Ohls

