Assessing the Food Security and Diet Quality Impacts of FNS Program Participation: Final Report

Contract 53-3198-2-026, Task No. 6

December 30, 2005

Contents

Contents

Contents

Contents

Contents

Contents

Acknowledgements

Acknowledgements   **Abt Associates Inc.**

Executive Summary

The Food and Nutrition Service (FNS) operates over a dozen nutrition assistance programs for children and low-income individuals. The programs aim to improve participants' access to nutritious food and promote healthier eating habits and lifestyles, by providing specific foods or food purchasing power, and by offering nutrition education. The Food Stamp Program (FSP), which serves nearly all types of low-income households, is the largest of these programs.

As the steward of public funds, FNS is responsible for assessing the extent to which the FSP achieves its stated missions. This involves evaluation activities of many types, including a periodic assessment using the Performance Assessment Rating Tool (PART) procedure specified by the Office of Management and Budget.

Over the last three decades, more than 100 studies have examined FSP impacts on program participants (Burstein *et al.*, 2004a). The studies have produced convincing evidence that the program leads participants to increase their expenditures on food. Findings regarding FSP impacts on individuals' diets and households' food security have been much less convincing, with weaker and inconsistent results. Consequently, a PART review led to the recommendation that USDA "develop studies to demonstrate the impact of program participation on hunger and dietary status."

The "gold standard" method for evaluating programs like the FSP would involve randomly assigning a sample of eligible households to two groups: a treatment group that would receive program benefits and a control group that would not. Such an evaluation would not be legal or ethical for the Food Stamp Program, however, because the authorizing legislation makes program benefits available to all eligible households that apply.

To explore other options for assessing impacts, FNS awarded a contract[1] to Abt Associates Inc. to consider the potential for using nonexperimental (survey-based) research designs. The objective of such research would be to provide FNS with new information on:

- Experiences and satisfaction of participants in FNS programs, and

- Impacts of program participation on food security, diet quality, and other indicators of household well-being.

Initial Steps

In the first phase, the project team sketched out five different potential study concepts varying in scope and complexity. The designs ranged from a single-wave survey of FSP participants and non-participants to a four-wave study that would assess FSP impacts and provide information about participant experiences in FSP, WIC, and the school meals programs. The project team presented the designs to an assembled group of outside experts in evaluation methodology, nutrition research, and

---

[1] *Assessing the Food Security and Diet Quality Impacts of FNS Program Participation: Survey Design Options*, Task Order 43-3198-4-3811 under Contract 53-3198-2-026.

evaluation of FNS programs (the Technical Working Group, or TWG), as well as USDA staff, in January 2005.

The consensus of the experts was that not even the most ambitious of the nonexperimental designs presented would predictably provide reliable estimates of program impact. They raised two major concerns:

- *Nonexperimental designs are vulnerable to selection bias*—that is, to pre-existing differences between program participants and eligible non-participants that may be related to the outcomes the program is trying to influence, such as individuals' dietary quality. A nonexperimental design would attempt to overcome selection bias by incorporating extensive information on factors associated with FSP participation, but the existing literature may not provide an adequate understanding of what those factors are. Moreover, even with good information on selection, no currently available methodology for nonexperimental research has consistently yielded "correct" results—i.e., the same results as randomized experiments—and researchers disagree about what analytic approaches are most likely to be effective.

- It is hypothesized that the FSP, by giving households greater food purchasing power, will cause increased food expenditures, which in turn will lead to a sequence of effects including improved household food supply, individual diets, and food security. ***Little is known, however, about whether the hypothesized chain of events is truly plausible. In fact, very little is known about the relationships between and among these outcomes in the low-income population in general***. Consequently, it is not possible to say how much impact, if any, should be expected from the incremental food purchasing power provided by the FSP. Thus, if an evaluation finds little or no impact on dietary quality at the household or individual level or on food security, which has been the result of most prior research, it is impossible to know whether null findings reflect reality or are caused by a flaw in design and/or measurement.

The TWG also noted that these obstacles would be magnified in any attempt to evaluate multiple nutrition assistance programs in a single study, which had been envisioned in some of the early designs. They recommended that any nonexperimental evaluation be focused on a single program.

In response to these concerns, the study team sketched out, and TWG members reviewed, a research program that focuses specifically on the Food Stamp Program. The program consists of nine preliminary studies that would culminate in one of two versions of a national study. The nine preliminary studies are divided into two groups. One group (five studies) addresses the issue of selection bias, and the other group (four studies) examines relationships among outcomes. Studies in both groups are sequenced, so that each study builds on what the prior studies have learned.

Implementing the two groups of preliminary studies is intended to provide a firm foundation for conducting a nonexperimental national evaluation of FSP impacts. That result cannot be guaranteed, however. The results might show that, even with the best possible understanding of FSP participation and relationships among potential FSP outcomes, a nonexperimental design will not predictably overcome selection bias. In that case, FNS could proceed with a national study of the experiences

and outcomes of FSP participants and nonparticipants, but could not expect to measure the program's impact.

Selection Bias Studies

The preliminary studies that focus on selection bias are designed to identify key factors that influence the decision to participate in the FSP and then test the effectiveness of a nonexperimental design in controlling for these factors. The program of research in this area includes five distinct studies, beginning with an analysis of FSP participation using the measures available in extant data. A subsequent qualitative study aims to develop detailed information on the circumstances that lead some eligible households to apply for FSP benefits and allow others to get by without participating. A third study incorporates information from the first two studies and uses a panel survey to estimate a comprehensive model of the factors related to FSP participation.

The last two studies in this group will test the effectiveness of a nonexperimental design. One study will be a random-assignment experiment, conducted with eligible non-participant households. The treatment group will be given intensive encouragement and support in applying for FSP benefits, while the control group will not receive special encouragement. A parallel nonexperimental study, which will incorporate information on participation obtained from the first three studies, will test the ability of the nonexperimental design to match the experimental results.

Outcomes Studies

The four studies in this set are designed to provide information about the relationships between low-income households' food expenditures, their household food supply, the dietary intake of individuals in the households, and shopping and other behaviors that may help explain the linkages (or absence of linkages) among these outcomes.

Three of the studies directly parallel the first three studies in the selection bias group. The sequence begins with analyses that mine the limited data available from previous surveys. This is followed by a qualitative study to develop an in-depth understanding of how individual households' food expenditures are connected to diet-related behaviors. Then a two-wave survey will examine these factors in a quantitative framework, measuring the extent to which changes in food expenditures are associated with changes in dietary outcomes and related behaviors.

The remaining study is a random-assignment experiment using households currently participating in the FSP. The treatment group would receive an increment to their normal FSP benefit—perhaps $30 per person per month—and the study would examine the extent to which this increment translates into differences in food expenditures and diet-related outcomes.

A National Study of the Food Stamp Program

The purpose of the preliminary studies is to provide guidance for a national study of the Food Stamp Program. If the preliminary findings demonstrate that a nonexperimental evaluation would be credible, the national study will use this approach to assess the Food Stamp Program's impact on participants' dietary quality and food security. This study would also capture information on the program experiences and program satisfaction of participants and eligible non-participants.

If the preliminary studies indicate that a nonexperimental evaluation design would not be reliable, even with extensive new information on program participation and relationships among outcomes, the national study would examine the program experiences and satisfaction of participants and eligible non-participants. It would also obtain descriptive information on diet and food security, but would not support an assessment of program impacts.

Time and Money Costs and Their Drivers

The full research program is long, complex, and expensive. We estimate that the nine preliminary studies require a period of about 12 years and expenditures of about $31 million, while the final national study would require an additional 4 to 5 years and $5 to $10 million to complete.

Most of the individual studies—all of the studies involving new survey data collection—are estimated to cost several million dollars. These relatively high costs are driven by four factors:

- **Intensive outcome measurement.** Most of the survey-based studies are designed to measure household food expenditures, household food use, individual dietary intake, food security, and household well-being. Each wave of data collection from a household requires two lengthy in-person interviews and one or two telephone interviews, and as many as three different people in the household may be involved. The subsequent coding and processing of the nutrient data for the household food supply and individuals' diets is also very labor-intensive and costly.

- **Multiple survey waves.** Nearly all of the survey-based studies envision multiple waves of data collection. When individuals' outcomes can be measured on multiple occasions, the analysis can adjust for characteristics that differ across individuals but are unmeasured, such as preferences for particular foods.

- **Random-digit-dialing surveys.** Most of the surveys include eligible nonparticipants or other low-income households as well as FSP participants. An RDD survey is necessary to draw these samples.

- **Large samples.** The initial sample sizes for most of the studies range between about 2,000 and 4,000 households. These sample sizes are not extraordinary, but they are quite substantial when considering that intensive data collection must occur over multiple waves.

Although precedents exist for research programs of this magnitude in evaluating Federal social programs, the recommended agenda would represent an extraordinary undertaking for FNS. This raises the question of whether the purpose might be achieved more quickly or cheaply, or indeed whether measuring overall program impact should be the highest priority for FNS' evaluation resources.

Options

The research agenda could be trimmed by making tradeoffs in several domains. These are:

1. **To omit some preliminary studies** based either on *a priori* decisions or on knowledge gained in the course of carrying out the earlier studies. This increases the risk that the preliminary studies will not produce a satisfactory nonexperimental evaluation design.

2. **To combine some studies.** Like the first option, this approach potentially saves time and money. But it substantially increases respondent burden in surveys that are already far more burdensome than what is usually considered feasible.

3. **To drop some dietary outcome measures.** Information on either household-level or individual-level outcomes could be eschewed, reducing respondent burden and survey costs but increasing the risk of uncertainty about the meaning of evaluation findings.

Choosing any of these options amounts to a gamble. Omitting preliminary studies could be effective if the information that they would provide turns out to be either unnecessary or insufficient to support a nonexperimental evaluation. Combining studies would gamble that the study would not collapse under the weight of respondent burden. Omitting measures rests on the hope that evaluation findings regarding diet quality and food security will not require explanation.

Finally, given the high costs and risks, it may be worth considering modified versions of the research objective. Rather than trying to learn the difference between participants' outcomes with and without the FSP, FNS might pose the question of whether the FSP is meeting specified targets. A target might be, for example, that at least 90 percent of all participating households should be food secure, or that no more than 15 percent should have "poor quality" diets based on HEI scores. Alternatively, research could focus on questions for which random assignment designs would be legal and ethical. These might be evaluations of particular program components, such as nutrition education, or evaluation of the program's impact on populations that are not currently eligible for the program but for whom eligibility might be considered. Such experiments would have direct policy relevance for the program component or population studied, and would generally produce lower-bound estimates of the effect of the FSP.

In sum, using a nonexperimental approach to evaluate the Food Stamp Program's impacts on participants' diets and food security will be extremely difficult, as illustrated by a long history of inconclusive research and the expert panel's rejection of the designs initially proposed in this project. The best hope of a credible nonexperimental evaluation lies in developing and applying extensive new knowledge about both the determinants of FSP participation and the relationships among hypothesized FSP outcomes. This will require a lengthy and expensive preliminary research program, estimated at 12 years and $31 million for the approach described here. Such a program would produce a great deal of useful information, but it cannot guarantee that a reliable nonexperimental evaluation will be possible.

Executive Summary **Abt Associates Inc.**

Chapter One: Introduction

The Food and Nutrition Service (FNS) operates over a dozen nutrition assistance programs for children and low-income individuals. The programs aim to improve participants' access to nutritious food and promote healthier eating habits and lifestyles, by providing specific foods or food purchasing power, and by offering nutrition education. The Food Stamp Program (FSP), which serves all low-income households, is the largest of these programs. In fiscal year (FY) 2003, the FSP provided benefits to more than 21 million participants per month. Other sizeable FNS programs serve specific population groups, including the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), which serves low-income infants and young children and low-income women who are pregnant or new mothers; and the National School Lunch Program (NSLP), the School Breakfast Program (SBP), the Summer Food Service Program (SFSP), and the Child and Adult Care Food Program (CACFP), which target preschool- and school-age children. The Food Distribution Program on Indian Reservations (FDPIR) and the Commodity Supplemental Food Program (CSFP) are smaller programs that distribute commodity food packages to Native Americans living on or near reservations (FDPIR) and to low-income women, children, and the elderly (CSFP).

As the steward of public funds, FNS is responsible for assessing the extent to which its programs achieve their stated missions as well as for assessing the level of satisfaction among program participants. To help explore options for addressing these requirements, FNS awarded a contract[2] to Abt Associates Inc. The purpose of this contract is to develop a menu of potential survey designs that can provide FNS with new information on:

- Experiences and satisfaction of participants in FNS programs, and

- Impacts of program participation on food security, diet quality, and other indicators of household well-being.

From the outset, it was understood that, while the best assessment of program impacts would be based on randomized trials, this approach is not feasible for FNS programs. The FSP and other major nutrition assistance programs are national entitlements,[3] and it is not ethically or legally permissible to deny benefits to eligible households or individuals in order to conduct an experiment.

Development of the Initial Menu

As a first step, Abt staff produced a design concepts memorandum (Burstein *et al.*, 2004a), which laid out key dimensions on which survey options could vary. Abt also recruited an external panel of technical experts to provide additional perspectives, comprising the following individuals:

- Tom Cook, Ph.D., Northwestern University, Departments of Sociology, Psychology, Education and Social Policy,

---

[2]   *Assessing the Food Security and Diet Quality Impacts of FNS Program Participation: Survey Design Options*, Task Order 43-3198-4-3811 under Contract 53-3198-2-026.

[3]   Although WIC is not an entitlement program, funding is generally sufficient to serve all eligible applicants.

- Jim Ohls, Ph.D., Mathematica Policy Research,
- Carol West Suitor, D.Sc., independent consultant, and
- Chris Winship, Ph.D., Harvard University, Department of Sociology.

Based on interactions with the panel members and FNS staff in response to the design concepts memorandum, Abt staff developed a *Draft Initial Menu of Design Options* (Burstein *et al.*, 2004b). A revised version of the *Initial Menu* (Burstein, *et al.*, 2005), incorporating comments from panel members and FNS, served as the basis of a meeting on January 26, 2005 of the technical experts, Abt staff, FNS staff, and other interested parties.

The Initial Menu took as its underlying premise that the vehicle for addressing the identified research questions would be a nonexperimental study using survey data collection methodology. It was expected to be the first step toward identifying, describing in detail, and comparing a set of three or four survey-based (nonexperimental) studies[4] that would:

- update existing information on experiences of FSP participants, which currently comes from the 1996 National Food Stamp Participant Survey (NFSPS), and provide comparable information for participants in other FNS programs; and

- advance, to varying degrees, the Agency's knowledge and understanding of how participation in the FSP and other FNS programs affects food security, diet quality, and other indicators of household well-being.

Although information about the four largest FNS programs—FSP, NSLP, SBP, and WIC—was deemed to be of the highest priority, the smaller targeted programs—SFSP, FDPIR, CSFP, and CACFP—were understood to be of interest as well.

Outcomes of the Technical Working Group Meeting

While the technical experts and other attendees had much to contribute on the issues of outcome measures, sample frame, and subgroups, their most significant comments pertained to the broad scope of the project and the intractability of selection bias. The consensus of the group was that the initial menu was misguided in three central aspects.

First, ***survey-based (nonexperimental) impact estimates unsupported by evidence of internal validity are of questionable value***. Selection bias is a serious threat, and the research community is becoming increasingly skeptical of claims that any given nonexperimental approach—e.g., use of covariates, propensity score analysis, instrumental variables, or two-stage procedures—will predictably and consistently yield valid estimates of program impacts. This argument is documented and elaborated in Chapter Two. Conversely, the validity of randomized trials has meant that in other fields, small randomized trials have had great persuasive power. In addition, paired with nonexperimental findings, they can provide confirmation of results. Hence research efforts can

---

[4]   For convenience, the term "survey-based" is used throughout this report to mean "nonexperimental," i.e., collecting data on extant populations. We note that either experimental or nonexperimental studies could use survey or administrative data.

profitably be focused on developing creative approaches to conducting legally and ethically permissible experiments.

Second, *we need to understand the causal links between outcomes before attempting to estimate program impacts on outcomes like food security and diet quality.* In a model of the effects of a program such as the FSP, pathways among intermediate, short-term, and longer-term outcomes are specified. For many of the links in the model, however, we do not have data to help us understand the strength of the relationship. As an example, the mechanism by which the FSP might affect food expenditures is understood, and it is generally accepted (in part because of the randomized cash-out demonstrations) that the FSP has a positive impact on food expenditures. The potential for impacts on more distal outcomes, however, such as household nutrient availability and individual dietary intake, is less obvious, and such impacts have never been clearly demonstrated. A well-designed study might find an impact on expenditures but not on dietary intake; we need to know whether that finding is attributable to (a) insufficient statistical power to measure second-order effects, or (b) a behavioral reality that an increase in food expenditures cannot be expected to lead to an increase in the quality of dietary intakes. The hypothesized relationships among outcomes are described in Chapter Two.

Finally, *it is highly desirable to focus attention on a single nutrition assistance program*. While we may be able to develop a design that leads to valid impact estimates for a single program, the resources (and luck) necessary to estimate impacts for multiple programs should not be counted on. The specific features of a particular program must drive the design of a study, and one study cannot expect to do justice to several sets of program features simultaneously. For example, the population of food stamp-eligible households is both too narrow for studying WIC and NSLP/SBP (as it excludes households between 135 and 185 percent of poverty) and too broad (as only minor fractions of the FSP population are categorically eligible for WIC and NSLP/SBP by virtue of age and pregnancy/lactation status). Furthermore, the previously mentioned difficulties of estimating single program impacts are dwarfed by those of investigating the impacts of multiple program participation, e.g., trying to understand why FSP and non-FSP households each choose to participate in WIC (or not), and how and why FSP impacts differ between WIC participants and WIC-eligible nonparticipants. Special issues surrounding impact evaluation of FNS programs other than the FSP are described in Chapter Sixteen.

In addition to these fundamental criticisms of the initial menu, the TWG meeting highlighted two methodological issues that need to be addressed for a successful evaluation: choice of appropriate outcome measures, and sufficiently large sample sizes to detect meaningful impacts. The challenges surrounding measurement of dietary outcomes are discussed in Chapter Two. Issues related to sample size are addressed in the later chapters of this report.

In response to the discussion at this meeting and subsequent correspondence, FNS modified the project objective. The revised objective is to identify, describe in detail, and compare three or four *research agendas* for studying the FSP that achieve the following:

- update the picture of the FSP participation experience, including information on how many and which of these households also receive WIC and/or school meals benefits,

reasons for participating or not participating in these programs, and the general satisfaction with these programs; and

*   advance our knowledge and understanding of FSP impacts on food expenditures, food security, diet quality, and other dimensions of household well-being.

The project objective was thus both narrowed and broadened. It was narrowed in three ways:

*   eliminating attention to smaller programs like the CACFP and SFSP altogether;

*   eliminating design proposals to assess WIC and school meal program impacts, but including a discussion of challenges for conducting impact assessments and ideas for addressing the challenges; and

*   confining design proposals that examine multiple program participation to FSP participants who also participate in WIC or school meal programs, assessing their level of satisfaction, and treating participation in these programs as a mediating variable in assessing FSP impacts.

The project objective was broadened in that the focus was expanded from a small number of options for a self-contained nonexperimental study to a small number of full research agendas that are each composed of multiple, linked studies, experimental and nonexperimental.

Development of the Final Menu

With this new understanding, Abt staff developed concepts for nine preliminary studies that address the concerns raised at the TWG meeting regarding selection bias and causal links. The findings of each of the preliminary studies are intended to provide guidance as to next steps in the research agenda, culminating ultimately in implementation of one of two national studies: (a) a survey-based FSP impact study, if the preliminary studies indicate that such an impact study is feasible, or (b) a descriptive study that assesses participant experiences and outcomes of interest but does not attempt to measure program impacts. The menu of research agendas comprises the full agenda just described, and three alternative agendas that proceed to a national study with fewer preliminaries.

The nine preliminary studies fall into two broad groups, corresponding to two major issues that must be understood before attempting to determine Food Stamp Program impacts. These issues are:

(1) Can FSP impacts be reliably estimated using survey-based, nonexperimental methods?

(2) Can we trace the pathways of FSP effects on household food expenditures through to effects on other outcomes, including household food security and individual dietary intake?

The remainder of this report is in two parts plus a coda. In Part One, we describe the challenges of conducting a national study of the FSP (Chapter Two); lay out a general research strategy for getting to either an impact study, or a decision that such a study should not be attempted (Chapter Three); and address the time and resource costs of the proposed research agenda (Chapter Four). In Part Two (Chapters Five through Fifteen), we describe in some detail the individual studies that comprise the

complete research agenda, concluding with descriptions of two versions of a national study of the FSP—one that focuses on program experiences, satisfaction, and outcomes, and one that supports impact estimates as well. The final chapter (Chapter Sixteen) discusses the issues that would need to be addressed in estimating impacts of other FNS programs.

Chapter Two: Challenges of Conducting a National Study of the Food Stamp Program

The twin objectives of the research agendas developed in this project are:

- To update the picture we have of the FSP participation experience, including information on how many and which FSP households also receive WIC and/or school meals benefits, reasons for participating or not participating in these programs, and the general level of satisfaction with these programs; and

- To advance our knowledge and understanding of FSP impacts on food expenditures, food security, diet quality, and other dimensions of household well-being.

In the sections that follow we begin by presenting a conceptual model of FSP impacts and a critical summary of research findings on program effects to date. We then discuss measurement issues pertaining to the domains of experience and outcomes subsumed in the two goals of the research agendas. We next describe the ideal random assignment experiment for measuring impacts, and contrast it with nonexperimental approaches. We conclude with a discussion of criteria for a successful impact evaluation.

Conceptual Model of FSP Impacts

Exhibit 2.1 shows the hypothesized causal chain between FSP participation and outcomes of interest. There are, of course, many other important influences on these outcomes, including characteristics of individuals, households, and communities, which are not depicted.

The potential links between the FSP and the various outcomes include:

1. The *dual intervention of FSP for participating households*, increasing their food purchasing power and providing nutrition education—including food budgeting, label reading, and efficient shopping practices.

2. *Greater food purchasing power* leading to increased expenditures on food, greater food security, and improved household well-being (e.g., resources freed for meeting other needs).

3. Households' improved *nutrition knowledge, attitudes, and skills* due to nutrition education leading to changes in household food expenditures, improvements in the nutritional quality of the household food supply, and improvements in the quality of the diets consumed by individuals in the household.

4. *Increased expenditures on food* leading to improvements in the nutritional quality of the household food supply.

5. *Improvements in the nutritional quality of the household food supply* leading to improvements in individual diet quality.[5]

---

[5]   For brevity, Exhibit 2.1 omits the role of food eaten away from home.

6. Improvements in diet quality leading to more appropriate weight status and improvements in other health outcomes.

7. Additional positive effects on food security from increased food purchasing power and food expenditures, improved household food supply, and improved diet quality.

Exhibit 2.1

Hypothesized Causal Chain of FSP Impacts
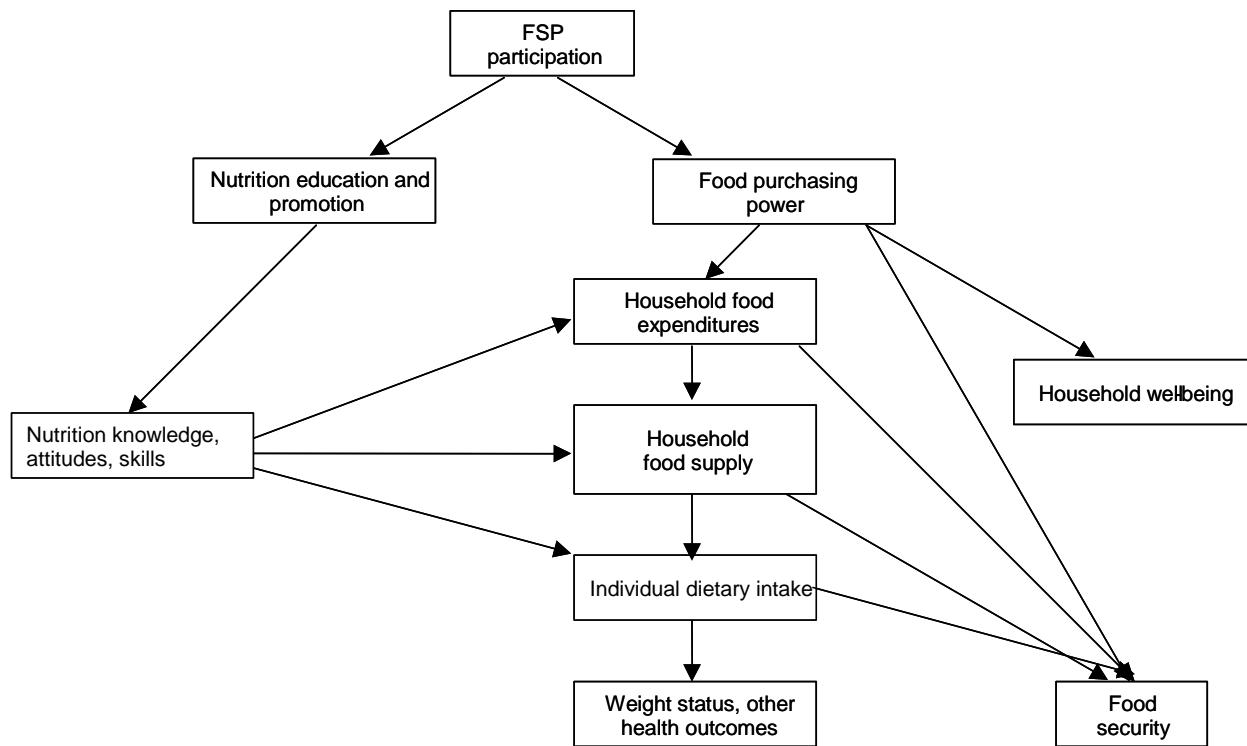
Research Findings to Date

Based on this causal chain, it may be supposed that FSP participation could have positive effects, direct or indirect, on a wide variety of outcomes. Well over 100 studies on FSP impacts were reviewed and summarized by Burstein *et al,* (2004a). Most of these studies were based on nonexperimental comparisons of participants and nonparticipants, some done extremely carefully, but nonetheless subject to significant threats to internal validity. Some were dose-response analyses, comparing outcomes for households receiving greater versus lesser food stamp allotments. Although this group of studies restricted the sample to FSP households, they were also vulnerable to selection bias, because households that choose to participate in the FSP when eligible for only a small benefit may differ in important ways from those that participate when eligible for a larger benefit. Two true experiments provided solid information in several areas, though they offered only lower-bound estimates of program effects.[6]

---

[6] These were the Alabama and San Diego Cash-out Demonstrations, which randomly assigned FSP participants to receive their benefits as food stamps or as unconstrained income. Differences in outcomes captured the effects of earmarking only, not of the increased purchasing power.

The studies of **household food expenditures** found marginal impacts of FSP benefits ranging from 0.17 to 0.47.[7] Remarkably, there is no evidence that this impact declined after the elimination of the purchase requirement (EPR) in 1979, despite the widespread belief that the EPR decreased the impact of participation by constraining fewer households in their food purchases.[8] A possible explanation is that the reduced constraint on households that would have participated pre-EPR was counterbalanced by the surge in FSP participation post-EPR, which brought in households who were constrained under the new régime (and who refused to participate at all under the old régime).

Findings from studies of **household nutrient availability** were less convincing. Many of these studies used data collected in 1980 or previously. The more recent studies, including the San Diego cash-out experiment, provided good evidence that the availability of energy and protein from household food supplies was increased by FSP participation. Support for impacts on the availability of vitamins and minerals from household food supplies was substantially weaker. These findings may have limited applicability today, however. The authors of the review note that "the American food supply has changed dramatically in the past 20-25 years, with important implications for both nutrient availability and individual dietary intake. Americans are eating substantially more grains than they were two decades ago, particularly refined grains, as well as record-high amounts of caloric sweeteners and some dairy products, and near-record amounts of added fats." Other changes in this time period include new food products, changes in food enrichment policies, and an increase in the number of meals obtained and eaten away from home.

There was little evidence of consistent impacts on **individual dietary intakes**, although scattered effects were found for young children; and it was impossible to assess impacts on dietary adequacy. Many studies used intake data for a single day and could not shed light on usual dietary intake. None of the remaining studies used the approach to estimating usual intake recommended by the Institute of Medicine (IOM) (IOM, 2001). The strongest study on patterns of food intake found negative effects of FSP participation for various age groups on servings of grains and grain products, servings of vegetables, dietary fiber, and the likelihood of obtaining less than 10 percent of total energy from saturated fat.

Potential salutary impacts of FSP participation on **food security** were swamped by intractable selection effects, because food insecurity leads households to seek food assistance.[9] Studies of other nutrition- and health-related outcomes (birthweight, height and weight, nutritional biochemistries, general measures of nutrition or health status) were at best suggestive because of their methodological limitations.

---

[7] This range, based on 29 studies, omits three outlier estimates from studies whose methodology was questionable.

[8] Prior to 1979, households were required to purchase their benefit allotment, which was a function of household size, paying an amount that depended on household income. The difference between the allotment and the amount paid was called the bonus amount. After 1979, households simply received the bonus amount. Post EPR, households were only constrained to spend the bonus amount on food rather than the entire allotment.

[9] Recent research at ERS has, however, established a strong positive link between food expenditures and food security. See, for example, Nord *et al*. 2005.

The conceptual model presented above showed pathways by which the FSP could, in principle, affect outcomes more distal than food expenditures, such as individual dietary intake and food security. Yet research to date has failed to show convincingly that such effects occur. Even at the household level, the available evidence suggests that FSP impacts on nutrition may be limited to increasing the availability of food energy and protein. It is essential for policy makers to distinguish between two competing explanations of these non-findings. One explanation is that the conceptual model is correct and the causal relationships are all there, but previous studies have failed to detect them due to weaknesses in study design. These weaknesses could be of several sorts:

1. ***Insufficient statistical power.*** Because the causal effects of FSP participation on distal outcomes are multiplicative, very large samples might be needed to detect them. Only some fraction of food stamps—perhaps 30 percent—is translated into increased food expenditures, and only some fraction of this increase goes to purchase more nutritious food. At present, FSP benefits amount to about $90 per person per month, so that food expenditures might increase by about $1 per day per person as a result of FSP participation. This increase in expenditures might yield a more nutritious household food supply, but it might also yield better tasting food, food that is easier or quicker to prepare, treats such as snacks or expensive cuts of meat, entertainment from eating out, or social value from feeding non-household members. The effect on diet quality, even if it were positive, might thus be quite small.

2. ***Intractable selection bias.*** Participant-nonparticipant comparisons could fail to detect FSP impacts if they were swamped by selection bias. Households may choose to participate in the FSP because they lack other resources possessed by eligible nonparticipants, or because they are prone to food insecurity for whatever reasons. Failure to control adequately for these differences could mask positive effects of participation.

3. ***Data deficiencies.*** The data needed to measure these effects are expensive and difficult to collect. In particular, the respondent burden associated with supplying information on both the household food supply and individual intake is very high.

Alternatively, it may be that the causal relationships assumed in the conceptual model are incorrect—that diet quality at the individual or even the household level is *not* positively related to food expenditures, or at least not at current levels of food abundance among even low-income American households. The failure to find beneficial effects of FSP participation may reflect a systematic tendency to use increased food purchasing power to buy items that are more convenient or taste better, rather than to buy healthier/more nutritious food. Many food items that are more convenient or "taste better" are higher in salt, sugar, and/or fat, and lower in fiber than their counterparts.

It is thus essential before carrying out an impact study to understand (a) whether the links in Exhibit 2.1 are operative and (b) if so, how a study should be designed to measure those links, with regard to data quality and sample sizes. That is, if there is good reason to believe that links exist between food expenditures and more distal outcomes, then the design for a study of FSP impacts needs to be strong enough to detect them. If the inherent expense or logistical practicalities preclude a sufficiently strong design, then the impact study should not be attempted.

If, on the other hand, it is concluded that *increased food expenditures* do not lead to better nutritional outcomes for low-income households or individuals, or to increased food security, then there is no reason to expect the *purchasing power of food stamps* would have these effects, and no point in an evaluation designed to detect such effects. Furthermore, given the persuasive evidence from prior studies that FSP does cause increased food expenditures, there is little value in assessing impacts on this outcome for its own sake in a new study. Failure to find a relationship between food expenditures and diet quality or food security would suggest that the focus of future FNS research should be on what, besides increased food expenditures, is needed to affect changes in these outcomes. A top priority for FSP research might therefore be to develop an approach to nutrition education that can strengthen these links.

Measuring Program Experiences and Satisfaction

The most basic information that policymakers need about a program is how it is viewed by participants, and by those who could participate but choose not to. The national study of the FSP will provide an opportunity to obtain information about low-income households' experiences and satisfaction with the FSP itself, and also about their experiences and satisfaction with other three major food assistance programs—WIC, SBP, and NSLP. The study will be limited to households eligible for the FSP, i.e., under 130 percent of the federal poverty level (FPL). It will therefore not provide a full picture of these other programs, which have higher income cutoffs for participation,[10] but will show how these programs affect those who do or could participate in the FSP.

The customer satisfaction component of the NFSPS assessed three dimensions of customer satisfaction: accessibility, costs of participation, and service (how well the program works). These broad dimensions can be used to assess customer service in virtually any FNS program.

Aspects of customer satisfaction that can be addressed in each domain are summarized below. This list is based on the NFSPS and refers primarily to the FSP, but most items are applicable with modification to all FNS programs.

Accessibility

- Circumstances that lead households to apply for program participation
- Reasons that some eligible households do not apply for participation
- Reasons that approved households never participate or drop out
- Number of visits to a program offices or other locations, and amount of time required, for an application, recertification, or issuance
- Alternative application procedures, including direct certification for SBP/NSLP, use of adjunctive eligibility for WIC, and expedited application in the FSP
- Participants and nonparticipants' understanding of program eligibility criteria

---

[10] WIC participants may have household incomes up to 185 percent of FPL, or higher if they are adjunctively eligible through Medicaid in a State with a higher eligibility threshold. SBP and NLSP meals are available for free to children in households under 130 percent of FPL, at reduced price for children in households between 130 and 185 percent of FPL, and at "full" price for children in households above 185 percent of FPL. (The "full" price may reflect a variety of subsidies.)

Costs of Participation

- Monetary costs (e.g., transportation, child or elder care) associated with the application and recertification processes
- Perceived stigma associated with applying for or using program benefits
- Stigma as a deterrent to program participation

Customer Service

- Participants' satisfaction with benefits
- Participants' satisfaction with program administration and operations, and suggestions for improving program administration/operations
- Customer service issues as a barrier to participation

We note that Abt Associates recently completed the Food Stamp Program Access Study, in which these aspects of customer satisfaction were addressed (Bartlett, Burstein, and Hamilton, 2004). Information on FSP barriers and experiences was collected from food-stamp eligible households of all type: nonparticipants, "near applicants" who had contacted the local office but had not filed an application, households that filed an application but did not complete the interview and verification process, households that applied and were approved for benefits, and ongoing participants.

Measuring Outcomes

The outcome domains of interest are those shown in Exhibit 2.1: household food expenditures, (quality of the) household food supply, individual dietary intake, weight status and other health outcomes, food security, and household well-being. The natural time frame for considering FSP impacts is one month, which is the period for which food stamp benefits are issued. Ideally, the study would use strong measures of food expenditures and diet quality covering *30-day* periods. As discussed below, however, the preferred measures for the outcome domains use reference periods of varying lengths. The accepted approach to measuring food expenditures and diet quality at the household level is the *7-day* food use record. The "gold standard" for measuring individual dietary intake, in contrast, is the *24-hour* recall, with a second recall for a subsample of the population, to allow for estimation of *usual* intake for population groups. Food security is most commonly measured for a *12-month* period, although the survey items can be worded to refer to a 30-day period.

Because strong measures with consistent reference time periods are not available, compromises must be made when assembling the pieces in order to estimate the relationships between food expenditures and diet quality, and between food expenditures and food security. If expenditures are measured for a different time period than diet quality or food security, measurement error will make the relationships appear weaker—that is, to have larger standard errors—than if the time periods were the same. This problem may be particularly severe for the relationship between household food expenditures and individual diet quality, which is also weakened by the fact that the two outcomes are measured at different levels. The only way to obtain consistent time frames for all outcomes would be to use measurement approaches that are not well validated, which would simply introduce different kinds of measurement error.

Our approach uses the optimal measurement approaches available at the time of this writing. Below, we describe available and recommended approaches to measuring each outcome. At the end of the section, we discuss the issue of response burden associated with measuring all outcomes in a single study.

Household Food Expenditures

Two approaches are available for collecting information on household food expenditures. These are the record-assisted recall method used in the NSFSP, which was based on the methods used in the USDA-sponsored Household Food Consumption Survey (NFCS) and Continuing Survey of Food Intake by Individuals (CSFII), and a detailed expenditure diary, such as the one used in the Consumer Expenditure Survey (CES).[11]

Two reference units for food expenditure data have been used in the field. The NFSPS obtained expenditure data for *foods **used** from the household food supply* during a specified 7-day period. This includes foods prepared or used at home, but excludes food obtained and eaten outside the home (such as restaurant meals or meals at friends' homes). In addition, foods taken from home food supplies but not actually eaten, such as waste in cooking and plate waste, are included. The CES, on the other hand, collects data on *foods **purchased** for at-home use and foods purchased and eaten away from home* during a 2-week period.

Both units for food expenditure data (food *use* and food *purchases*) provide an indirect indicator of the resources that households actually expend on food, which results in understatement of actual value of food acquired or used by a household. For example, both approaches exclude food received through food/meal-based nutrition assistance programs such as the NSLP, SBP, CACFP, SFSP, elderly feeding sites, and soup kitchens, as well as food received through meals eaten at friends' or relatives' homes. Food use data additionally fail to consider costs for foods that are purchased and consumed away from home (snacks, restaurant meals, etc.). Food purchase data also fail to consider food received through WIC vouchers, commodity distribution programs (such as the FDPIR and the CSFP and comparable programs sponsored by local civic and charitable organizations), and food acquired through gardening, hunting, and fishing.

The consensus at the TWG meeting was that the methods used in the NFSPS continue to be the most appropriate for this research agenda. Food use data provide a better picture of the household food supply—the food source that food assistance programs such as the FSP, WIC, and commodity-based programs are designed to affect—and may be less susceptible than food purchase data to the influence of the benefit cycle. For example, 42 percent of all FSP households make major food shopping trips only once per month; and Wilde and Ranney (2000) found that mean food spending by FSP households peaks sharply in the first 3 days after food stamp benefits are received. While the distribution of expenditure diaries across the month could ensure that mean outcomes are unbiased, this additional source of variation for FSP households would make it more difficult to compare outcomes with nonparticipants. In addition, the individual intakes of household members are likely to

---

[11] Participants in the TWG meeting concluded that simple recall-based data on food expenditures, such as that collected in the Current Population Survey Food Security Supplement (CPS–FSS) are neither reliable nor detailed enough to meet the needs of the research agenda.

be more closely linked to information on foods used from household food supplies than to data on food purchases.

It was recommended that the NFSPS methods be carefully examined and updated to reflect changes that have occurred in food consumption behaviors and the marketplace over time. A potentially significant change is the fact that many food packages no longer have individual price tags that respondents can refer to in reporting cost data. Having respondents keep register tapes will partially solve this problem, for foods that are purchased within the timeframe of the data collection in stores that use scanners. However, for foods that were purchased prior to data collection, or from sources that do not use scanners, prices may ultimately have to be imputed from data compiled by private companies. In addition, given the increasing role of away-from-home foods, it makes sense to supplement measures of household food use with questions that assess these expenditures. The burden associated with collecting food use data is substantial, for both data collection and analysis. Collection of detailed data on food expenditures requires in-person data collection and the use of respondent-maintained records. An obvious way to decrease burden would be to collect data for a shorter period of time. However, NSFSP researchers studied this issue and found that the length of the data collection period had a noteworthy impact on expenditure estimates. Given that there is no way to determine which estimates were more accurate, it does not seem advisable to curtail the data collection period without research that would support this change.[12]

Quality of the Household Food Supply

The food use data described above can also be used to assess the quality of the household food supply. The only additional data collection requirement would be that the instruments be designed to capture information about the characteristics of food that affect nutrient content (e.g., fat content of milk and ground beef). The quality of the household food supply can be assessed in a very detailed way, by estimating household nutrient availability, as was done in the NFSPS, or in a more general way, by assessing key characteristics of the foods used, e.g., use of low-fat dairy products, lean meats, and whole grains; use of fruits and vegetables (all types and fresh produce); use of sweetened fruit drinks and carbonated beverages; use of sweets, desserts, candy and snack foods; and use of added fats and oils. Another potentially interesting approach mentioned at the TWG meeting is estimation of a household-level version of the Healthy Eating Index (HEI). (The HEI is described in the next section).

Individual Dietary Intake

At the TWG meeting, it was agreed that assessment of individual dietary intake should consider both the adequacy of nutrient intake (whether individuals are getting enough of the nutrients they need to

---

[12]   The NSFSP included a subsample of participants who provided data for a period of 4 rather than 7 days. Analysts compared results of 4-day and 7-day records and found that estimates of food expenditures were consistently greater for the 4-day sample than the 7-day sample (Cohen *et al.*, 1999). The average 4-day estimate of total household expenditures was 11 percent greater than the average 7-day estimate, and the average estimate of the total value of food used per person was 3.5 percent higher for the 4-day sample than the 7-day sample. Four-day estimates could be more accurate because respondents did a better job of reporting food use over a shorter period of time (due to better recall or lack of "respondent fatigue"). On the other hand, the 4-day estimates could be less accurate because of "telescoping" error (respondents' tendency to report more events in a limited amount of time than actually occurred).

maintain good health) and the healthfulness of food choices (how well individuals' diets conform to accepted recommendations for healthy eating, including avoidance of excessive intakes of energy as well as nutrients and other dietary components known to be associated with chronic disease).

Assessing Nutrient Adequacy. Beginning in 1999, the Institute of Medicine (IOM) issued new nutrient intake standards—the Dietary Reference Intakes (DRIs) (IOM, 1999, 2000a, 2000b, 2002a, 2002b, 2004a). These standards replaced the Recommended Dietary Allowances, which were used in most previous research on nutrition assistance programs. For many important nutrients, the DRIs include a new reference standard: the Estimated Average Requirement (EAR). The EAR can be used to assess the prevalence of inadequate intakes in population groups. The IOM recommended a specific approach for collecting and analyzing nutrient intake data that will ultimately be assessed using the EAR. The recommended approach requires one 24-hour recall (or diet record) for the entire group, and two non-consecutive days or three consecutive days of data for a subsample of the population. The additional day(s) of data are used to estimate and control for intra-individual variability in intake, resulting in more reliable estimates of the group's usual intake of key nutrients.

In making this recommendation, the IOM concluded that other approaches to measuring nutrient intake, particularly semi-quantitative food frequencies, may be useful for other purposes but are not accurate enough to assess reliably the nutrient intakes of either individuals or groups (IOM, 2001). Key limitations include lack of information on portion sizes, restricted focus on a limited number of foods/food groups, and the use of composite food composition data. In addition, quantified food frequency instruments are generally appropriate only for the population for which they were designed, based on food consumption patterns at a specific point in time.

At the TWG meeting, it was agreed that it is advisable to include the IOM-recommended approach in any study design that includes adequacy of nutrient intake as a key outcome.[13] There are no firm recommendations on the number of replicate recalls needed. However, to support subgroup analyses, it may be necessary to collect second recalls on up to 25 percent of the sample. The protocol for the ongoing National Health and Nutrition Examination Survey (NHANES) calls for collection of second recalls from all individuals. Resources permitting, this is an attractive approach. The ultimate

---

[13]   It was suggested that, if resource constraints preclude collection of the second 24-hour recalls required for application of the EARs, assessment of nutrient adequacy could be limited to nutrients for which Adequate Intake levels (AIs) rather than EARs have been defined (calcium, vitamin D, others). For these nutrients, assessment of nutrient adequacy is based on comparison of group mean intakes to AIs. Reliable estimates of group means can be obtained with a single 24-hour recall. Most attendees did not consider this to be an attractive option because (a) such an approach would severely limit the nutrients that could be examined and (b) comparison of group mean intake to an AI does not provide definitive information about the adequacy of diets being consumed by the population under question (IOM, 2001). When group means are equivalent to or greater than the AI, the prevalence of inadequacy can be assumed to be low. However, when group means are below the AI, it is not possible to draw any conclusions about the prevalence of inadequate intakes (IOM, 2001).

---

decision about the percentage of the sample asked to provide replicate recalls will likely be driven by resource constraints.[14]

Assessing Diet Quality. Nutrition researchers have developed a number of different indices that measure overall diet quality. The index used to measure diet quality should provide broad coverage of the nutrition principles included in the *Dietary Guidelines for Americans* and the Food Guide Pyramid (recently updated and renamed the MyPyramid food guidance system).[15] USDA is committed to encouraging nutrition assistance program participants to consume diets that are consistent with these recommendations.

Potential options for a comprehensive summary measure of diet quality include:
- Healthy Eating Index (HEI) (Kennedy *et al.*, 1995),
- Diet Quality Index-Revised (DQI-R) (Haines *et al.*, 1999),
- Alternative Healthy Eating Index (AHEI) (McCullough *et al.*, 2002; Kant *et al.*, 2000),
- Recommended Food Score (RFS) (McCullough *et al.*, 2002; Kant *et al.*, 2000), and
- Diet Status Index (DSI) (Basiotis et al., 1995).

The characteristics of the different indices are described in a previous report (Burstein *et al.*, 2005). The consensus at the TWG meeting was that the HEI is the optimal summary measure of diet quality

---

[14]   We note that in 2003, NHANES added a non-quantified food frequency questionnaire (FFQ) to the protocol for dietary data collection. The FFQ, which includes 139 individual items, is mailed to respondents after they complete their other data collection requirements, as a supplement to the two 24-hour recalls already being collected. As noted at the TWG meeting, however, knowledge of usual intake of foods by itself does not provide information on such aspects of dietary quality as nutrient adequacy or intake of fats. That is, the FFQ is not a substitute for a 24-hour recall. As a supplement to the 24-hour recall, it would increase the respondent burden.

We corresponded with Kevin Dodd, who is leading the efforts at NCI to develop estimates of usual food intake, about the valued added by the FFQ. Dr. Dodd explained that the statistical modeling method used by NCI considers the 24-hour recall as the primary survey instrument. He reiterated that the FFQ could not substitute for the 24-hour recall, but might offer more or less supplemental information. Contrary to NCI's expectations, the estimates of the distributions of usual intake when the FFQ was included as supplementary information were not substantially different from the estimates obtained using only the 24-hour intake data. Estimates of individual usual intake were however substantially improved when the FFQ information was included.

He wrote:

Our feeling is that, when interest is in estimating individual usual intake, the ultimate goal (in general) is to relate usual intake to a health outcome; i.e., usual intake is a predictor, not the outcome of interest. In certain specialized applications, such as assessing diet in a clinical setting, there may be motivation to estimate a particular individual's usual intake for its own sake, but in such cases, there is a need for much greater amounts of information (be it many 24-hour recalls, food diaries, or whatever) to produce a good measure of diet.

He concluded:

The FFQ alone does not track the intake measure well, and when looking at summary measures of diet across a group (if that is your outcome measure), it doesn't make much difference whether you use just the 24-hour data or use 24-hour data with the FFQ as a covariate; you get roughly the same answer.

[15]   http://www.usda.gov/cnpp/pyramid.html.

---

for the proposed research agenda. The HEI is the standard used in FNS's strategic plan for 2000–2005 (USDA/FNS, 2000). In addition, USDA's Center for Nutrition Policy and Promotion (CNPP) is currently revising the HEI to incorporate recent updates in the *Dietary Guidelines* and the new MyPyramid food guidance system.[16] Finally, CNPP publishes reference data on HEI scores for the total population and key subgroups, based on national survey data.

Brief Assessments of Key Dietary Behaviors. An alternative to the detailed estimates of food and nutrient intake needed to support estimation of nutrient adequacy and the HEI is assessment of specific dietary behaviors associated with superior diet quality—for example, consumption of fruits and vegetables—using survey questions. FNS is currently working with the Economic Research Service on development and validation of a questionnaire to assess dietary behaviors of low-income populations (see Burstein *et al.*, 2005). It is possible that this validation work might be completed by the time FNS needs to finalize a design for the studies that would include assessment of individual dietary outcomes.

Food Security

USDA has developed four different scales that can be used to assess household food security: the full 18-item U.S. Household Food Security Scale; the 6-item Short Form Scale; the 10-item Adult Scale; and the Children's Scale. All of the food security scales, which are described in detail in a previous report (Burstein *et al.*, 2005), are well-suited to either in-person or telephone administration.

The survey module used for the full 18-item scale includes screeners that keep respondent burden to the minimum level needed to obtain a reliable assessment of household food security. Most households without children are asked only three questions; most households with children are asked five. Survey items may be asked with a 12-month or 30-day reference period.

When overall response burden precludes use of the full 18-item scale, either the 10-item Adult Scale or the 6-item Short Form scale can be used. Because of the sensitive nature of the questions about children, the 10-item Adult Scale is also recommended for use in studies that are not specifically focused on children or households with children. Field reports from data collectors using the full 18-item scale have indicated that administration of the child-specific questions sometimes causes stress or anxiety for the respondent or the data collector. This scale comprises 10 items from the full 18-item scale that do not specifically concern children. It results in the same four-category food security status measure as the full scale. The 6-item Short Form has two disadvantages relative to the 10-item Adult Scale. It provides only a three-category food security status measure and does not measure the most severe levels of hunger. It is also less precise and somewhat less reliable than the 18-item scale in its ability to assign correct food security status measures to individual households.

For the current study, the 10-item scale is probably the best choice. Omission of the sensitive child-focused questions reduces the possibility of non-response and the overall response burden, and the

---

[16]   The revised HEI is expected to address aspects of diet quality that are considered in both the DRIs and the new *Dietary Guidelines for Americans*, namely, intake of total fat (and potentially other macro nutrients) and sodium intake. The DRIs have defined Acceptable Macronutrient Distribution Ranges for fat, carbohydrate, and protein and an upper limit for sodium. The revised HEI may also address fiber intake. If not, this dietary constituent can be examined separately.

adult-based measure is still comparable across households with and without children. If response burden is a major concern, the 6-item Short Form scale could be used.

Household Well-Being

Survey items that assess household well-being are available from the Adult Well-Being module in the Survey of Income and Program Participation. This module covers a wide variety of topics, all of which are of potential interest for the proposed studies. Given the wide-ranging goals of the research agenda, it is probably necessary to focus on "basic needs," with perhaps consideration of unsafe housing conditions and overcrowding.[17]

Weight Status

Research has shown that both cross-sectionally and prospectively, the determinants of weight change are multifactorial (Sherwood *et al.*, 2000). Key determinants may include gender, race/ethnicity, age, energy intake, fat intake, physical activity, presence of restrained eating behaviors, household food security, and, among children, parents' weight, eating patterns, and physical activity. A recent analysis from the Framingham Heart Study showed that both short- and long-term risks of becoming overweight or obese were substantial for both males and females (Vasan *et al.*, 2005).[18]

Clearly, any attempt to assess the independent impact of the FSP on weight gain would require collection of a substantial amount of data and careful statistical modeling. Given the uncertainty about whether the earlier and, arguably, less complex links in the causal chain are operational, it seems premature to expend resources on this endeavor. At the same time, concerns about the ongoing obesity epidemic, in the population at large and among the low-income population that participates in FNS programs, suggests that studies focused on diet-related impacts should not ignore the issue of weight status. The general consensus at the TWG meeting was that, resources permitting, weight status should be assessed for descriptive purposes as well as to provide a potentially useful covariate (e.g., weight status may influence individual dietary intake).

Self-reported heights and weights are used in several ongoing national monitoring surveys (e.g., the Behavioral Risk Factor Surveillance System (BRFSS) and the Youth Risk Behavior Surveillance System (YRBSS)). However, researchers have documented significant discrepancies between self-reported weights and/or heights and actual measurements among adults (Paccaud, Wietlisbach, and Rickenbach, 2001; Villanueva, 2001; Spencer *et al.*, 2002) as well as adolescents (Brener *et al.*, 2003; Himes and Faricy, 2001). In an analysis of NHANES–III data, Villanueva (2001) found that weight discrepancy was positively associated with age, and negatively associated with measured weight and BMI. Other factors associated with the validity of self-reported weight were gender and, for one or both genders, race/ethnicity, education, cigarette smoking, a desire to lose weight, marital status,

---

[17]   Using the 1996 SIPP, Abt researchers found that even among households under 100 percent of the federal poverty level, 98 percent had a stove and 99 percent had a refrigerator, and fewer than 5 percent lived more than 1.5 persons per room (Ouellette, Burstein, *et al.*, 2004). Other housing problems were, however, more frequent: 25 percent of households living in poverty had problems with pests ("rats, mice, roaches, or other insects") and 11 percent had a leaking roof or ceiling

[18]   The observed 4-year rates of developing overweight were 14 percent to 19 percent in women and 26 percent to 30 percent in men. Four-year rates of developing obesity were 5 percent to 7 percent in women and 7 percent to 9 percent in men.

---

income, physical activity level, and number of months since the last doctor's visit. Spencer *et al.,* (2002) found that use of self-reported heights and weights resulted in inappropriate classification of BMIs for 22.4 percent of women and 18 percent of men. In a recent study of maternal employment and children's nutrition, Abt researchers found that proxy-reported heights of young children in the CSFII were unusable, because they were implausibly concentrated at rounded values such as 24 inches, 30 inches, and 36 inches (Crepinsek and Burstein, 2004). A similar experience was reported by Devaney and her colleagues (2004) with self-reported data collected for the 2002 Feeding Infants and Toddlers Study (FITS).

To get a better understanding of the discrepancies between the two sources of data and implications for the proposed study, we analyzed NHANES-III data, which includes both self-reported and field-measured heights and weights, for adults and for children 12 and older (self-reported or caregiver-reported data were not collected for children under 12). We compared self-reported and measured heights and weights and computed the percentage difference [(Measured – self-report) / Measured] for height, weight, and Body Mass Index (BMI).

We found that, on average, males overreport both height and weight, while females tend to overreport height and all but the oldest groups of females tend to underreport weight. With few exceptions, comparable patterns were observed for FSP participants and both income-eligible and higher-income nonparticipants. For each variable (height, weight, BMI), only one of the 48 between-group comparisons (24 age and gender groups × two FSP *versus* non-FSP comparisons) was found to be statistically significant, and these significant differences were not consistent.

These findings indicate that, among individuals 12 and older, there is no difference between FSP participants and nonparticipants in the relative validity of self-reported data. Consequently, self-reported data could be considered for these age groups. Based on the patterns Abt researchers have observed in the CSFII data, however, we strongly recommend field data collection rather than proxy reports for younger children. If food expenditures and individual dietary intake are included as outcomes, several in-person visits will be required. Having field staff obtain data on heights and weights is a marginal incremental cost that will yield substantially better data.

Other Nutrition and Health Outcomes

A study of FSP impacts could include other nutrition- and health-related outcomes, such as nutritional biochemistries or information on birth outcomes. Given the other broad-ranging goals for the proposed research, however, the merit of including such outcomes is questionable. All of the other outcomes can be measured for all households or individuals. Incorporating outcomes that are relevant only for one group of individuals (e.g., pregnant women) goes beyond the scope of the proposed research. Although nutritional biochemistries could be assessed for all population groups, the prevalence of nutritional deficiencies in the general population is low (Fox and Cole, 2004). Hence sample sizes would need to be quite large in order to have an adequate sample of "nutritionally deficient" individuals on which program participation could have a discernable impact.

Response Burden of Measuring Outcomes

We have assumed that any national study of the FSP would measure food expenditures, quality of the household food supply, individual dietary intake, food security, and household well-being. We have made this assumption because we believe that it is essential that FNS have information on

---

expenditures *and* household food use *and* individual intake in order to fully understand the impacts of the FSP. FNS's strategic goals call for "improving the food choices made by program participants," and the standard used to assess performance on this goal is the HEI (FNS, 2000). This focus makes it is essential that individual dietary outcomes be assessed. However, if a study were to focus exclusively on individual intakes, FNS would not be able to explain null findings: Did the FSP fail to improve individual intake because it failed to increase household food expenditures and/or the quality of food available in the household, or because a better household food supply does not guarantee higher quality diets at the individual level?

To ensure that any of the planned or potential studies has the power to detect relationships that may exist between and among these outcomes, outcomes must be measured accurately. If cost and/or response burden become concerns, it would be better to drop an outcome entirely than to use a substandard, but perhaps less burdensome or costly approach to measuring it.

To collect data on all of the "links" in the causal chain using the methodologies recommended above puts a large response burden on sampled households. Data collection would proceed as follows. Once a household is selected for inclusion in the study, a household roster is completed and up to three individuals within the household are selected for the various data collection components. To maintain consistency with the NFSPS, the primary respondent for the household survey (socio-demographic characteristics, program participation and experiences, etc.) will be "the person most responsible for the finances of the household," and the respondent for the food use data will be "the person who is responsible for buying and preparing most of the meals for the household" (Cohen *et al.*, 1999). In some households, these criteria will lead to one respondent; in others, there will be two respondents. If neither of these respondents is sampled for the individual dietary intake component, a third household member will be a respondent.

Each survey wave will typically involve two in-person interviews and one telephone interview for each household. A subsample of households will be asked to complete a second telephone interview.

The first in-person interview will involve both the primary respondent and the food manager. The primary respondent will complete a brief interview that will collect data on socio-demographic characteristics; program participation, experiences and satisfaction; food security and household well-being; and all covariates except nutrition knowledge, attitudes, and practices. (Those data will be provided later by the food manager as well as by the person sampled for individual dietary intake data.) The first in-person visit will also be used to train the food manager to maintain the food use data. Materials required for this task (e.g., forms, envelopes or baggies for storing labels and receipts, and writing utensils) will be provided. In instances where the food manager is not available, instructions and materials will be reviewed with the primary respondent. A telephone follow-up will be completed with the food manager to ensure receipt of materials and understanding of data collection requirements.

The second in-person interview will take place, ideally, the day after the 7-day-food-use period is over. In this interview, the records and materials maintained by the food manager will be reviewed and the interviewer will work the food manager to complete a detailed food use record. The food manager will also complete a brief interview that will collect data on non-food expenditures and on nutrition-related knowledge, attitudes, and practices. At the conclusion of the interview, the interviewer will leave a response aid to be used in reporting portions in the 24-hour recall. One to

two days after the food-use interview is completed, the person sampled for the individual dietary intake component will complete 24-hour recall by phone. Respondents included in the subsample for the second 24-hour recall will be re-contacted a week to 10 days later. The goal will be to have the second recall cover a different day of the week than the first recall.

To summarize, the data collection includes an in-person interview with both the primary respondent (30 minutes) and with the food manager (15 minutes); a second in-person interview with the food manager (150 minutes)[19]; a 30-minute telephone interview with the sampled individual or adult proxy; and in 25 percent of cases, a second 30-minute telephone interview with the sampled individual or adult proxy. An additional 10-minute module will collect information about nutrition knowledge, attitudes, and practices. If the food manager is not sampled for individual data collection, this module will be added to the second in-person interview. If the food manager is sampled for individual data collection, this module will be added to one of the telephone interviews. We assume that incentive payments would be necessary to gain households' cooperation, especially in multiwave studies that collect this information repeatedly.[20]

Several of the preliminary studies described in Part Two of this report, as well as both proposed versions of the national study, require collection of all of this information. As noted above, we have made this assumption because we believe that it is essential that FNS have information on expenditures *and* household food use *and* individual intake in order to fully understand the impacts of the FSP. An alternative to the approach we have taken would be to build the outcomes studies sequentially. For example, an initial outcomes study could focus on household food expenditures and household food use.[21] If the results of this study establish that the FSP influences these outcomes, a subsequent study could evaluate impacts on individual dietary intake. If no impact is found in the initial study, the study focusing on individual intakes would not be attempted.

The Ideal Random Assignment Experiment

The research community has increasingly urged the use of randomized trials for evaluating program effectiveness. In their review of research designs for assessing the effects of food assistance and nutrition programs on nutrition and health, Hamilton and Rossi (2004) wrote:

> The randomized experiment is the "gold standard" of program evaluation. The scientific community is not completely unanimous on this point, but the consensus is strong enough

---

[19] This time estimate is based on the value used in planning data collection for the NFSPS, plus an additional margin (approximately 30 minutes) to allow for the collection of information on foods purchased and consumed away from home (by all household members) as well as additional information that may be needed to impute prices (e.g., type of store, type of unit (bulk *versus* pre-packaged), use of coupons or sales, and so on).

[20] We have included small incentive payments in costing all of the surveys described later in this report: $10 for RDD screenings, and $20 to $45 per wave of data collection.

[21] We see little reason to avoid assessing both household level outcomes in one study solely to reduce respondent burden. Based on existing research, we have reasonable confidence that impacts on food expenditures exist. In addition, the data required to assess household food expenditures are similar to those required to assess the quality of the household food supply.

---

that, for example, pharmaceutical companies must conduct randomized trials of new drugs in order for the products to be approved for marketing in the United States.

They concluded:

> For programs that deliver services and benefits directly to individuals and families, randomized experimentation is the only design that, properly applied, is guaranteed to produce unbiased estimates of program impact. All other designs are vulnerable to some bias. Their sources of bias can sometimes be described, but the direction and magnitude of the bias cannot be measured reliably. Thus, all the nonexperimental designs have some substantial probability of producing answers that are far from the truth—which can lead to inappropriate policy decisions that may affect millions of people and billions of dollars of public expenditure.

Equally strong language was used in the Report of the Coalition for Evidence-Based Policy to the U.S. Department of Education (2002):

> Education is a field in which a vast number of interventions, such as ability grouping and grade retention, have gone in or out of fashion over time with little regard to rigorous evidence. As a result, over the past 30 years the United States has made almost no progress in raising the achievement of elementary and secondary school students, according to the National Assessment of Educational Progress, despite a 90 percent increase in real public spending per student. Our nation's extraordinary inability to raise educational achievement stands in stark contrast to our remarkable progress in improving human health over the same time period—progress which, as discussed in this report, is largely the result of evidence-based government policies in the field of medicine.

They make clear in their report that "evidence-based policies" are those based on large-scale randomized trials. Like Rossi and Hamilton, they note that

> Randomized controlled trials are widely considered the "gold standard" for measuring the effect of a particular intervention in medicine, psychology, welfare policy, and other areas. This is because the process of randomizing subjects into either a treatment group or a control group, if properly executed (e.g., with a large enough sample size), ensures that the two groups are statistically comparable in all factors other than the intervention; therefore, the resulting difference in outcomes between the treatment and control groups can be causally attributed to the intervention.
>
> Other study designs often lead to erroneous conclusions. For example:
> a. "Pre-post" study designs, commonly used in education research, often produce seriously biased results. […]
> b. Most "comparison group" study designs, including quasi-experimental designs, also lead to erroneous conclusions in many cases.

A study design that avoids the obvious pitfalls of both pre-post and comparison group designs is the use of double differences—estimating impacts as differences between participants and nonparticipants in *changes* in outcomes from one period to the next. Hamilton and Rossi describe

this as the "strongest of the quasi-experiments". If done well this type of study can be quite persuasive. Nonetheless, its potential weakness is that outcomes for the two groups might naturally change at different rates, among other reasons because individuals typically enter programs when their lives are in some type of transition.

A randomized trial of a social program or intervention typically follows one of two patterns. One model is to recruit samples of *eligible nonparticipant households* and offer half of them the opportunity to participate, while quarantining the other half from participation for some period of time (e.g., a year). A second model would be to select a sample of *approved applicants* and randomly quarantine half of them from participation for a year. The second model has more statistical power than the first because all of the treatment group would start out as participants, rather than slowly trickling in over time.

These models, which have been used for interventions and programs with limited slots, are obviously impossible to implement for the FSP. The FSP is an entitlement program, and for both legal and ethical reasons eligible households cannot be barred from receiving benefits to which they are entitled.

Later in this report (Chapter Eight) we describe a randomized experiment to measure FSP impacts which seems to be neither illegal nor unethical. This study eliminates selection bias and achieves internal validity. The price for these advantages is however considerable. First, the external validity of the study is questionable, because it measures impacts for households that would not normally participate in the FSP, rather than for current participants. Second, the study is extremely expensive in terms of sample size, because households cannot be directly assigned to participate or not in the FSP; they can only be specially encouraged, or not specially encouraged, to do so, leading to a substantial loss of statistical power.

Selection Bias in Nonexperimental Approaches

Absent random assignment, survey designs attempt to reduce the risk of selection bias to acceptable levels through two means:

- Drawing samples of participants and nonparticipants that are similar on as many dimensions as possible, so that the outcomes of the nonparticipants can be taken to be what those of the participants would have been, absent the program;

- Collecting data on covariates (characteristics of participants and nonparticipants) that allow researchers to make statistical adjustments that, in effect, make the samples more similar.

Commonly used approaches for using information on covariates to reduce selection bias include:

1. Single equation (regression) methods. Outcomes are related to program participation and to respondent characteristics, including those that specifically tend to differ between participants and nonparticipants (i.e., determinants of participation).

2. Propensity score analysis (PSA). A program participation model is estimated first. Then outcomes are compared between participants and nonparticipants with similar

participation propensities. Respondents who are nearly certain of participating or not participating are usually dropped from the analysis because they nearly all fall in one or the other treatment group.

3. Instrumental variables analysis (IV). A program participation model is estimated first, including one or more covariates that are assumed to be unrelated to the outcome except through their effect on participation. Then outcomes are related to respondent characteristics (other than the instrument(s)) and to predicted participation.

4. Two-step (or two-stage) procedures. A program participation model is estimated first. The purpose of the first stage is to obtain a correction factor called the inverse Mills ratio which is used in the second stage to take account of possible selection bias. In the second stage, impacts are estimated from a model that includes a dependent variable, an indicator variable for participation, the selection bias correction variable (inverse Mills ratio), other independent variables, and an error term. This approach also requires data on determinants of participation that are not directly related to outcomes.

Starting with the seminal work of LaLonde (1986), a substantial literature has developed documenting the success of these methods in controlling selection bias—where success is measured by comparing impact estimates with those obtained through randomized experiments. The results have been mixed, and reasonable people may differ in their interpretation of the evidence (Smith and Todd, 2004; Dehejia, 2004; Glazerman, Levy, and Myers, 2003; Orr, Bell, and Kornfeld, 2003; Dehejia and Wahba, 2002; Smith, 2000; HRDC, 1998).

Design Factors for Reducing Selection Bias

There appears to be a consensus in the literature that, short of a randomized experiment, no single method for controlling selection bias can be expected to eliminate bias consistently under all conditions. There also appears to be some consensus about conditions under which methods for controlling selection bias would be expected to have greater success. Some of these conditions would obviously met in a national study of the FSP, and do not require further discussion here:

- The comparison group is drawn from the same local population as the participant group;
- Identical measurement instruments are used for participant and comparison groups; and
- The sample size is large.

Other conditions pertain to the data analysis, and can be tabled in the design phase:

- Specification tests are employed; and
- OLS regression and/or matching techniques are used rather than instrumental variables, selection correction, or simple mean differences

A final pair of conditions, however, has significant implications for this research agenda, and these are discussed below.

Availability of Background Data Relevant to the Program Participation Decision
Design replication studies use data from randomized experiments to estimate treatment impacts, re-estimate impacts using one or more nonexperimental methods, and then compare results from the two methods. Early design replication studies focused on the question of the choice of impact estimator,

given the available data. More recently, the focus has shifted to consideration of both data quality and choice of impact estimator. Many authors now claim that data quality is as important or more important than research design (Smith and Todd, 2004; Glazerman, Levy, and Myers, 2003; Smith, 2000; HRDC, 1998). Replication studies conclude that a superior nonexperimental design cannot substitute for the availability of good data on selection factors in helping researchers overcome the effects of selection bias. Rich data regarding the determinants of program participation seems to be indispensable.

The implication for the current study design is that, prior to conducting an impact study, serious consideration should be given to conducting a study (or studies) that would provide data that could be used to develop a strong participation model.

Availability of Pre-treatment Data and Multiple Pre- and Post-treatment Measurements
The availability of pre-treatment data can significantly strengthen nonexperimental designs. Pre-treatment data can be expected to reduce or eliminate selection bias when program participation is dependent on unobserved person-level characteristics that vary from person to person, but which remain constant within a person over time. The difference-in-differences approach has been described as particularly promising when the data are expected to conform to those conditions (Smith and Todd, 2004; Glazerman, Levy, and Meyers, 2003; HRDC, 1998). In their design replication study, Dehejia and Wahba (2002) concluded that the propensity score method they evaluated performed adequately only for the subset of observations that had two or more years of pre-treatment measures on their dependent variable. They also cite work done by Ashenfelter (1978) and Ashenfelter and Card (1985) as suggesting that (in the context of impacts on earnings) two or more years of pre-treatment earnings data are necessary to estimate a treatment effect because people who volunteer for training programs are likely to experience a drop in earnings just prior to program entry.

The research showing that two pre-treatment measurements produce more stable pre-treatment measures than a single measurement, and may even be necessary to obtain an accurate assessment of impacts, is compelling. The same argument can be easily applied to the number of measures taken during the treatment period. More measurements will provide more stable measures and will produce more accurate estimates of impact.

The implication for the current study design is that the optimal nonexperimental approach would include multiple observations of at least some households as FSP participants and eligible nonparticipants—suggesting a minimum of three or four waves of data collection.

A Test for the Importance of Omitted Factors

A useful tool for assessing the potential of omitted selection factors to bias estimates of impacts is a test described by Harding (2003). Suppose that there was some variable Z that affected both participation and an outcome, which was omitted from the analysis. How much could the omission bias the estimated parameter? To answer this question, one can compare three contingency tables. The first (Exhibit 2.2) is the observed relationship between participation and the outcome, where for simplicity the outcome has been dichotomized as "high" and "low." The entries in the cells are the counts of the population in each combination.

Exhibit 2.2

Observed Relationships

	Participation		All Eligibles
	Yes	**No**	
High outcome	A + E	B + F	A + E + B + F
Low outcome	C + G	D + H	C + G + D + H
ALL	A + E + C + G	B + F + D + H	A through H

The other two tables (Exhibits 2.3 and 2.4) are the underlying latent tables that add up to Exhibit 2.2, for "high" and "low" values of the omitted variable Z.

Exhibit 2.3

Low Value of Z

	Participation		All Eligibles
	Yes	**No**	
High outcome	A	B	A + B
Low outcome	C	D	C + D
ALL	A + C	B + D	A + B + C + D

Exhibit 2.4

High Value of Z

	Participation		All Eligibles
	Yes	**No**	
High outcome	E	F	E + F
Low outcome	G	H	G + H
ALL	E G	F H	E + F + G + H

From the observed values in Exhibit 2.2, we know the values of A+E, B+F, C+G, and D+H. We then postulate two more facts: that the effect of Z on the outcome (expressed in odds-ratio terms) is M, and the effect of Z on the likelihood of getting the treatment is N. We furthermore assume that the effect of Z on the outcome is the same for both treatment group and control group members (T's and C's), and the effect of Z on the likelihood of getting the treatment is the same for both high- and low-outcome individuals. These assumptions give us the following:

$E/G \div A/C = M$ (effect of Z on the outcome for T's)
$F/H \div B/D = M$ (effect of Z on the outcome for the C's)

$E/F \div A/B = N$ (effect of Z on participation for high-outcome individuals)
$G/H \div C/D = N$ (effect of Z on participation for low-outcome individuals)

That enables us to solve for the unobserved values of A through H, conditional on what we assume about M and N. We then infer the "real" effect of treatment on outcome (i.e., if we had data on Z) as $A/C \div B/D$ (which as can be seen from the equations above is equal to $E/G \div F/H$), rather than our previous biased estimate from Table 1 of $((A+E)/(C+G)) \div ((B+F)/(D+H))$.

To apply this method, we suppose that the omitted variable Z increases participation by, say, 5 percent, and that it also increases the likelihood of a good outcome by 5 percent. Benchmarks for these assumptions could come from examination of included variables; we could argue that the omitted variable was unlikely to be *more* influential than some variables already in the models. Using values of M=1.05 and N=1.05 (or whatever values were chosen) would place bounds on the potential

remaining selection bias. This procedure was applied by Abt staff in an evaluation of participation in AmeriCorps (Jastrzab *et al.,* 2004).

How Good Is Good Enough?

Heckman and Smith (1995) marshaled several arguments favoring econometric analysis of quasi-experiments over randomized experiments. Two of these arguments centered on aspects of external validity. First, randomization may alter the pool of participants. Those willing to participate in a randomized experiment may, in fact, form such a special population that conclusions about program impacts on this group may be of little practical interest. Second, because the control group may receive services that vary widely, depending on the environment, impact estimates have little meaning outside the study sites. And if the experiment makes these services either more or less available to control group members than to the population as a whole, the impact estimates may not even be meaningful within the selected sites. Depending on the intervention and the evaluation design, study staff sometimes offer control group members information on how to obtain alternative services, effectively reducing the difference between treatment and control group members in services received. Alternatively, if only a fixed pool of services is available in a community, which is monopolized by treatment group members, control group members may have fewer options than they would have, absent the experiment.

Heckman and Smith further argue that the replication studies of Lalonde and others were not a fair test of the ability of econometric models to overcome selection bias because they made limited use of information on covariates: "The most convincing way to solve the selection problem is to collect better data." Finally, they note that randomized experiments can answer only a narrowly focused question of whether a program had an impact given how it was implemented, rather than adding to cumulative knowledge about the effects of program variations.

A reader might draw the following conclusions:

1. Randomized experiments, if conducted properly, achieve internal validity. Their external validity, i.e., generalizability to the full population, must to some extent be taken on faith. Impacts may vary widely among sites because of differing availability of alternatives, and even within a site, the sample may not be representative of the population of interest. Furthermore, by their nature, they measure the impacts given that the program exists, but not the impacts of the program on nonparticipants.

2. Impact studies using participant-nonparticipant comparisons, in contrast, can be designed to achieve national representativeness of the eligible population. But, here, internal validity must to some extent be taken on faith. Any previously postulated source of selection bias can be addressed by better data collection. One might still be skeptical of the results, however, because:

   a. Participation models generally leave a large amount of variance unexplained, and even a small selection effect might swamp the true program impact.
   b. Replication studies do not give assurance that statistical methods can compensate for selection.
   c. Non-experimental impact studies of FNS programs find "wrong" results sufficiently often (e.g., FSP participation increasing food insecurity, WIC participation decreasing birthweight) that doubt is cast on results that go in the hypothesized direction.

It is thus a live question whether, absent an experiment, the risk of selection bias can be reduced "to acceptable levels." One TWG member previously commented:

> Remember that we have to be confident that selection bias is almost completely controlled in order to have confidence that a model accurately estimates the ameliorative effect of the program. Just reducing selection bias by some unknown amount is not good enough.

Nonexperimental impact studies are still common in many research environments. Whether the results of such a study of the FSP would be credible depends on the audience. A stronger design will convince more readers than a weaker design.

Criteria for a Successful Impact Evaluation

Two conditions must be met for a national nonexperimental study of FSP impacts to be worth attempting. The purpose of the research agenda described in the next chapter is to provide information on whether these conditions are met.

1. FNS must be confident that a national nonexperimental study will generate impact estimates that are not invalidated by selection bias.

2. FNS must have reason to believe that increasing food expenditures improves the quality of individual dietary intake or household food security, at least among FSP participants.

If both conditions are met, the impact evaluation can be justified.

If either one of the two conditions is not met, an impact evaluation is not worthwhile, but the preliminary research steps may still provide valuable information.

If Criterion 1 is not met, but Criterion 2 is met, there is reasonably strong evidence that the FSP has beneficial effects. The effects of FSP participation on food expenditures are well established. Consequently, the preliminary studies indicate that increased food expenditures generally result in improved dietary intake and/or food security, one might reasonably infer that the FSP has positive effects on these outcomes.

If Criterion 2 is not met—assuming that sample sizes were sufficient to detect effects if they existed, and the preliminary studies were otherwise appropriately designed—the implication is that the FSP as currently constituted is not a sufficiently intensive intervention to improve the distal outcomes of interest. FNS might logically focus its energies on changing the nutrition education component of the program, which is beyond the scope of this research agenda. This alternative is explored further in Chapter Four, under the heading of "Changing the Research Question."

Chapter Three: Research Program for an Evaluation of the FSP

In this chapter we describe the overall research strategy that culminates in either (a) a national study of FSP experiences, satisfaction and impacts, or (b) a national study limited to FSP experiences, satisfaction, and outcomes. The overall research agenda comprises three components:

1. ***Approaches for addressing selection bias*** in estimating impacts of the FSP. We propose five individual studies that test three approaches for addressing bias. These are:

    a. Estimation of participation models using extant data (Study 1).
    b. A pair of studies that respectively generate hypotheses about the determinants of participation and test our ability to predict it (Studies 2 and 3).
    c. A paired randomized experiment and replication that test whether a well-designed quasi-experiment can consistently match the experimental results (Studies 4 and 5).

2. ***Approaches for testing the links in the causal chain between food expenditures*** (which the FSP is confidently believed to affect) ***and diet quality*** (for which evidence of FSP impacts is weak at best). We propose four individual studies that test three approaches for estimating the links. These are:

    a. Exploration of relationships between outcomes using extant data (Study 6).
    b. A pair of studies that respectively generate hypotheses about these interrelationships and model the relationships (Studies 7 and 8).
    c. An experiment that uses food vouchers and perhaps nutrition education to increase food expenditures and/or improve diet quality among low-income households (Study 9).

3. ***A national study of FSP experiences and outcomes whose design may vary in two ways depending on the results of the nine preliminary studies.*** These are:

    a. Exclusion or inclusion of an impact study component (Studies 10 and 11).
    b. A broader or narrower range of outcomes studied.

The relationship among the studies is shown in Exhibit 3.1. The overall research strategy is as follows. Two separate strands of research lead to the final design for the national FSP survey. In one strand, preliminary studies increase our understanding of how to address selection bias, or possibly persuade us that the selection bias problem is intractable. Based on the conclusions drawn from these studies, FNS chooses a national study that includes or excludes an impacts component. In the second strand, other studies increase our understanding of the relationships among outcomes of interest, including household food expenditures, household food security, quality of the household food supply, and individual dietary intake. The conclusions drawn from these relational studies will also contribute to the decision by FNS whether to conduct an impact study. If impacts on the distal outcomes are not expected, an impact study is of little interest. If impacts are expected, then the

Exhibit 3.1: Proposed Research Agenda
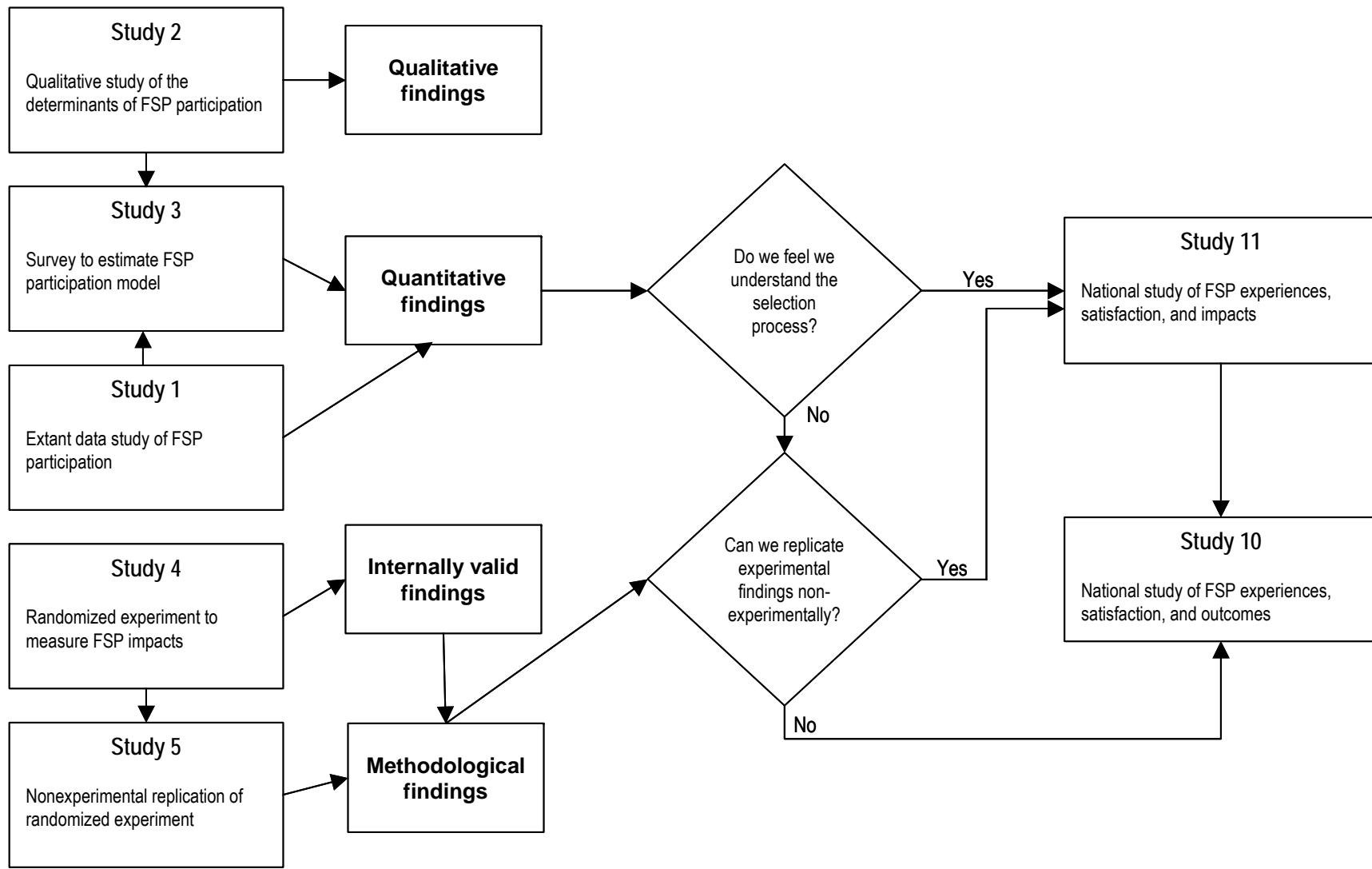
Exhibit 3.1: Proposed Research Agenda *(continued)*

Study 7

Qualitative study of diet-related implications of changing food expenditures

→ **Qualitative findings**

Study 8

Survey of food expenditures and diet-related outcomes

→ **Quantitative findings**

Study 6

Study of outcomes in extant data

Study 9

Food expenditures experiment

→ **Internally valid findings**

Do we believe household food expenditures → household food supply?

No → Study 10

Yes ↓

Do we believe household food supply → individual diet quality?

No → Analyze household food expenditures and household food supply only

↓ Study 10   Study 11

Yes ↓

Analyze household food expenditures, household food supply, and individual food intake

↓ Study 10   Study 11

preliminary studies will provide information on how best to measure some or all of these dimensions for inclusion in the national study, and on the necessary sample sizes for detecting effects.

In the sections that follow, we describe the alternative versions of the national FSP study (Studies 10 and 11), and sketch out the approaches used in the preliminary studies to address selection bias (Studies 1 through 5) and to explore relationships among outcomes (Studies 6 through 9). Each of these studies is described in detail in a subsequent chapter. We then discuss the time and resource requirements for the overall research program, and propose three alternative versions of the agenda that reduce resource requirements by omitting some components.

National Studies of the Food Stamp Program

The NSFSP collected information on FSP *experiences* from both participants and eligible nonparticipants, and information on *outcomes* (household food expenditures and food use) for participants. At a minimum, the new national study will collect comparable information on experiences from both participants and eligible nonparticipants. At FNS's option, the study may also collect information on outcomes from participants and nonparticipants, and may include an impact estimation component.

Study 11: National Study of FSP Experiences, Satisfaction, and Impacts

The research questions for this study are:

- What is the impact of the FSP on household food expenditures, diet quality, food security, and household well-being?
- What are FSP-eligible households' experiences and satisfaction with the FSP, WIC, and school meals programs?

A multiwave survey will be used to collect longitudinal information on outcomes of interest among participants and eligible nonparticipants. Comparisons of these outcomes between the participant and nonparticipant groups will yield FSP impact estimates. Information on program experiences and satisfaction will be collected in the first wave.

Study 10: National Study of FSP Experiences, Satisfaction, and Outcomes

If impacts are not to be evaluated, then Study 10 will be carried out as a fallback. The research questions are:

- What are FSP-eligible households' experiences and satisfaction with the FSP, WIC, and school meals programs?
- How do FSP participants compare with eligible nonparticipants in terms of household food expenditures, diet quality, food security, and household well-being?

Information will be collected on program experiences, satisfaction, and outcomes among participants and eligible nonparticipants in a single-wave survey.

Specifications for Studies 10 and 11 appear in Chapters Fourteen and Fifteen, respectively.

Approaches to Addressing Selection Bias

This research program includes several approaches for studying and assessing the extent of selection bias in an FSP impact study: analyzing extant longitudinal data to explore participant decisions (Study 1); a qualitative study of determinants of participation followed by a survey-based study that tests the hypotheses generated (Studies 2 and 3); and a comparison of narrowly focused experimental and nonexperimental estimates of FSP impacts (Studies 4 and 5).

Study 1: Extant Data Study of FSP Participation

Study 1 (described in Chapter Five) will use data from longitudinal surveys to attempt to explain FSP participation decisions. The research questions are:

- What are the determinants of FSP participation among eligible households?

- How well can econometric models *using extant data* classify eligible households as participants and nonparticipants?

Study 1 could conceivably generate a sufficiently good model that no further research on participation or nonexperimental methods would be deemed necessary. Determining a benchmark for "sufficiently good" is clearly a judgment call, based on thoughtful consideration of excluded factors and their likely importance relative to included factors. A more likely outcome is that the best model will still have a significant amount of unexplained variation. Assuming that this is the case, FNS would want to develop better models of participation (Studies 2 and 3) or demonstrate that valid impact estimates could be derived despite the remaining gaps in our understanding of participation (Studies 4 and 5).

Study 2: Qualitative Study of the Determinants of FSP Participation

Study 2 (described in Chapter Six) uses in-depth retrospective interviewing with relatively small numbers of households to generate hypotheses about FSP participation. Because of its qualitative nature, this study can go beyond the "usual suspects" found in general-purpose surveys to explore the events, circumstances, attitudes, and perceptions that come into play as households decide whether to apply for or to continue receiving food stamps. The research question is:

- What are the determinants of FSP participation among eligible households?

The study will have two products. The first will be a collection of detailed stories about individual low-income households that identifies as many factors as possible that seem related to participation behavior. The second will be a synthesis of the patterns in the individual stories to generate hypotheses as to which variables are the best candidates to predict participation and how they interact with each other. The study can be expected to enrich our understanding of how people decide to participate in the FSP, a critical piece of information for the ultimate decision of whether to implement Study 10 or 11.

Study 3: Survey to Estimate FSP Participation Model

Study 3 (described in Chapter Seven) uses a panel survey with multiple waves of data, conducted in a limited number of sites. Its research questions are:

- What are the determinants of FSP participation among eligible households?

- How well can FSP participation be predicted *using a specially designed survey*?

Study 3 provides an opportunity to test hypotheses regarding FSP participation that have been generated in Studies 1 and 2. While a secondary objective of Study 3 is to add to our knowledge about the determinants of participation, its primary objective is to see if it is possible to understand the selection process, and, therefore, move on to a large-scale national study of program impacts, using specially designed survey items. Studies 3 and 8 could potentially be combined, as discussed in Chapter Four. While this could reduce total costs, it would come at the expense of some compromises in both study designs and increasing the respondent burden.

Study 4: Randomized Experiment to Measure FSP impacts and Study 5: Nonexperimental Replications of Randomized Experiment

Studies 4 and 5, described in Chapters Eight and Nine respectively, comprise a randomized experiment designed to measure FSP impacts in special circumstances, coupled with a nonexperimental study in the same sites and during the same time frames. The research questions are:

- Study 4: If eligible nonparticipants can be induced to participate in the FSP, what is the impact of FSP participation on household food security, household food expenditures, the quality of the household food supply, household food expenditures, and individual dietary intake?

- Study 5: Can nonexperimental methods produce valid impact estimates?

It is not legally or ethically acceptable to deny food stamp benefits to eligible households for the purposes of an experiment. In Study 4, therefore, experiments are based on interventions that reduce barriers to FSP participation, e.g., by providing assistance in the application process or lengthening the certification period. The experiments are conducted on a small scale in purposively selected sites.

The replication in Study 5 will attempt to measure FSP impacts using a quasi-experimental design that incorporates knowledge about selection acquired from Studies 1 through 3. If the replication consistently duplicates the patterns of results observed in the experiments, this would provide evidence that Study 11, a national survey-based impact study, is worth doing. A significant discrepancy would be a strong argument against implementing Study 11.

Understanding Relationships Among Outcomes

The second branch of the research agenda explores relationships between the outcomes of interest to help guide the design of Study 10 or 11.

Study 6: Study of Outcomes in Extant Data

The goal of Study 6 (discussed in Chapter Ten) is to learn about the relationships between the outcomes in the hypothesized causal chain, as shown in Exhibit 2.1 on page 6. Ideally, we would like to explore all the hypothesized links. As discussed in Appendix A, however, available extant data only support exploration of two links in the chain: the relationship between household food expenditures and the quality of the household food supply and the relationship between household food expenditures and food security.

The relationship between food expenditures and quality of the household food supply will be explored using data from the NSFSP and CES. The research questions are:

- Among low-income households, is an increase in food expenditures associated with
  - an increase in household nutrient availability?
  - an increase in the nutritional quality of foods used at home?

- What factors or household characteristics mediate these relationships?
  - Is the relationship between food expenditures and the quality of the household food supply different for FSP households than for other low-income households?

Study 7: Qualitative Study of Diet-Related Implications of Changing Food Expenditures

Study 7, described in Chapter Eleven, will conduct repeated in-depth interviews with a small group of households over an extended period. The objective is to observe and learn about patterns of food expenditures, diet, and food security before and after external shocks (e.g., FSP entry or exit, job loss or gain, new family member, illness, vehicle breakdown) and at varying points in the relevant cycles (e.g., early and late in the month). The research questions are:

- How and why do households change their food purchasing patterns and diet when they increase or reduce their food expenditures?
- What factors influence the household in changing its level of food expenditures, and how are differing *reasons* for change associated with different *consequences* of the change?

As is the case for Study 2, the study will have two products. The first will be a collection of detailed stories about individual low-income households. The stories will describe processes that occur as households change their food expenditure patterns, including those that lead to the changed expenditures, and those through which changed expenditures affect dietary patterns and food security. The second will be a synthesis of the patterns in the individual stories to generate hypotheses about relationships among the processes and diet-related outcomes, and develop appropriate measures.

This qualitative study will support the design of a subsequent study of the quantitative relationships among income (including FSP benefits), needs, food expenditures, household and individual dietary patterns, and food security (Study 8). It is expected to enrich substantially the understanding of the connection between low-income households' food expenditures and their dietary patterns, and to help provide guidance in designing Study 10 or 11.

Study 8: Survey of Food Expenditures and Diet-Related Outcomes

Study 8 (Chapter Twelve) uses primary data from a multiwave survey to test hypotheses about the relationships among food expenditures, short-term food security, and diet quality. Like Study 3, this study need not be based on a nationally representative sample, as long as there is sufficient variety among the locales to cover relevant situations. The research questions are:

- To what extent are greater food expenditures among low-income households associated with better diet quality?
- If food expenditures are not closely related to diet quality, what are the results of higher expenditures?
- To what extent are greater food expenditures among low-income households associated with greater food security?

If only weak relationships are found among the various outcomes, the study will clarify why the FSP has not been found to have strong impacts on diet-related outcomes in the past, will temper expectations for future impact studies, and will support increased attention to nutrition education in the FSP. If strong relationships are found, the study will support collection of data on these distal outcomes in a national study.

Study 9: Food Expenditures Experiment

Study 9 (described in Chapter Thirteen) is an experiment focused on the relationships *between* potential FSP outcomes (increased food expenditures and changes in diet quality and/or food security) rather than on the impact of the FSP on these outcomes *per se*. Low-income households will be randomly assigned to receive vouchers that can be used only for food at home. The vouchers will need to be of sufficient value and provided over a sufficient time period to have a perceptible impact on food expenditures. The effects on a variety of distal outcomes can then be examined. An attractive variant of the experiment would be to offer *intensive nutrition education* in addition to the food vouchers to a random subset of the treatment group. The research questions are:

- What is the impact of increased food purchasing power on
  - food expenditures
  - quality of the household food supply
  - individual intake
  - food security?
- How do these impacts differ in the presence of intensive nutrition education?

Sequencing of Studies

It is not necessary that all of the preliminary studies be implemented. Those that are selected should be conducted in a particular order, however, so that later studies can take advantage of the findings of earlier studies. The key relationships are as follows:

1. The studies on selection bias include three ***participation*** studies: Study 3 (survey), Study 1 (extant data) and Study 2 (qualitative). Study 3 should be preceded by both Studies 1 and 2, and Study 1 should also precede Study 2. The reason is that the survey, which collects data for estimating models, should incorporate in its design all possible information on determinants of participation; and that the new qualitative study which feeds into it should make use of findings from previous research in framing its questions.

2. The studies on selection bias also include paired ***experiments and replications*** (Studies 4 and 5). These two studies should happen concurrently to maximize the replications' chance of success. They should be preceded by Studies 1 through 3, so that the replication design can be based on the best possible model of selection.

3. The outcomes studies on selection bias include three ***nonexperimental outcomes studies***, Study 8 (survey), Study 6 (extant data) and Study 7 (qualitative). Study 8 should be preceded by both Studies 6 and 7, and Study 6 should precede Study 7. The logic is the same as for Studies 1 through 3: the survey design should be guided by what is known about the relations among outcomes, and the new qualitative study should take advantage of the findings from previous research.

4. Study 9 (food expenditures experiment) does not depend temporally on any of the others.

It is not necessary for the participation studies to precede the outcomes studies. A connection between program participation and diet quality and/or food security could be strongly inferred if the outcome studies demonstrate that increased food expenditures improve diet quality and/or food security, even though causation is not directly established. Therefore, even if the paired experiments and replications suggest that selection bias is intractable and that an impact evaluation should not be attempted, it is still worthwhile to carry out the outcomes studies.

Assuming that all the preliminary studies will be conducted, two critical paths lead to Study 10 or 11, as illustrated in Exhibit 3.2. One path runs in sequence through participation studies 1, 2, 3, and 4 and 5. The second path runs in sequence through outcomes studies 6, 7 and 8. The two groups of studies can be conducted at the same time, and the food expenditures experiment (Study 9) can be conducted at any time, independently of the other two sequences.

Exhibit 3.2

Proposed Sequence of Studies
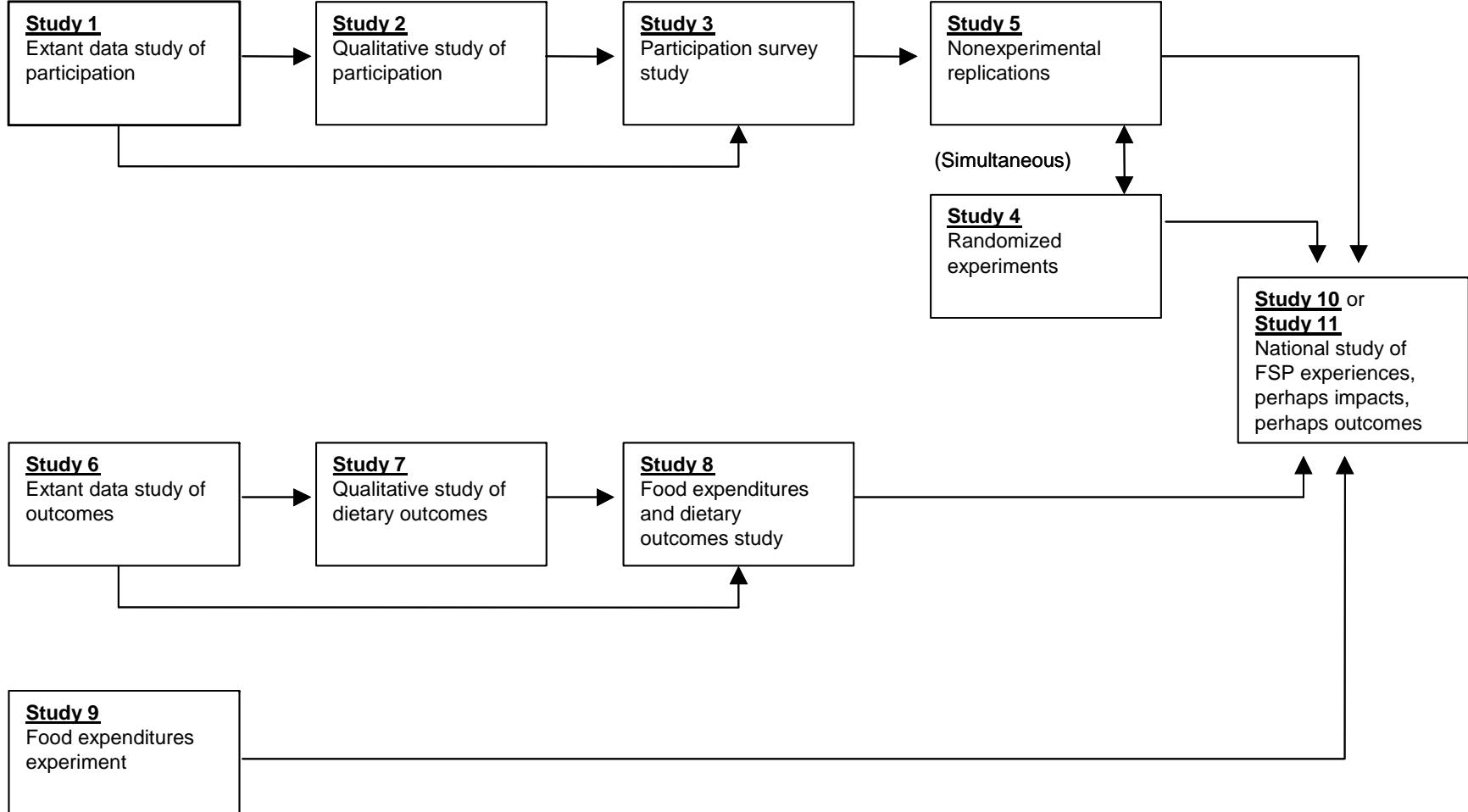
Note: Arrows show links between results of earlier studies and design of later studies.

Chapter Four: Costs and Options for the Research Agenda

We estimate that the research program outlined in Chapter Three would require a period of about 12 years and expenditures of about $31 million for the nine preliminary studies, plus an additional 4 to 5 years and $5 to $10 million for the final national study.

An evaluation research agenda of this scale would by no means be unprecedented for the Federal government. For example,

- The Income Maintenance Experiments conducted from 1970 to 1978 cost nearly $80 million in 1975 dollars.

- The Housing Allowance Experiments conducted in the mid-1970s cost over $30 million in 1976 dollars.

- The Institute of Education Sciences is currently spending about $26 million on evaluating Even Start and about $28 million on an evaluation of Reading First

- The Adminstration for Children and Families is spending about $24 million on evaluating Head Start and $10 million on evaluating subsidy policy experiments.

- The Social Security Administration initiated two experiments in 2005 evaluating potential policies for Social Security Disability Insurance (SSDI) program, awarding separate research contracts with an estimated value of $30 and $45 million.

Moreover, even if the total cost of an FSP evaluation substantially exceeded $40 million, it would still be far less than 1 percent of the Food Stamp Program's annual budget.

Nonetheless, we recognize that time and money commitment required for the recommended research program be extraordinary for FNS. This chapter therefore discusses the reasons for the heavy resource requirements and considers some alternative strategies that would reduce the cost and/or shorten the time to complete the program. We note at the outset, however, that all of these strategies would increase the risk that the research would fail to obtain conclusive evidence of the presence or absence of FSP impacts.

Time and Resource Costs of the Research Agenda

Time and resource costs of the nine preliminary studies and two alternative national studies are summarized in Exhibit 4.1. More detailed discussion of the individual studies is presented in Chapters Five through Fifteen.

Study durations range from one year to five years. The briefer studies—one to two years—are those based on extant data analysis or qualitative research. The studies involving new surveys generally take four to five years.

The extant data and qualitative research studies, with costs estimated at $200,000 to $1 million, are also much less costly than the studies with new surveys. The survey-based studies have estimated costs ranging from $2.4 million to $10.2 million.

Exhibit 4.1

Estimated Time and Resource Costs of Preliminary Studies

Study	Duration (years)	Cost
Study 1: Extant data study of FSP participation	1.0	$200,000
Study 2: Qualitative study of the determinants of FSP participation	2.0	700,000
Study 3: Survey to estimate FSP participation model	4.2	6,200,000
Study 4: Randomized experiment to measure FSP impacts	4.0	10,200,000
Study 5: Nonexperimental replication of randomized experiments	5.0	2,400,000
Study 6: Study of outcomes in extant data	1.5	400,000
Study 7: Qualitative study of diet-related implications of changing food expenditures	2.0	1,000,000
Study 8: Survey of food expenditures and diet-related outcomes	4.7	7,200,000
Study 9: Food expenditures experiment	3.5	2,900,000
Subtotal		*31,200,000*
Study 10: National study of FSP experiences, satisfaction, and outcomes	4.0	4,800,000
Study 11: National study of FSP experiences and impacts	5.0	10,000,000

Time and cost estimates were developed using the procedures that Abt Associates normally uses in preparing proposals for federally-funded research. This involves considering the sequence of tasks that must be performed, defining a staffing plan, estimating the time requirements for each staff member, and estimating non-personnel costs such as travel and computer. Standard Abt Associates rates for salaries and indirect costs were applied in calculating total costs. Each study's costs were estimated independently except the Nonexperimental Replication (Study 5), which can only be conducted in conjunction with the Randomized Experiment (Study 4). First-year costs for each study were based on 2005 salary levels, with increases in subsequent study years to reflect inflation.

Although the estimation procedures were careful, the resulting estimates must be considered first approximations. "Real" estimates would require much more detailed definition of research objectives and scope. These definitions will be influenced not only by the results of previous studies in the sequence, but by policy context at the time when FNS is initiating any given study and by substantive and methodological developments in the relevant fields.

Exhibit 4.2 shows the overall time requirement for the nine preliminary studies, assuming that the studies are conducted in the sequence described in Chapter Three. The series of participation studies takes 12 years to complete. The outcomes studies, which are conducted independently of the participation studies, are completed four years earlier. When the program of preliminary studies is completed, FNS will decide which of the two national studies to conduct. Since these studies take 4 and 5 years, respectively, completing the full study program takes 16 or 17 years.

Exhibit 4.2

Timeline for Preliminary Studies
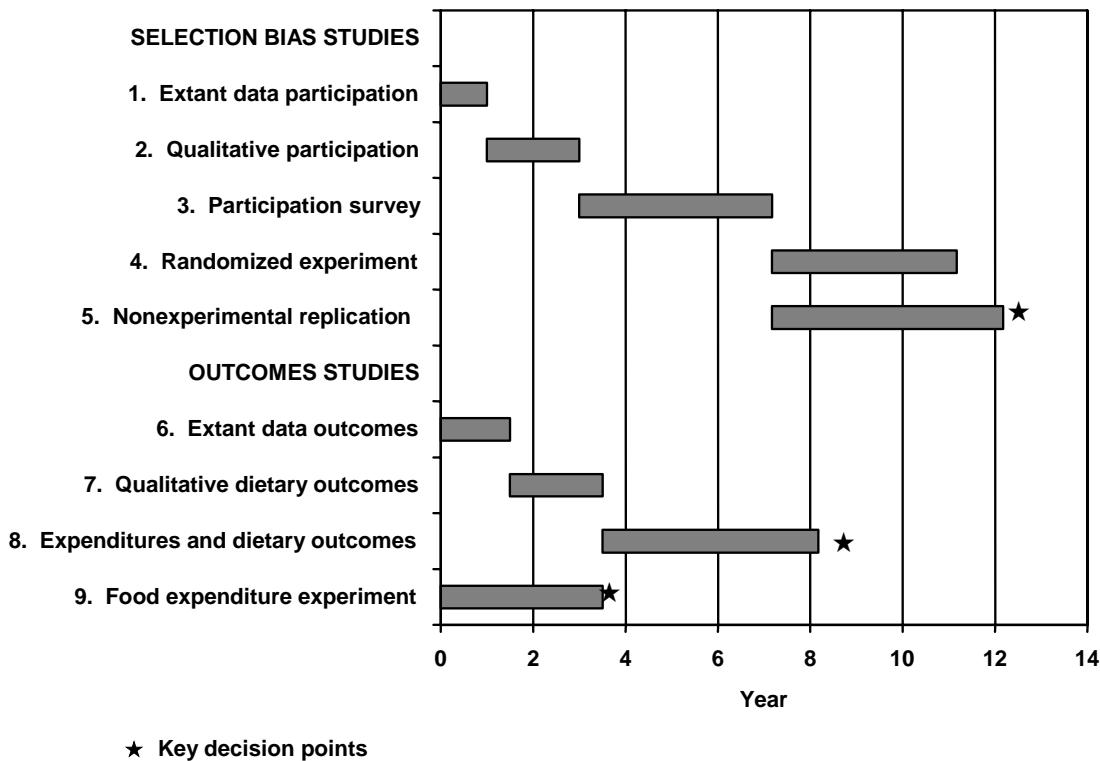
Why Are the Studies So Expensive?

These studies are quite expensive, individually and collectively. This is in part the price of not being able to use a random assignment evaluation approach. Although we have not attempted to estimate the cost of a national random assignment evaluation, it is easy to imagine a design that would cost well under half of the $40 million potential total. If random assignment were possible, it would not be necessary to conduct any of the nine preliminary studies. Selection bias would not be an issue, and the impact estimates for the various outcomes would be conclusive without a detailed understanding of the relationships among them. A single random assignment study would doubtless be a large undertaking, and two or more studies might be needed, but the total price tag would undoubtedly be much lower and the results available much sooner than with the program envisioned here.

Most of the individual studies—all of the studies involving new survey data collection—are estimated to cost several million dollars. These relatively high costs are driven by four factors: the intensive procedures required for measuring diet-related outcomes; multiple survey waves; random-digit-dialing surveys to identify study participants; and relatively large sample sizes. These four issues are described in turn.

Intensive outcome measurement. All of the survey-based studies except the participation survey are designed to measure household food expenditures, household food use, individual dietary intake, food security, and household well-being. While the last two of these are based on relatively short batteries of survey questions, data in the other three domains are expensive to collect and analyze. As described in Chapter Two, each wave of data collection from a household requires two lengthy in-person interviews and one or two by telephone, and as many as three different people in the household may be involved. After the surveys are completed, coding and processing the nutrient data for the household food supply and individual diets is also very labor-intensive and costly. As a result, the cost per completed survey is vastly more—on the order of ten times as much—as the cost for a single telephone interview.

Multiple survey waves. Nearly all of the survey-based studies envision multiple waves of data collection.[22] When individuals' outcomes can be measured on multiple occasions, the analysis can adjust for characteristics that differ across individuals but are unmeasured, such as preferences for particular foods. This strengthens the impact estimates from the two experiments (Studies 4 and 9), each of which is designed with two waves of outcome measurement. In the two studies designed to assess FSP impacts through nonexperimental designs (Studies 5 and 10), and in the Participation Survey (Study 3), it is important to observe as many individuals as possible both while they are participants and while they are eligible nonparticipants. Ideally, households in these studies would be observed at least twice in each status. Accordingly these studies have four or five waves.

Random-digit-dialing surveys. All but one of the studies that include surveys are expected to survey eligible nonparticipants or other low-income households as well as FSP participants.[23] Because nonparticipant households cannot be sampled from any existing list, an RDD survey is the least expensive way to draw a probability sample. The RDD survey is quite costly even so, because 85 to 90 percent of the households reached will not meet the criteria for inclusion in the study. (In the Participation Survey, which calls for stratification of particular subgroups, the screen-out rate is even higher.) As a result, the RDD component alone costs over $1 million in most of the studies.

Large samples. The initial sample sizes for most of the studies range between about 2,000 and 4,000 households. These sample sizes are not extraordinary, but they are quite substantial when considering that intensive data collection must occur over multiple waves. The sample sizes generally increase when smaller effects must be detected or when more waves of interviews will occur (because attrition occurs between waves).

The Randomized Experiment to measure FSP impacts (Study 4) requires a substantially larger sample, comprising about 7,200 households. This larger sample size requirement stems from the fact, although households will be randomly assigned to treatment and control groups, not all treatment group members will participate and some control group members will do so.

---

[22] The exception is the National Study of FSP Experiences, Satisfaction, and Outcomes (Study 10), which is a single-wave survey of participants and nonparticipants.

[23] The exception is the Expenditures Experiment (Study 9), for which the suggested design involves FSP participants only.

This study randomly assigns eligible nonparticipant households to receive an intervention that is anticipated to increase FSP participation rates markedly. Suppose that we wish to detect a 5 percentage point impact of the program on the likelihood of a "poor quality" diet (based on HEI scores) for *FSP participants*. We believe that the most aggressive assumption that can plausibly be made about the impact on participation is that, over a four month period, the intervention would cause a *50 percentage point differential* in FSP participation between treatment and control group members—the difference between 5 percent entering the program in the control group, and 55 percent doing so in the treatment group. To detect the aforementioned 5 percentage point impact on participants, we need to be able to detect an impact on the entire *treatment group* of only 2.5 percentage points.[24] Halving the impact to be detected multiplies the required sample size by four. That is, we need four times as much sample to detect a 5 percentage point impact on FSP participants in the Study 4 design, as we would need in a if we could directly assign households at random to receive food stamps or not.

Possibilities for Reducing the Time and Money Requirements

The remainder of this chapter considers ways that the research agenda might be modified to reduce costs and/or shorten the timeline. Before that discussion, however, we wish to emphasize that the agenda we have described thus far does not reflect a "blue sky" approach. Recognizing that the costs could be prohibitively high, we have tried to limit the research designs to the minimum that would allow the research to be reasonably convincing to most of the professional community. Other researchers might argue that we have applied too much restraint on particular points, and indeed some of the designs leave us less than fully satisfied. Three examples may illustrate the point.

The first example involves the Randomized Experiment and Nonexperimental Replication (Studies 4 and 5). Comparing the results of these two studies is supposed to indicate whether a nonexperimental approach to a national impact study would be feasible. The design involves only one pair of studies, however, conducted in a single site. We had initially envisioned a set of three pairs of studies, but this would nearly triple the estimated $12.6 million cost for Studies 4 and 5, adding around $20 million to the total. The design therefore assumes that comparing impact estimates across multiple outcomes and multiple subgroups within a single pair of studies will yield sufficient information to judge the reliability of the nonexperimental approach. It is a more aggressive assumption than we would like.

The second example concerns the many population subgroups that are of special interest within the FSP: families with children, the elderly, individuals living alone, the working poor, households under the poverty line, and so on. Sample sizes are not set at levels that would support separate estimation of subgroup impacts. Increasing sample sizes to allow separate subgroup estimates would add a major increment to the cost of the research agenda.

---

[24] The impact is zero for the 45 percent of treatment group members who do not enroll in FSP, and is also zero for the 5 percent who enroll and whose control group counterparts also enroll. Hence only for the remaining 50 percent of the treatment group do we find the FSP effect of 5 percentage points. The effect for the entire treatment group is $(0.45 \times 0) + (0.05 \times 0) + (0.50 \times 5$ percentage points), or 2.5 percentage points.

The third example relates to the exclusion from the national study of households without telephones, especially the homeless. As noted in Chapter Fourteen, about 15 percent of FSP participants do not have land-line telephones. Adjustments to the telephone screener may be possible to reach the 6 percent that have cell phones. The 9 percent with no telephones could only be reached by drawing an area probability sample and conducting door-to-door interviews—a very expensive procedure.

In addition to these design compromises that save money, we have also been aggressive in our assumptions about the speed with which the preliminary studies could be implemented sequentially. We have scheduled each to start as soon as its necessary predecessor has ended. In reality, substantial time might elapse between successive studies as FNS considers the findings, writes and issues a new RFP, and awards a contract for the next part of the research agenda.

Decision Points within the Research Agenda

If the research agenda is implemented along the lines suggested, each study that is conducted will provide some information about what needs to be done in future studies. Three points stand out, however, as occasions for FNS to make major decisions about the remaining agenda. These decision points are indicated in Exhibit 4.2 and described below.

One major decision point occurs *at the completion of the Randomized Experiment and the Nonexperimental Replication* (Studies 4 and 5). Based on these studies, FNS will decide whether the nonexperimental approach appears sufficiently reliable—that is, sufficiently immune to selection bias—to proceed with the National Experiences and Impacts study (Study 11). If selection bias still poses a major threat to the credibility of study results, FNS will presumably proceed with the National Experiences and Outcomes descriptive study (Study 10).

To reach this decision point, FNS will have invested 12 years and about $20 million in the five studies attempting to deal with selection bias. Even if selection bias proves intractable, however, the studies will have produced a great deal of substantive knowledge about the situations in which households participate in the FSP and the factors that allow them to cope without participating.

The most important information might come from the Randomized Experiment. This experiment will determine the impact of the FSP on households who, though eligible, would not normally participate. One would expect the FSP to have less impact on these households than on "normal" participants— that is households who participate in the absence of the special intervention—because nonparticipating households presumably see less value in the program. The experimental results may thus be seen as a lower-bound estimate of the program's impacts on normal participants. Finding significant impacts nonparticipants would create a strong presumption that the FSP has similar or greater impacts on participants. If this study does not find significant impacts, however, no conclusion can be drawn about the program's impact on normal participants.

Within the set of studies examining relationships among outcomes, the earliest decision point occurs *after completing the Expenditures Experiment* (Study 9). This experiment randomly assigns some FSP households to receive a modest increase in food purchasing power and measures their food expenditures, household food supply, and individual dietary intake. If this study shows that increased purchasing power has clear positive effects on all of these outcomes, it will eliminate the need to

conduct the study of Expenditures and Dietary Outcomes (Study 8) before proceeding with the National Experiences and Impacts study (Study 11). Study 8 measures the extent to which increments in food expenditure (which are expected to result from FSP participation) translate into changes in household food supply and individual diet. If the Expenditures Experiment provides good estimates of these relationships, Study 8 is unnecessary.

The end of the Expenditures Experiment also marks a methodological decision point. The experiment provides an early test of the feasibility of measuring household expenditures, household food supply, and individual dietary intake in a single study. Nearly all of the subsequent studies (Studies 4, 5, 8, 10, and 11) are designed to use this measurement approach, which entails an extraordinary respondent burden and a substantial logistical challenge as well as very high operating cost. If the respondent burden and operational complexity make it impossible to obtain reliable data, these study designs and perhaps the overall research agenda will need major reassessment and revision.

The Expenditures Experiment can be considered to yield a lower-bound estimate of the potential effects of the FSP. The suggested increment in purchasing power—about $30 per person per month—is smaller than the average food stamp benefit. And since it is added to the FSP benefit, which is designed to meet the household's basic food needs, the experimental increment may have a smaller or different effect than the FSP benefit. Positive experimental effects on household food supply and individual diets would therefore create a strong presumption that the FSP has equal or greater impacts. If the experiment finds no significant effects, it provides no information about the FSP and increases the importance of conducting the study of Expenditures and Dietary Outcomes.

If any portion of the research agenda can be considered a bargain, it is the Expenditures Experiment and the two other outcomes studies conducted in the same time frame: the Extant Data Outcomes study and the Qualitative Outcomes study (Studies 6 and 7). These three studies could be completed in three years at a combined cost of about $4.3 million. Each of the three studies would yield valuable additions to the remarkably weak existing literature on the relationships between food expenditures, household food supply, and individual dietary outcomes.

The third key decision point occurs *at the end of the Expenditures and Dietary Outcomes study* (Study 8). This study will measure, among other things, the rate at which an increase in low-income households' food expenditures translates into improvements in individuals' dietary intake. One possible finding is that an expenditure increase equivalent to that expected with the average FSP benefit does not normally lead to a measurable improvement in a low-income person's diet, or leads to an improvement that is too small to have policy significance. In this case, FNS would have to decide whether and how to proceed with subsequent studies, especially the National Experiences and Impacts study (Study 11). FNS might consider, for example, conducting the study only in locations with strong nutrition education programs, or measuring outcomes related to dietary knowledge and preferences rather than individual intake.

On the other hand, the Expenditures and Dietary Outcomes study may show that expenditure increments of the size expected with food stamps do normally lead to meaningful improvements in individual diets. In this case it will be appropriate to proceed with the agenda as conceived, using information from this study to calculate appropriate sample sizes.

The Expenditures and Dietary Outcomes study will also provide a wealth of substantive information that will be useful for policy makers and researchers. Perhaps most importantly, it will measure a wide array of behaviors that may influence the link between food expenditures and individual diets, such as eating at restaurants, eating at the home of friends or relatives, serving food to non-household members, shopping for bargains, growing or hunting food, and obtaining food from food pantries or government programs. This information will be especially important if the link between food expenditures and individual diets proves weak or non-existent, in which case policy makers will want to know what behavioral changes may be needed to achieve dietary improvements.

Omitting Some Studies from the Research Agenda

We believe that Studies 1 through 9 all have merit, and if resources are available it would be desirable to implement the full agenda. Recognizing that trade-offs may be necessary, however, we have attempted to identify ways to accomplish the basic purpose without conducting all of the studies in the agenda. As a guide to making these decisions, we used the following criteria.

First, regarding *approaches to addressing selection bias:*

1. ***If participation is to be studied, one should do it thoroughly.*** Rather than designing a participation survey based on *a priori* notions of what should be in the models, it is worthwhile to see first what can be learned about determinants of participation from extant data sources and from open-ended interviews with low-income households.

2. The methodological value of the experiment comes from the replication, so ***one should plan to do both the experiment and the replication, or neither***.

3. We believe that the experiment and replication are essential for judging the reliability of a nonexperimental evaluation approach. We do not know whether existing information on participation is sufficient for a successful replication, but it may be. Hence ***the experiment and replication*** (Studies 4 and 5) ***are higher priority than the participation studies*** (Studies 1-3).

Second, regarding *approaches to addressing relationships among outcomes:*

4. ***The extant data studies of relationships between food expenditures and diet-related outcomes*** (Study 6) ***are low-cost*** and potentially illuminating. We recommend their inclusion in any research agenda.

5. Our current knowledge about how households meet their food needs is insufficient to design a good survey of food expenditures and diet-related outcomes (Study 8). Such a survey should be based on a prior qualitative study (Study 7). On the other hand, the qualitative study is not of much use by itself. Hence ***one should do both the qualitative study and the survey, or neither***.

6. The qualitative study and survey of food expenditures and diet-related outcomes can jointly provide unique and valuable descriptive data about dynamic patterns of food use in low-income households. Nonetheless, it must be acknowledged that ***the food expenditures experiment*** (Study 9) ***can provide stronger evidence on the potential of the***

FSP to affect the quality of the household food supply, individual intake, and food security.

If these criteria are accepted, the implications are as follows.

1. For addressing *selection bias*, every research agenda should include Studies 4 and 5. Resources permitting, agendas can also include Studies 1, 2, and 3 as a group.

2. For addressing *relationships among outcomes*, every research agenda should include Studies 6 and 9. Resources permitting, agendas can also include Studies 7 and 8 as a pair.

This leads to the four alternative research agendas shown in Exhibit 4.3. All agendas ultimately include Study 10 or 11, but which of these is carried out depends on the results of the preceding studies.

Exhibit 4.3

Alternatives to the Full Research Agenda

	Full Agenda	Alternative 1	Alternative 2	Alternative 3
Selection Bias Studies				
1. Extant data participation	✓		✓	
2. Qualitative participation	✓		✓	
3. Participation survey	✓		✓	
4. Participation experiment	✓	✓	✓	✓
5. Nonexperimental replication	✓	✓	✓	✓
Outcomes Studies				
6. Extant data outcomes	✓	✓	✓	✓
7. Qualitative dietary outcomes	✓	✓		
8. Expenditures and dietary outcomes	✓	✓		
9. Food expenditures experiment	✓	✓	✓	✓
Time required	12 years	8 years	12 years	5 years
Cost	$31M	$24M	$23M	$16M
10. National FSP experiences, or 11. National impact	✓	✓	✓	✓

Full research agenda. The full agenda includes all nine of the preliminary studies. By implementing all of the preliminary studies, FNS maximizes the likelihood of establishing that impacts can be estimated nonexperimentally. If FNS chooses this agenda, it will gain the following benefits:

- From the studies addressing selection bias:
    - A fuller understanding of why eligible households enter and exit the FSP, based on new analyses of extant data, a qualitative study of triggers and barriers, and a multi-wave (non-national) study that tracks households prospectively.
    - Suggestive findings about FSP impacts based on a random-assignment experiment.
    - An answer to the question of whether our best attempt to replicate experimental results nonexperimentally can succeed in overcoming selection bias.

- From the studies addressing relationships among outcomes:
    - A fuller understanding of the relationships among food expenditures, household food supplies, individual intake, weight status, and food security based on an ethnographic study of a small number of households and a multi-wave (non-national) study that tracks households prospectively.
    - Suggestive findings about relationships among diet-related outcomes, and possibly FSP impacts, based on a random-assignment experiment.

- A national study of FSP experiences and possibly outcomes, describing participants and eligible participants.

- Either nationally-based impact estimates whose credibility can be supported based on previous studies, or a well-supported conclusion that such estimates should not be attempted with the current state of the art.

Alternative 1: Limited Selection Bias Studies with Complete Outcomes Studies. This alternative omits Studies 1, 2, and 3 on selection bias. In effect, this approach gambles that learning more about FSP participation would not make the difference between success and failure in addressing selection bias. What is lost in this alternative agenda is the fuller and richer understanding of why households participate in the FSP at some times but not others, and why some households never participate. There is no way to know in advance whether this additional information is necessary or sufficient to deal successfully with selection bias.

What is retained in this agenda is:

- The paired experiment and replication that address selection bias which, in our judgment, provide the essential information needed to support or reject the idea of measuring FSP impacts nonexperimentally, and

- The full analysis of relationships among diet-related outcomes, to provide the best possible guidance in designing the national study.

This alternative cuts about $7 million in costs from the research agenda. It shortens the timeline from 12 to 8 years, assuming that the Randomized Experiment and Nonexperimental Replication begin immediately with existing outcome measures rather than waiting for any information from the outcomes studies.

Alternative 2: Complete Selection Bias Studies with Limited Outcomes Studies. This approach takes a different gamble. It assumes that we do not need to know more about the relationships among outcomes than we can learn from the Extant Data Outcomes study and the Food Expenditures Experiment (Studies 6 and 9). If these studies find weak or non-existent links between food expenditures and dietary outcomes, the agenda will provide only limited information to explain the result.

What is retained in this agenda is:

- The full set of selection bias studies, which will provide rich information about FSP participation as well as testing the viability of a nonexperimental evaluation approach.

- A random-assignment test of how an increase in food purchasing power affects food expenditures, household food supply, and individual dietary outcomes, as well as an operational test of the feasibility of measuring all of these outcomes in a single study.

This alternative cuts about $8 million from the cost of the research agenda. It does not affect the timeline, because completing the full set of selection bias studies takes 12 years.

Alternative 3: The Minimalist Agenda. We view this agenda as the minimum necessary to get directly to the goals of this project. It combines the approaches taken in Alternatives 1 and 2, taking both gambles in order to save more time and money. This alternative cuts the cost of the preliminary studies roughly in half, from $31 million to $16 million, and would be completed in 5 rather than 12 years.

What this agenda still includes is:

- The paired experiment and replication, to support the decision of whether to attempt to estimate impacts nonexperimentally in the national study.
- The Food Expenditures Experiment and extant data analyses to support the design of the national study with regard to outcome measures.

Combining Studies

The mandate of the present project was to consider alternative agendas formed by omitting some studies from the full research program. Other approaches to saving time and money could also be considered, but they generally entail even greater risk of failure to provide convincing evidence about the impacts of the Food Stamp Program. The remainder of this chapter discusses three such approaches: combining studies, measuring outcomes less intensively, and changing the research question.

If studies were to be combined, the logical candidates would be:

- Participation Survey (Study 3) combined with Expenditures and Dietary Outcomes (Study 8).

- Participation Survey (Study 3) combined with Randomized Experiment and Nonexperimental Replication (Studies 4 and 5).

Several differences in the designs of these studies would have to be overcome in order to combine them effectively. These include differences in sample size, sample stratification, the number of waves of data collection, and the outcomes and covariates measured. The most important problem, however, is respondent burden. Accomplishing all of the objectives of the separate studies would require collecting all of the data that each requires. We believe that the recommended study designs already stretch the limits of what can be achieved without unacceptable levels of non-response. Increasing the respondent burden even further seems very unlikely to produce reliable data.

An even more aggressive approach involves both omitting and combining studies. The idea is to begin immediately with the National Experiences and Impact study (Study 11), combining it with the Randomized Experiment (Study 4). This program could in principle be completed quite quickly at far less cost than the full research agenda.

This approach entails risks that we would consider unacceptably large. It amounts to the same design that the Technical Working Group rejected, with the sole exception that the Randomized Experiment has been added. If the results of the Randomized Experiment are not solidly convincing about the reliability of the nonexperimental approach, FNS will have to consider the findings of the national study as unreliable. This means not only that substantial sums will have been spent unnecessarily, but that unreliable findings will still be widely quoted by groups that do not understand or choose to ignore their limitations.

Less Intensive Measurement

One cause of research agenda's high cost is the intensive survey procedure recommended for measuring food expenditures, household food supply, and individual dietary intake. Using less intensive measurement procedures or measuring fewer outcomes would substantially reduce study costs.

The idea of simpler measurement procedures is appealing not only for cost reasons, but because the recommended procedures use time frames that differ from each other (one week for food expenditures and household food supply *versus* 24 hours for dietary intake) and from the food stamp cycle (one month). The mismatch in time periods adds measurement error to the estimated relationships among outcomes and to the estimated impact of FSP participation. Simpler procedures using a 30-day time frame might reduce cost while improving the precision of estimates.

Unfortunately, no simpler procedures for measuring these outcomes have been validated. Until validated alternatives are available, only the recommended measurement approaches can be assumed to yield reliable results.

Measuring fewer outcomes also has some appeal. If the key research question is whether the FSP improves individuals' diets, it is arguably unnecessary to know whether the program affects the intermediate outcomes of food expenditures and household food supply. This approach becomes unsatisfactory, however, if the study finds little or no FSP impact on dietary intake (which has been

the general result of previous research). Presented with such a finding, policy makers would want to think about ways to enhance impacts on dietary intake, which would require, or at least benefit greatly from, an understanding of program effects on food expenditures and household food supply.

To give an idea of the savings that can be obtained by measuring fewer outcomes, we recalculated the cost of the survey component of Study 10, the National Study of FSP Experiences, Outcomes, and Satisfaction, under several alternative data collection assumptions. The one-wave survey of this study (excluding the RDD component) is estimated to cost around $2.6 million. It collects information on the full set of dietary outcomes.

1. If data collection of *individual intake* were dropped, about $50,000 could be saved. (This component is relatively inexpensive, as it is done by telephone.)

2. If the household food manager were queried about *food expenditures*, but no information was collected about the *quality of the household food supply* (while retaining the individual intake component), greater savings of $100,000 relative to collecting the full set of outcomes could be achieved.

3. Finally, if *all household-level measures* were dropped, other than food security, about $400,000 could be saved on the survey.

Changing the Research Question

One might summarize the preceding discussion by saying that, given the impossibility of random assignment and given current research methodologies, evaluating the impact of the Food Stamp Program will take a long time and cost a lot of money, and even after all that time and money may be subject to challenge.

This leads us to ask whether the government might get a better return on its research investment by changing the question—that is, by focusing on questions that are more tractable but still she light on whether the FSP is accomplishing what it should and whether it could be improved. Two ways of changing the question might serve this end.

One approach is to ask whether the FSP is meeting specified objectives. For example, policy makers might set as program goals that at least 90 percent of all participating households should be food secure, or that no more than 15 percent should have "poor quality" diets based on HEI scores. The research needed to determine whether the program is meeting such standards would be straightforward, reliable, and far less costly than most of the studies described here.

The second approach is to look for opportunities to apply random assignment evaluations of particular program components or particular populations. Any program component that is not fully mandated in legislation or that is permitted to vary across States or localities is a potential candidate for such an evaluation. For example, it would be legal and ethical to use randomized experiments to test the effect of nutrition education on FSP participants' dietary intake. One important feature of such experiments is that they produce lower-bound estimates of the effect of the FSP (provided that the FSP normally includes the tested component). The food stamp cashout experiments in San Diego and Alabama are precedents for this approach, and are the source of the most conclusive available evidence that the FSP causes participants to increase their food expenditures.

Similarly, experiments can be conducted with any population that is not legally entitled to benefits. The Randomized Experiment proposed here (Study 4) focuses on households that would be entitled to benefits but have not applied for them, and it is legal because it is designed not to deny anyone benefits. Another approach is to focus on "near-eligible" populations, such as households that apply for benefits but are found not to be eligible. If legislation is being contemplated that would extend FSP eligibility to some new group, randomly selected members of that group could be offered participation before the legislation takes effect. Similarly, if policy changes in a way that would exclude some currently eligible group, the eligibility cutoff date could be extended for randomly selected households in the group. Particularly if experiments can be implemented with several different groups, this approach might be the most cost-effective way to obtain reliable insights into the effects of the FSP.

Chapter Five: Study 1—Extant Data Study of FSP Participation

This study will begin with a comprehensive review of the literature on predictors or determinants of FSP participation. This information will be used, in concert with extant longitudinal datasets, to develop models of FSP participation and assess their predictive ability.

Research Questions

- What are the determinants of FSP participation among eligible households?
- How well can econometric models *using extant data* classify eligible households as participants and nonparticipants?

Objectives

The primary objectives of Study 1 are to document current knowledge about determinants of FSP participation, and to assess the ability of econometric models to predict FSP participation decisions. A secondary objective is to identify types of FSP-eligible households that are extremely unlikely to participate in the program, and other types that are extremely likely to participate.

Rationale

Before collecting new data to study the topic, and even before attempting to build new prediction models from extant data, it is appropriate to review and consolidate findings of previous researchers pertaining to the basic socioeconomic determinants of FSP participation. McKernan and Ratcliffe (2003) identified a large number of factors that had significant associations with the probability of FSP participation for working-aged adults. Their model, which used 1996–1999 SIPP data, included factors for household employment characteristics, income volatility, FSP policies, household composition, demographic characteristics, economic conditions, geographic characteristics, and year. Hisnanick and Walker (2000) used SIPP data and logistic regression to assess the likelihood of participating in the FSP in 1999 (Wave 10), given that an individual initially reported participating in the program in 1996 (Wave 1). While both of these studies provide important information on predictors of participation, neither provides details of the predictive ability of the models.

Historically, longitudinal data sets have been used to study the dynamics of FSP participation and the effects of socioeconomic characteristics and occurrences on program entry and exit (Burstein and Visher, 1989; Burstein, 1993; Gleason *et al.,* 1998; Cody *et al.*, 2005). The focus of Study 1, in contrast, is on participation status. Since the research agenda may culminate in an impact study that will compare outcomes for participants and nonparticipants, it is important to be identify predictors that distinguish between eligible households that are in and out of the program, rather than between households that do and do not enter or exit at a given time.

Obviously point-in-time participation could be analyzed using cross-sectional data. The advantage of longitudinal data is that it allows for household-level random effects. A household's past behavior may be a very good predictor of its future behavior, because it captures the effects of stable unmeasured characteristics.

Study 1 will provide a benchmark for the predictive ability of FSP participation models. It could conceivably generate a sufficiently good prediction model that no further research on participation or nonexperimental methods would be deemed necessary. We believe that the more likely outcome is that we will be left with a significant amount of unexplained variation. Before proceeding with a large-scale impact study, FNS would then want to develop better models of participation (by conducting Studies 2 and 3) or demonstrate that valid impact estimates could be derived despite the remaining gaps in our understanding of participation (Studies 4 and 5).

Sample

The longitudinal dataset recommended for Study 1 is the 2001 panel of the SIPP. In addition to information about the timing of households' FSP entries and exits, the SIPP covers changes in households' economic circumstances and the contexts of their lives, including assets and liabilities, work schedules, education and training, marital history, major expense categories, and household members' functional limitations and disabilities. The analytic files will be limited to low-income households in each wave. The criteria for defining low-income households should approximate FSP eligibility criteria.

The SIPP has been in operation continuously since 1984. Each SIPP panel comprises a nationally representative sample of the non-institutionalized U.S. population. Panel members are interviewed every four months, for a total time span ranging from 2½ to 4 years depending on the panel. Sample sizes range from about 14,000 to 36,700 households per panel.

The SIPP interview comprises three components: the control card, the core questionnaire, and topical modules. The *control card* contains information about the type of housing and the household roster with basic demographics (date of birth, race/ethnicity, gender, and education). The relationship of each household member to the reference person is shown, and additional variables identify members' spouses and parents when they are in the same household. The *core questionnaire* covers labor force participation, earnings, sources and amounts of unearned income, assets, health insurance, program participation, and education activities. This information is collected for all members aged 15 and older in every wave. *Topical modules* vary by wave. These modules collect information on events that occurred prior to the initiation of the panel and characteristics that tend to change slowly and can be summarized annually. Modules for the current (2004) panel include:

- recipiency history
- employment history
- work disability history
- education and training history
- marital history
- fertility history

- migration history
- household relationships
- medical expenses/utilization of health care—adults and children (four times)
- work related expenses/child support paid (four times)
- assets, liabilities, and eligibility (four times)
- child and adult well-being (twice)
- work schedule (twice)
- child care (twice)
- annual income and retirement accounts (three times)
- taxes (three times)
- school enrollment and financing
- child support agreements (twice)
- support for non-household members (twice)
- functional limitations and disability—adults and children (twice)
- employer provided health benefits
- informal caregiving
- retirement and pension plan coverage
- welfare reform.

Data Elements

The outcome variable will be an indicator of *FSP participation by a household* at a particular time. It is possible for some household members to be FSP participants while others are nonparticipants, e.g., if some individuals are sanctioned, or if some prepare and consume their meals separately. Typically, however, the participation decision is a household-level decision. Modeling the individual-level data as if they were independent, not accounting for the correlation of the participation decision within households, could result in substantial underestimates of the standard errors of model coefficients. Ideally the unit of analysis would be the "food stamp unit," but this level of detail is not available.

The SIPP asks respondents about FSP participation in each of the four months preceding the interview. Prior research has suggested that the quality of participation data collected retrospectively may be far below the quality of the data collected in the interview month (see, e.g., Burstein, 1993, Burstein *et al.*, 2000). It may therefore be desirable to limit the analyses to participation data for the most recent month in each wave.

The process of limiting the vast number of potential explanatory variables to a more manageable set that have a reasonable likelihood of being important predictors of the FSP participation decision will be guided by the review of the FSP participation literature. The list would certainly include variables that measure constructs such as:

- prior participation in FSP
- participation in other programs
- demographics of head of household: age, gender, race/ethnicity, education level, citizenship status, marital status
- health/disability status

- household composition
- marital status
- household income
- employment.

In addition, measures of FSP policies, region, and the local economy could be valuable additions—though we note that the main purpose of the participation model is to explain selection in a given time and place.

The SIPP also contains some survey items related to reasons for applying for food stamps and reasons for stopping food stamps that may be of use.

Analysis

The analysis will develop logistic regression models of the probability of participation by a household as a function of characteristics or conditions that may be time invariant, or that represent characteristics or conditions at that time or a prior time. The predictive ability of the model will be assessed by summary statistics such as the area under the ROC curve.[25]

The models that are estimated should account for the repeated observations on individual units (persons or households) over time. This could be done by using hierarchical generalized linear mixed models (HGLMMs) with random effects for individual units.

By re-estimating the participation model with varying numbers of waves, the study can also provide information on the incremental value of additional observations on a sample of households. This information would be useful input to the design of Study 3. To further assist in the design of Study 3, the analysis should include an assessment of the timing and frequency of substantial changes in household circumstances over time.

Timeline

This study will have two phases, lasting 9 to 12 months in total. In the design phase, the analytic files will be constructed and the draft and final analysis plans will be prepared. In the analysis phase, the analysis will be carried out and the results will be presented. The study tasks are as follows:

---

[25] A receiver operating characteristics (ROC) curve is a plot of the true positive rate (sensitivity) against the false positive rate (1 − specificity) for various cutpoints of a diagnostic test. In this situation, it would plot the proportion of households that were correctly versus incorrectly classified as FSP participants for different values of the cutpoint—e.g., if we classify everyone as participants for whom the model yielded a predicted probability of participation of 20 percent, 30 percent, 40 percent, etc. An ROC curve that lies along the 45° line has no predictive power; one that lies along the left hand and top borders of the graph has perfect predictive power.

1. Design (3–4 months)
   a. Initial meeting with FNS and followup memorandum
   b. Build SIPP data base
   c. Draft and final analysis plan

2. Analysis and reporting (6–8 months)
   a. Analysis
   b. Draft and final report
   c. Briefing
   d. Submit data and codebook

3. Management
   a. Prepare monthly progress reports

Estimated Cost

The total estimated cost of this study is between $180,000 and $240,000.

Limitations and Risks

The chief limitation of extant data analysis is that the relationships that can be studied are restricted to the variables measured in the available data. Hence, Study 1 falls short of Studies 2 and 3 in its ability to assess other determinants of participation. In particular, the SIPP (and other existing data sources) lacks information about individuals' attitudes and perceptions about the FSP as well as experiences they may have had with application or previous participation.

Variations

Although we believe that the SIPP holds the most promise for extant data analysis of FSP participation patterns, two other data sources deserve mention: the Survey of Program Dynamics (SPD) and the Panel Study of Income Dynamics. The costs of analyzing each of these data sets are likely to be similar to the costs of analyzing the SIPP.

SPD

The Survey of Program Dynamics (SPD) is a longitudinal survey initiated in 1997 that collected data on the economic, household and social characteristics of a nationally representative sample of the US population over time. It was created in response to PRWORA, in which Congress required the Census Bureau to continue collecting data on the 1992/1993 panels of the Survey of Income and Program Participation (SIPP). The primary goals of the SPD were to provide information on spells of actual and potential program participation over a ten-year period; to examine the causes of program participation and its long-term changes that result from implementing welfare reform; and to assess the effects of national welfare reforms, how these reforms interact with each other, and how they interact with employment, income and family circumstances. It was sponsored and administered by the U.S. Census Bureau.

The SPD offers similarly rich data to the SIPP. In addition to the 1992/1993 SIPP data sets, it included a 1997 bridge survey and a 1998–2002 survey created to be compatible with the prior SIPP surveys.[26] Information on the sample members therefore includes the topics covered in the core of the SIPP and the topical modules for the 1992/1993 panels. The later SPD surveys include several questions on food expenditures and a detailed food security module, as well as cognitive and attitudinal batteries lacking in the SIPP (e.g. social competence, parental depression). Of greatest interest are the questions about FSP exit and entry:

What set of circumstances led (you/NAME) to apply for food stamps in (month/year)?
[multiple responses permitted]
1. Needed money
2. Pregnancy/birth of child
3. Began receiving for another dependent
4. Separated or divorced from spouse/partner
5. Loss of job/wages/other income
6. Loss of other support income
7. Just learned about the program
8. Just got around to applying
9. Became disabled
10. Other (specify)

Why did (you/NAME) stop receiving food stamps in (month/year)?
1. Food stamp benefit cut off
2. Because of family changes
3. Still eligible but chose not to collect
4. Other (specify)

What reasons were given for (your/NAME's) food stamps being cut off? [multiple responses permitted]
1. Not eligible—income or other resources too high to qualify
2. Not eligible—due to penalty from previous program participation (sanctioned)
3. Not eligible—did not meet health or disability requirement
4. Not eligible—immigration status
5. Not eligible—no reason specified or some other reason given
6. Did not provide all the information requested
7. Non-cooperation with work requirements
8. Non-cooperation with child support requirements
9. Not residing in an adult-supervised household
10. Failed substance abuse requirements (testing or any other related)
11. Had already received maximum assistance (time and $ limit)
12. Lack of program funding
13. Other reason (specify)

---

[26] Budget cuts necessitated a reduction in sample size of about half between the Bridge Survey and the 1998 SPD.

The presence of these items and the long time span covered are attractive features of the SPD. Two disadvantages, however, are that the SPD is becoming outdated, as it is no longer collecting data; and that attrition over its 10-year course was considerable. The response rate for the 1992/1993 SIPP was 73.4 percent. The rates for the Bridge Survey and subsequent waves of SPD data collection through 2001 were 58.7 percent, 50.0 percent, 50.2 percent, 55.7 percent, and 63 percent. There has been particular concern about the higher attrition rates for low-income households. User Notes for the second (interim) SPD Longitudinal File include this caveat:

> The original SPD sample is five to six years old and has undergone a substantial amount of attrition (sample loss) from 1992 (or 1993) through 1998. As a result, estimates from this file are not as representative of the U.S. population as a cross-sectional survey or a longitudinal survey with less sample attrition would be. Therefore, results should be viewed with caution.[27]

Another challenge is that the SPD longitudinal file contains data collected using three different survey collection vehicles: the 1992/93 SIPP paper instruments that were used to produce data for calendar years (CY) 1992, 1993, and 1994, a modified March CPS CAPI instrument that was used to collect data for calendar year 1996, and the 1998 CAPI instrument that was used to collect data for calendar year 1997. Therefore, the different questions and modes of interview used to produce the estimates should be considered when analyzing changes over time.

PSID

A salient feature of FSP participation is that eligible households may continue as nonparticipants for quite some time before they are tipped into participation by some life occurrence. It seems that individuals have latent thresholds for participation, and may require larger or smaller external shocks to induce them to change their status. This suggests a secondary line of analysis that could identify subgroups of individuals that have very high or very low probabilities of participation. This analysis would use the data from the Panel Study of Income Dynamics (PSID), which includes extensive time series on some individuals. (Whether it contains extensive series on households depends on the analyst's view of what constitutes a longitudinal household, i.e., if the "same" household remains in existence for decades as individual members come and go.) Some individuals may be observed never to participate in the FSP over several decades, despite many years of apparent eligibility. Others might be observed to participate in every year in which they were eligible. Characterizing such individuals might shed light on the problem of selection bias. A tree-based methodology, i.e. splitting the sample repeatedly on different variables to categorize sample members as participants or nonparticipants, may represent a better strategy for identifying these groups than regression.

The PSID panel began in 1968 with a nationally representative sample of 4,800 families, and has interviewed them and their successors continuously ever since. Prior to 1997, families were interviewed on an annual basis. Starting in 1997, the interview schedule changed to one interview every two years. Over the years, there have been some reductions and some additions to the original core sample. As of 1999, the panel sample included 6,434 families.

---

[27] www.sipp.census.gov/spd/long/usernote_2lgt.htm.

Similar to the SIPP, the PSID has core surveys, which are administered at every data collection, and topical modules, which are administered only in selected years. The core topics include:

- income sources and amounts
- poverty status
- food security
- public assistance in the form of food or housing
- other financial matters (e.g., taxes, inter-household transfers)
- family structure and demographics (e.g., marital events; birth and adoptions; children forming households)
- labor market work (e.g., employment status, work/unemployment/vacation/sick time; occupation, industry; work experience)
- housework time
- housing (e.g., own/rent, house value/rent payment, size)
- geographic mobility (e.g., when and why moved; where head grew up; all states head has lived in)
- socioeconomic background (e.g., education, ethnicity, religion, military service; parents' education, occupation, poverty status)
- health (e.g., general health status, disability, 30-day emotional distress).

Supplemental topical modules over the years have included:

- housing and neighborhood characteristics (1968–1972, 1977–1987)
- achievement motivation (1972)
- estimating risk tolerance (1996)
- child care (1977), child support and child development (1997, 2002)
- job training and job acquisition (1978)
- retirement plans (1981–1983)
- health: health status, health expenditures, health care of the elderly and parent's health (1986, 1990, 1991, 1993-1995, 1999–2003)
- kinship: financial situation of parents, time and money help to and from parents (1980, 1988)
- wealth: assets, savings, pension plans, fringe benefits (1984, 1989, 1994, 1999–2003)
- education: grade failure, private/public school, extracurricular activities, school detention, special education, head start programs, criminal offense (1995)
- military combat experience (1994)
- risk tolerance (1996)
- immigration history (1997)
- time use (1997, 2002)
- philanthropic giving (2001–2003).

Chapter Six: Study 2—Qualitative Study of the Determinants of FSP Participation

Study 2 uses in-depth retrospective interviewing with a relatively small sample of households to generate hypotheses about the determinants of FSP participation. Because of its qualitative nature, Study 2 can go beyond the predictors typically found in general-purpose surveys (such as demographics, household composition, income sources, changes in socioeconomic circumstances) in developing an understanding of participation patterns.

Research Question

- What are the determinants of FSP participation among eligible households?

Objectives

The objective of this study is to use conversations with households to identify factors that influence households' decisions about FSP participation. The approach must be sufficiently systematic and comprehensive that it is unlikely to overlook any factor that importantly affects the FSP participation decisions of large numbers of households.

The study will explore the events, circumstances, attitudes, relationships, motivations, opinions, beliefs, and perceptions that come into play as households decide whether to apply for or to continue receiving food stamps. Aspects to be considered will include the following:

1) When a household's needs and/or resources change in ways that make the household eligible for food stamps, what adjustments does the household make or consider making? What resources does the household call on? What expenditures are cut or deferred?

2) How long does the adjustment process go on? Does it go through phases?

3) How does FSP participation fit into this picture—which adjustments are preferred to food stamps, which are less desirable than food stamps? What rationales drive the priority rankings?

4) How are food-related needs, as opposed to other financial requirements, considered in the decision-making? What priority is given to food-related needs? What adjustments are seen as particularly responsive to food-related needs?

5) How long does the household think it will be before its financial situation improves? How has that perception evolved? How does the expectation affect adjustment choices, including FSP participation?

Rationale

Dealing effectively with selection bias requires that two conditions be met. First, the determinants of participation must be sufficiently well known that one can be reasonably certain that no important factors have been omitted from the study. Second, the analytic model must include all determinants of participation that are also correlated with the outcomes of interest (such as changes in food expenditures, diet quality, and food security).

A number of studies, most recently Bartlett and Burstein (2004) and Cunnyngham (2005), have compared participation rates for various population subgroups and have found that eligible households are less likely to participate if the household head is white or Hispanic, is elderly, or has more education, or if the household has higher income, earnings, or assets, or does not participate in TANF. McConnell and Nixon (1996) summarized findings from earlier studies and found them largely consistent. Other research has shown participating households are likely to have experienced a recent precipitating event, such as a job loss (Burstein, 1993; Gleason *et al*,. 1998). Some studies have asked apparently eligible households why they did not participate. These studies have typically found that households were not aware that they might be eligible, felt they could get along without food stamps, were averse to receiving government help, were deterred by stigma, or were confused or put off by the application process or program requirements (Coe 1983; U.S. GAO, 1988; Blaylock and Smallwood, 1984; Ponza *et al.*, 1999; Bartlett and Burstein, 2004).

Existing literature cannot be assumed to have identified all of the important determinants of participation. It is quite possible that food stamp participation depends importantly on factors that have not yet been considered or examined systematically. Participation may be conditioned by people's attitudes toward government programs or food and diet, or by their feelings of food security. Only a few research efforts have focused on these issues. One can hypothesize a large number of potential influencing factors that have not been explored, such as the extent to which a household can call on family and friends for help, its ongoing and short-term expenditure requirements, and its chance encounters with information or with formal or informal referral agents.

Sample

A purposive sample of households entering the FSP will be recruited for the study. Households will be recruited at or around the time of application. For example, with permission from a local FSP office, an interviewer might be stationed in the reception area. The interviewer would identify applicants whose households met the study criteria, explain the study and the incentives, and schedule a subsequent time for an interview.

A total of 60 households will be selected, with 20 in each of the following three groups: households consisting of elderly persons, single-adult households with children, and multiple-adult working poor households. (While there are no hard and fast rules about sample sizes for this type of research, a rule of thumb is that 20 subjects are needed for each group that is separately of interest.) Samples will be drawn in three different locations for geographic diversity, including two urban areas and one small town or rural location. Interviews within each major group will be roughly evenly divided among the three locations. Within each group in each location, interviewers will seek a racial/ethnic balance roughly similar to that of the FSP caseload.

Data Collection

Persons selected for study will be interviewed twice. The first interview will occur as soon as possible after recruitment into the study. The second will occur at the end of the household's certification period, or six months after the initial interview if the certification period is longer than six months.

Highly trained interviewers will conduct exploratory interviews, loosely structured around the topics noted under "Objectives." Typically, the interviewer will first attempt to determine the point at which the household reached its current state of eligibility for food stamps (income, household composition, and assets). This point might be some weeks or months, or possibly even years, prior to the interview. The interviewer will take this as the starting point for learning the household's story, briefly exploring the events that occurred to bring about that status, but principally focusing on the household's actions and perceptions in dealing with the situation.

The second interview will serve two purposes. Most importantly, the interviewer will explore the household's situation and its decision to continue or discontinue participation at the recertification. In addition, the interview will provide an opportunity to revisit issues raised in the first interview, to see whether the respondents have any new perspectives, and clarify or get more detail on points that the interviewer found interesting. To the extent possible the same interviewer will conduct the initial and followup interviews, for purposes of continuity in relationship and developing a deeper understanding of the household's experience and actions.

Interviews will have varying duration, but first interviews are generally expected to last two to three hours. Second interviews may be a bit shorter. Respondents will be offered an incentive payment. Based on practices in 2005, an incentive of $40 per interview would be appropriate. We estimate that about 80 percent of the first-round sample will complete the second interview.

Analysis

The analysis will have two parts. The first part is household-specific, developing a compendium of stories of the 60 households in the study. Each household's story will be told in a way that identifies all of the factors that the subject presented or the interviewer interpreted as having a bearing on the household's FSP participation. To the extent possible the individuals who conduct the interviews will write up these stories because they are closest to the data and can provide important context and interpretation.

The second line of analysis is synthetic, looking across households to provide an organized statement of the factors apparently influencing FSP participation. The intent of this report is to be comprehensive, including all factors in addition to those deemed to be most common or most important. At the same time, the report will have a strong interpretative component, as the study team offers hypotheses about what factors are most important, how factors interact, and what key elements have been missing in previous quantitative research.

Timeline

The overall study is expected to take 20 to 26 months, in three major phases. The design phase will require six to eight months, the field phase will take eight to 10 months, and six to eight months will be needed for the analysis and reporting phase. Key tasks within the phases are listed below.

1. Design phase (6–8 months)
   a. Initial meeting with FNS and followup memorandum
   b. Develop research protocols
   c. Develop procedures and obtain permissions for recruiting sample
   d. OMB clearance and design revisions

2. Field phase (8–10 months)
   a. Train interviewers
   b. Recruit sample
   c. Schedule and conduct initial interviews
   d. Transcribe interviews, write up interview notes
   e. Schedule and conduct second interviews
   f. Transcribe interviews, write up interview notes

3. Analysis and reporting phase (6–8 months)
   a. Household stories report
   b. Synthesis report
   c. Final briefing

Estimated Cost

This study is estimated to cost between $625,000 and $700,000.

Limitations and Risks

The important limitations of this study are its retrospective design and its limitation to food stamp participants. The ideal design might be to select families who had just become circumstantially eligible for food stamps and follow them through the process of deciding whether or not to participate. This would avoid the recall problems and selective memories that are inherent in retrospective interviews. It would provide as much information on why some people decide not to apply for benefits as why other people do apply.

The retrospective approach to sampling is based primarily on cost considerations. Although we might be able to identify a reasonably representative group of families just as they became eligible, it would be extremely costly to observe them over the weeks or months that could pass before their participation decision was clear. It would be necessary to interview the family at least every few days in order to avoid reverting to a fully retrospective approach.

The exclusion of nonparticipants from the study design also is based principally on cost considerations. Including nonparticipants would double the study sample, and while this would not

fully double the overall study cost, it would cause a substantial increase. It seems likely that most factors influencing participation decisions can be learned by focusing on participants only, especially because many participants will have been eligible nonparticipants for some weeks or months before deciding to apply for benefits.

The risk associated with both of the design limitations is that the study will fail to identify one or more key participation determinants that would have been identified if the sample were larger, was more representative, or included nonparticipants.

Variations

The most useful way to strengthen the design would be to add eligible nonparticipants to the study. The decision to exclude nonparticipants from the basic design stemmed from the hypothesis that households can be divided into three types: those that always participate in the FSP any time they are eligible; those that never participate, regardless of circumstances; and those that sometimes participate. The latter group may participate during some but not all spells of eligibility, or they may participate during part but not all of a given spell. If many or most food stamp participants are in the "sometimes" category, then the interviews with current participants should reveal information about parts of their eligibility spells when they did not participate, and the considerations that led them to switch from the state of eligible nonparticipation to eligible participation. Similarly, the second interview is expected to encounter some households who have stopped participating but remain eligible, and the interview will identify factors underlying that switch. Around half of eligible nonparticipants are former FSP participants.

Given this logic, the weakness of a study based solely on current participants is that we do not learn about the considerations influencing households in the "never" (or "hardly ever") category. If such households amount to a substantial proportion of the eligible nonparticipants at a given point in time, and if they are influenced by considerations not found in participant households, then the study would fail to capture information that could be important in dealing with selection bias.

If nonparticipants were added to the study design, the basic structure of the study would be the same for these subjects as for FSP participants, with two exceptions. First, the nonparticipants would have to be recruited by other means than stationing interviewers at food stamp offices. One possible approach would be for interviewers to ask FSP participants if they knew of people whose circumstances were similar to their own but who were not participating in the program. Another approach would be to work with community organizations such as food pantries, churches, or senior citizens centers to identify households that meet the study criteria. Recruiting nonparticipants will be more cumbersome and costly than recruiting participants, because FSP eligibility cannot be fully determined in advance of the interview, which means that some interviewed households will have to be excluded from the analysis.

The second design difference is that each nonparticipant household would be interviewed only once. The interview would establish that the household is indeed apparently eligible for food stamp benefits, and would seek to learn the history of the household's coping strategies and perceptions since the point at which it became eligible. A second interview is not suggested because of the

---

Abt Associates Inc.   **Chapter Six**   **65**

possibility that the first interview itself, by raising questions related to FSP participation, might influence people's behavior or perspective.

The nonparticipant sample would comprise 60 households, with 20 in each of the three household types used in sampling participants.

Chapter Seven: Study 3—Survey to Estimate FSP Participation Model

Study 3 uses a panel survey with multiple waves of data collection, conducted in three locations, to model the probability of FSP participation among apparently eligible households.

Research Questions

- What are the determinants of FSP participation?

- How well can FSP participation be predicted using a specially designed survey?

Objectives

The key requirement for avoiding selection bias in the potential impact study (Study 11) is to include those factors which are not only significant predictors of participation but also are significantly correlated with key outcomes (food expenditures, diet quality, etc.). Based on Studies 1 and 2, Study 3 will examine a large number of factors as significant predictors of participation. The objective of Study 3 is to winnow down the list of potential predictors to those that are actually important, and to show that these explain participation well—that is, to determine which factors are necessary and sufficient for understanding FSP participation. Study 3 will not, however, measure outcomes and hence will not be able to test the attendant correlations.[28]

Rationale

This study uses the information on predictors of participation obtained in the other studies of FSP participation—Study 1, which uses extant data and Study 2, which uses qualitative research. The extant data study will test the limits of existing survey data in predicting FSP participation. The Study 3 survey will need to be conducted only if Study 1 shows, as expected, that existing data are not sufficient to predict participation very well. If, in Study 1, models based on extant data do not predict participation well enough to make it clear that selection bias can be avoided, the qualitative research in Study 2 will be examined to identify additional survey measures that may add strength to participation models. (It is theoretically possible that the qualitative research would identify no important factors beyond those included in the extant data analysis. If this occurs, which seems unlikely, it would be another reason that Study 3 would not be conducted.)

---

[28] Several strategies could be used in determining which of the participation predictors are correlated with outcomes. Existing literature on the outcomes will already have examined some of the predictors. If the timing of the studies permits, one of the later surveys in the studies of food expenditures and diet quality (e.g., Study 8) could include the predictors identified in this study, with the intent of examining the correlations.

The impact study will measure all significant participation predictors that have not been shown to be uncorrelated with the outcomes of interest. In practice, it is hard to imagine excluding any important known predictor of participation from an impact study.

The determinants of FSP participation include both long-term and short-term household attributes and circumstances. Some eligible households may never participate, even in protracted periods of eligibility, while others may apply for food stamp benefits whenever they become eligible. The "never participate" households may be distinguished by characteristics that never or seldom change, such as whether they are elderly, are strongly averse to accepting government assistance, or have a strong local family network. Other households may seek benefits during some portions of a spell of FSP eligibility, or may participate during some eligibility spells but not others. The behavior of these households is presumably influenced by events and conditions that can change relatively rapidly. Such factors would include (a) events that alter household composition and income (e.g., birth of a child, losing a job), (b) events that change expenditure requirements (illness, an automobile breakdown), (c) events that focus the household's attention on food needs (food security), (d)attitudes regarding the seriousness of the household's condition ("we can get by"), (e) expectations about the future ("things will be better next month" *versus* "there's no jobs out there"), and (f) receipt of information or assistance in accessing food stamp benefits or alternative resources.

This study will seek to identify and examine a comprehensive set of predictors of participation, covering all of the factors mentioned above and any others suggested by the literature search conducted under Study 1 and/or findings of the qualitative study (Study 2). The study will include both current participant and eligible nonparticipant populations, and sample members will be interviewed multiple times. The sample will therefore include "never" and "always" participants, as well as "sometimes" participants in both their participating and nonparticipating states.

Sample

The sample will be drawn in three locations, which is assumed to provide sufficient diversity to ensure that the estimated relationships are not determined by unique location-specific factors. Because the study will focus on household-level attributes, it is not necessary to include the large number of locations that would be required to test hypotheses about community-level factors. As in Study 2, we expect that two urban locations and one small town or rural location would be used. Roughly equal amounts of the total sample will be allocated to each location. Although an argument could be made for including more than three locations, because in-person interviews are required for this study, additional locations would significantly increase costs,

A Random Digit Dialing (RDD) strategy will be used to identify and recruit households into the sample. Respondents will be screened to identify separate groups of current FSP participants and eligible nonparticipants in each location.

FSP participants will be further stratified into two groups: those who began participating less than six months before the interview, and those who have participated for six months or longer. This will avoid the possibility that the sample will consist mainly of long-term participants, many of who would not be observed in a nonparticipant condition during the course of the study.

Eligible nonparticipants will be stratified into those with household income above and below 80 percent of the poverty line. This will avoid the possibility that the sample is dominated by households that would be eligible for only small FSP allotments.

One person will be interviewed in each selected household. The screening protocol will be designed to identify the adult in the household who is expected to be most knowledgeable about decisions to participate or not participate in the FSP.

The survey will include multiple waves. The number of survey waves will be decided partly on the basis of information developed in the extant data study of participation (Study 1). That study will provide information about how multiple waves of data affect the performance of participation models. The ultimate decision about the number of waves to include in this study will balance the cost of each additional wave against its value in strengthening the model. We anticipate that a minimum of three waves will be necessary in order to observe a substantial number of households in both the participant and the eligible nonparticipant states. For cost estimation purposes, we assume five survey waves.

A 4-month interval between surveys is planned. This is expected to be a long enough interval to allow meaningful changes in household circumstances or perspectives (e.g., a job change, or a re-assessment of how long the current situation is likely to last), but short enough that few households will experience more than one such meaningful change between surveys. Assuming the appropriate analysis is implemented, Study 1 will provide information about whether four months appears to be the appropriate interval.

Five survey waves conducted at 4-month intervals will provide a 16-month observation period (or 20 months, including the months covered retrospectively in the first interview). We would expect about a third of sample members who are FSP participants at baseline to become eligible nonparticipants during this period. Further, we would expect about 8 or 9 percent of those who are eligible nonparticipants at baseline to become participants within 16 months.[29]

The suggested sample size for each location is 496 respondents in the final wave of interviews for each of the two initial groups (participants and nonparticipants). This translates into final sample sizes of 992. Allowing for 15 percent attrition between interview waves, the initial sample must include 950 participant households and 950 eligible nonparticipant households in the three sites, for a total of 1,900 households. The sample calculation is based on the amount of predictive power gained by adding a single predictor variable to a logit model of participation.[30] The most stringent requirement of this study is to determine whether a particular battery of items is, in fact, necessary for inclusion, i.e., whether it adds meaningfully to the predictive power of the model. If the sample size is sufficient to answer this question, the study will be able to estimate overall predictive power with accuracy.

---

[29] These estimates are based on rough tabulations of the first four waves of the 2001 SIPP panel.

[30] The sample calculation was done following the method of Hanley and McNeil (1983) and assumes 80 percent statistical power to be able to detect an increase of 2 percentage points in the area under the ROC curve (significance level of 0.05 in a one-tailed test). We assume that the area under the curve is 0.70 before adding the variable, and that the correlation between predicted participation values before and after adding the variable is 0.90.

Data Collection

An extensive set of potential predictors of participation will be measured, including long-term and more transient characteristics of the household. These characteristics will be selected on the basis of findings from Studies 1 (literature review and extant data analysis) and 2 (qualitative research).

It is highly likely that the qualitative research will identify possible determinants of participation that have not previously been measured in surveys. New survey items will have to be developed and validated. At the same time, new items could be developed for determining participation in the FSP, and for the identification of eligible nonparticipant households based on income and household size. Survey items on program participation and income are known to lead to misclassifications.

Pretesting methods for new items include expert panel review, conventional pretesting (with or without respondent debriefings), behavior coding, cognitive interviewing, usability testing, and other laboratory procedures. The cognitive interview approach is an especially effective tool for survey instrument testing because it can provide possible reasons why a question is not performing as intended. This information is valuable in the revision process and in the conduct of a conventional pretest. The technique uses respondents' verbal reports about each phase of the response process: comprehension, recall, response formation, and reporting. Cognitive interview protocols are designed to combine concurrent or retrospective "think aloud" techniques with other procedures such as scripted or free probing. The testing occurs at a special facility (a Cognitive Testing Laboratory) that allows for monitoring of interviewers and respondents and for other testing procedures, such as the use of vignettes and different types of debriefings.

We anticipate that the respondent interview will take about an hour and a quarter. For an interview this long, in-person interviewing is necessary to avoid a high rate of break-offs in the initial interview and non-response in subsequent waves. There is some possibility that the later interview waves could be shorter and therefore could be conducted by telephone. Deleting items from the survey would be based on preliminary analysis of the first two or three waves of data to identify survey questions that seem unlikely to contribute to the final participation model. Although we recommend that such an analysis be conducted, we cannot assume that the interview can be shortened sufficiently to permit telephone interviewing. Cost estimates are therefore based on in-person interviewing for all waves.

In order to minimize non-response and wave-to-wave attrition, techniques to encourage cooperation will be necessary. A small incentive payment of $20 per interview has been assumed for costing this study. Incentives have been used for many years in market research surveys and are becoming more common in government-sponsored surveys. The survey methodology literature shows that up-front monetary incentives are more effective in raising response rates than promised monetary incentives (Singer *et al.*, 2000). Also, non-monetary incentives (e.g., a phone card) are generally less effective than monetary incentives.

Analysis

The main analysis will model FSP participation in the month of the interview as a function of household-level predictors. The modeling approach will need to account for the hierarchical nature of the data: multiple interviews are conducted with each household, and respondents are clustered within

communities. A logistic regression with random effects for households and fixed effects for communities is one logical approach. Tree-based modeling techniques such as CART or CHAID may be appropriate if the data suggest that the predictive factors differ across population subgroups.

A key question is whether the analysis indicates that a study of FSP impacts, by including variables used in this study, has a reasonable likelihood of avoiding distortion from selection bias. Part of the answer to this question lies in how well the model actually predicts participation—i.e., its goodness of fit as measured by a statistic like the area under the ROC curve.   In general, greater predictive power means less concern about selection bias, but there is no specific threshold of acceptability. The other key issue, an entirely judgmental one, is whether a reasonable observer can think of factors excluded from the model that seem likely to be important sources of selection bias. This judgment can be made only when researchers have had an opportunity to think about the study results.

If the analysis suggests a sufficient likelihood of avoiding selection problems, an important question is what predictors of participation are needed in an impact study. Alternatively, if uncertainty remains, we may ask if an impact study can nonetheless proceed. The methods described by Harding (2003) (see Chapter Two) can be used to assess the importance of omitted factors.

Timeline

The duration of the study will depend, in part, on the number of waves of survey data collection and the interval between waves. For present purposes, we assume a total of five waves of in-person data collection. Given this assumption, the overall study would be expected to take 48 to 52 months, in three major phases. The design phase will require 8 to 10 months, and the field phase will take 30 to 32 months. There will be two rounds of analysis and reporting, one based on the first two waves of data collection, and one using all waves of data. Key tasks within the phases are listed below.

1. Design phase (8–10 months)
    a. Initial meeting with FNS and follow-up memorandum
    b. Development and cognitive testing of new survey items
    c. Design, field test, and revise survey instrument and procedures
    d. Prepare OMB submission and make required revisions

2. Field phase (30–32 months)
    a. CATI/CAPI programming
    b. Train interviewers
    c. RDD survey to recruit sample
    d. Wave 1 (baseline) interviews
    e. Wave 2 interviews
    f. Wave 3 interviews
    g. Wave 4 interviews
    h. Wave 5 interviews

3. Analysis and reporting phase (24–26 months; overlaps with field phase, starting in month 24–26)
    a. Preliminary analysis (after Wave 2)
    b. Preliminary report

---

c. Final analysis
d. Final report
e. Final briefing

Estimated Cost

The estimated cost of this study is between $6,000,000 and $6,500,000. More than 80 percent of this is attributable to the survey. Most of the non-survey costs occur in the analysis and reporting phase.

The survey costs are very large because of the assumption of five rounds of in-person interviews. In addition, the RDD screening to find sample members costs between $600,000 and $700,000.

Limitations and Risks

The most important limitation of this study is its reliance on self-reports by respondents of their FSP participation. National surveys and comparisons of survey responses with administrative records have shown that individuals frequently misreport their participation status at a point in time. Measurement error in the dependent variable would generally be expected to reduce the predictive power of the model and could distort some of the specific relationships estimated.

Another limitation is that the study's ability to model long-term participation will be constrained by the fact that the survey will include only items that are expected to predict participation in the current month, many of which would not be hypothesized to predict longer-term participation.

The major risk is that the study would fail to identify some important source of selection bias—that is, some factor that is correlated with participation and also with one or more of the outcomes that the FSP is hypothesized to affect. This could occur if the factor is overlooked in choosing constructs for inclusion in the study, if the factor is poorly measured by the survey items, or if the general relationship between the factor and participation does not happen to exist in the locations chosen for the study.

Variations

An obvious way to reduce the cost of the study is to reduce the number of waves of data collection. Analysis of the SIPP data in Study 1 might support the idea of a shorter followup period.

The study design could be strengthened by verifying the FSP participation status of households in the study sample through examination of administrative records. Agreement would have to be obtained from state or local food stamp offices to provide the data. Households would then be requested— probably in one of the later interview waves—to provide their Social Security numbers and permission to access their records. An extra incentive might be needed to promote cooperation. For those households willing to have their records checked, the analysis would compare the self-reports with administrative records, and also compare participation models estimated with the alternative measures.

Chapter Eight: Study 4—Randomized Experiment to Measure FSP Impacts

Study 4 is a randomized experiment designed to measure FSP impacts in special circumstances. In the research agenda this study is coupled with Study 5, which calls for a nonexperimental study addressing the same question, in the same sites and during the same time frame as the experimental study.

Research Question

- If eligible nonparticipants are induced to participate in the FSP, what is the impact of FSP participation on their household food security, household food expenditures, and individual dietary intake?

Objectives

The main objective of this study is to provide a benchmark for comparison with nonexperimental results (Study 5). The secondary objective is to obtain information about FSP impacts, net of selection bias.

Rationale

Because FSP is an entitlement program, it is not legally or ethically permissible to deny benefits to eligible households. Therefore, an experiment must be based on some intervention that raises participation rates among eligible households in a randomly assigned treatment group while leaving participation at its usual level in a control group. For the results of this experiment to be substantively useful to FNS, the intervention should not radically change the nature of the program. For example, an intervention that doubled FSP benefits, offered benefits in cash rather than as EBT cards, or eliminated work requirements would not provide good information about the effects of the FSP as it actually operates. Likewise, an intervention focused on a small segment of FSP eligibles (e.g., undocumented aliens) would not be as useful as one that included a more broadly defined population.

In such experiments, estimated program impacts are based entirely on effects among those treatment group members who do participate, but would not have participated if they had been in the control group. For this reason, it is essential that the intervention have a marked impact on participation rates. We note that the Elderly Nutrition Demonstrations found impacts in the 20 to 35 percentage point range among the better-performing sites implementing simplified eligibility, commodity alternative benefits, and assisted application (Cody, 2004). Increases of this magnitude were possible because participation rates among the elderly are historically low. It would be hard to achieve comparable results in demographic groups whose normal rates of participation are substantially higher.

We recommend that the experiment be conducted in sites with historically low participation rates, that they include the full range of FSP eligibles, and that the intervention provide the strongest possible

incentives for participation consistent with national FSP regulations. If the selected sites have significant barriers to participation and stringent regulations, then the intervention can succeed in raising participation rates by offering treatment group members both assistance in overcoming the barriers and waivers for State and local regulations. For example, in contrast to control group households, treatment group households could be offered home visits by caseworkers who encourage them to apply for food stamps, in-home certification interviews, assistance in obtaining verification documents, and longer certification periods. (We note, however, that variations in State and local regulations are often related to error control. Extending certification periods poses the risk that treatment group members could continue to receive FSP benefits when no longer eligible.)

The intervention features are obviously not part of the FSP. Their sole role in this study is to create a wedge in participation rates between treatment and control group households, without altering program impacts. Positive incentives to participate might also be considered, as long as they did not pose a danger of affecting eligibility status or the outcomes being measured. These incentives could be neutral in-kind gifts, e.g., magazine subscriptions or housewares.

Sample

The sample will be drawn in three sites, using RDD screening of households to identify eligible nonparticipants. For face validity, these sites should include different regions of the country, varying racial/ethnic group concentrations, and both urban and rural locations. The advantage of more sites is greater representativeness with respect to environmental factors and household characteristics that could affect dietary outcomes, with and without receipt of food stamps. The disadvantage is increased cost for the in-person interviews. Random samples will be drawn in each site, so the sample will be representative of the full population of eligible nonparticipant households in those geographic areas. Treatment group members will be exposed to the intervention immediately after baseline data are collected. The intervention, which will be carried out by a caseworker, will comprise an explanation of the FSP and encouragement to apply, including an offer of an in-home certification interview and assistance with obtaining necessary documentation.

In calculating the sample size, it is important to take into account the fact that many treatment group members will not choose to enroll in FSP, while some control group members will enroll. This has a serious impact on required sample sizes. Under the reasonable assumption that treatment group itself does not affect dietary intake except through changing the likelihood that households enroll in the FSP, the impact of the intervention on the dietary outcomes of the entire treatment group is equal to the impact of FSP participation *multiplied by the incremental likelihood of participating*. Hence if we want to be able to detect a 5 percentage point difference between FSP participants and eligible nonparticipants in the proportion with "poor quality" diets based on HEI scores, we will need to be able to detect a smaller difference in this measure between the treatment and control groups in their entirety.

For purposes of costing this study, we assume that over a six-month period about 5 percent of eligible nonparticipants in the control group will enroll, and that the intensive intervention will increase this proportion by *50 percentage points* in the treatment group. It will be seen that even this aggressive assumption the required sample size, and consequent study costs, are quite large. A more conservative estimate of the effect of the intervention would imply still larger samples and costs.

We assume that our covariates, which will include baseline outcomes, will explain 25 percent of the variance. This implies 3,088 final wave completed interviews in each group for each site or group of sites for which impacts are estimated. Thus the total final wave sample size per site is 6,176 interviews. Allowing for 15 percent attrition the initial number of completed interviews in a site is 7,264 (3,632 per group).

Data Collection

The design assumes, for both treatment and control groups, a baseline interview and a six-month followup interview. In addition to collecting information on basic demographics, the baseline contact will serve four additional functions. First, the information collected will screen out some households that appeared to be for eligible for food stamps based on the brief RDD screener but, in actuality, are not eligible. These households will be dropped from the study.[31] Second, collection of outcomes data at baseline as well as at followup will support estimation of random effects models of program impacts. Third, for those assigned to the treatment group, this contact will offer the intervention, including in-home certification. Finally, for those in the control group, it will collect information which will be needed in Study 5, as described in the following chapter.

Both baseline and followup interviews will collect information on household food expenditures, quality of the household food supply, food security, and individual dietary intake. In each sampled household, one individual will be randomly sampled for collection of individual intake data. As discussed in Chapter Two, the recommended measures for these outcomes are a 7-day record of household food use (supplemented with questions about away-from-home food expenditures), the 10-question food security module, a 24-hour recall, and a second recall for 25 percent of the population. The burden associated with this data collection will be substantial. Incentive payments to participating households, for both baseline and followup data collections, will be essential. For costing purposes, we have assumed a small payment of $20 per wave.

Analysis

The analysis will be performed for all sites combined. The sample size requirements for a single pooled estimate are too large to consider developing separate estimates by site.

Impact estimates will be based on comparison of outcomes between treatment and control group households (T's and C's), corrected for "crossovers" and "dropouts." We put these terms in quotes because we are referring to FSP participation among the C's and FSP nonparticipation among the T's, rather than the usual reference to receiving or not receiving intervention services.

The randomly assigned groups will include three types of households: those that would normally participate in the FSP, those that only participate in the FSP because of the intervention, and those

---

[31] Even after this second pass, the focused eligibility interviews conducted by caseworkers with treatment group members as part of the intervention will undoubtedly result in some denials. These more subtly ineligible households in the treatment group have counterparts in the control group, so they must be retained in the analysis sample.

---

that do not participate, regardless. *The impact estimate will be based exclusively on those households that participated only as a result of the intervention*—not on those that would have participated normally. The mean outcomes for the entire treatment and control groups are appropriately weighted combinations of the outcomes for these types of households:

$$y_T = p_1 \, y_1^P + p_2 \, y_2^P + p_3 \, y_3^N$$
$$y_C = p_1 \, y_1^P + p_2 \, y_2^N + p_3 \, y_3^N$$

where y_T is the mean observed outcome for the treatment group,

y_C is the mean observed outcome for the control group,

p_1, p_2, and p_3 are respectively the proportions of eligible households that would participate in any event, would participate only under the intervention, and would not participate in any event,

y_j^P is the mean outcome for household type *j* conditional on FSP participation, and

y_j^N is the mean outcome for household type *j* conditional on FSP nonparticipation.

Taking the difference between the two equations, we find

$$y_T - y_C = p_2 \, y_2^P - p_2 \, y_2^N \,,$$

i.e., the impact *for household type 2* is

$$y_2^P - y_2^N = (y_T - y_C) / p_2.$$

This is all we can learn from the experiment—the impact for households that would participate only under the intervention.

An additional complication is that participation is not dichotomous, but rather continuous. T's and C's may have participated in the FSP for any number of months between zero and six inclusive. We suggest that several alternative measures of participation be constructed, such as:

    a) participation in the most recent month;

    b) any participation in the past six months;

    c) fraction of months participating out of the past six;

    d) weighted fraction of months participating, with higher weight given to more recent months (e.g., $0.5 \, P_t + 0.25 \, P_{t-1} + 0.125 \, P_{t-2} + 0.0625 \, P_{t-3} + \ldots$, where P_{t-k} is an indicator that the household participated *k* months prior to the interview).

For outcomes for which FSP is expected to have immediate impacts (e.g., household food expenditures), definition (a) would be appropriate. For outcomes for which persistent impacts were anticipated (perhaps household food use), definition (b) could be preferable, while analysis of longer-run effects might use definitions (c) or (d) depending on the hypothesized shape of the response function.

Faute de mieux, we would suggest that the standard correction for crossovers and dropouts be used based on continuous participation measures. That is, the participation rate for the treatment group would be calculated as the mean of the continuous participation measure over treatment group members, and similarly for the control group.

Timeline

This study runs for 44 to 48 months. The design phase, falling entirely in the first year, includes site recruitment and selection, development of instruments and procedures, and the OMB submission. The field phase, beginning at the end of Year 1 and running through Year 2, includes an RDD survey to recruit the sample, the baseline survey and random assignment, and the follow-up survey. The analysis and reporting phase begins when the baseline survey is complete, and runs through Year 4. The tasks for the study are:

1. Design phase (8–10 months)
   a. Initial meeting with FNS and followup memorandum
   b. Site recruitment and selection
   c. Design, field test, and revise survey instrument and procedures
   d. Prepare OMB submission and make required revisions

2. Field phase (20–22 months)
   a. CATI/CAPI programming
   b. Train interviewers
   c. RDD screening to recruit sample
   d. Baseline survey
   e. Random assignment
   f. Followup survey

3. Analysis and reporting phase (18–20 months; overlaps with field phase, starting in month 26–28)
   a. Preliminary analysis (baseline data)
   b. Preliminary report
   c. Final analysis
   d. Final report
   e. Final briefing

Estimated Cost

Exclusive of the survey, the costs of this study are between $1,200,000 and $1,500,000, the great bulk of which is in the analysis and reporting phase. The survey costs are very large, because the sample size is so large, and because the study includes several length in-person and telephone contacts in each wave. The RDD screening to identify sample members is estimated to cost between $1,400,000 and $1,600,000; the baseline survey, between $3,800,000 and $4,600,000; and the follow-up survey, between $2,800,000 and $3,300,000. Thus, the entire study will cost between $9,000,000 and $11,000,000.

Limitations and Risks

The chief limitation of this study is that it provides a single set of impact estimates. While the combination of three sites will ensure that the estimate is not driven by idiosyncrasies of one or another locale, it will not be possible to obtain reliable estimates for each site separately.

A second limitation is that impact estimates will be based entirely on FSP effects for households that are induced to participate by the intervention. These may not be typical of FSP participants. Arguably, program impacts are greater for households that are already participating than for households that would participate only if offered extraordinary encouragement to do so. Study refusals, sample frame exclusions (e.g., households lacking land-line telephones), and sample attrition may further reduce generalizability. In addition, the study will be conducted in only a few purposively selected sites, rather than being based on a nationally representative sample.

The chief risk for this study is that the intervention would be insufficient to increase the participation rate of T's sufficiently above that of C's to allow precision in estimating impacts. While information is available on the efficacy of assisted participation for the elderly, there is no such information for other population groups.

Variations

In developing the design for this study, many alternatives were considered. Although they were deemed less practical or less strong than the design presented in this chapter, they are worth describing.

One variation that could be very informative, if resources permitted its implementation, is three-way random assignment, with the third group receiving intensive nutrition education in addition to the usual FSP benefits. This experiment would shed light on the impacts that could be expected if the nutrition education component of the FSP was substantially strengthened. This would however increase the survey costs by nearly fifty percent.

Another variation would include ongoing FSP participants in the study, offering them longer certification periods, in-home interviews, and assistance with verification when they came due for certification, with the intent of creating a wedge in the recertification rate. The drawback of this variation is that relatively few eligible participants fail to recertify.

A third variation would draw the sample from administrative lists of participants in other means-tested programs (e.g., TANF, WIC, Medicaid, free NSLP). Barriers and drawbacks to this approach include the following:

- The samples would have to be screened further for FSP eligibility because income cut-offs are higher than 130 percent of the federal poverty line for some of these programs in some or all States;

- These groups are nonrepresentative of FSP participants in general;

- Administrative staff would likely be reluctant to provide the necessary contact information, especially for school meals participants.

A fourth variation that was considered was to draw convenience samples of FSP eligibles from among WIC exiters or attendees at senior centers. This had the similar drawback of nonrepresentativeness.

A more attractive idea is to randomly assign explicitly denied applicants. One could set up a procedure in which certain households that are about to be denied are offered the opportunity to participate in the experiment, with half getting food stamp benefits and the other half getting some other incentive (e.g. housewares). Because the treatment group is already in the office applying for benefits, most of them could be expected to participate. The control group could be denied benefits legally and ethically, as long as they stay ineligible. Thus one can achieve close to 100 percent participation in the treatment group (because they are interested in receiving food stamps) and 100 percent nonparticipation in the control group (because they have been found not to be eligible). This approach eliminates the need for RDD or other expensive procedures to track people down, substantially reducing the cost. On the other hand, because the sample members are all ineligible, the estimated impacts are not of great intrinsic interest. While this approach has considerable appeal, we ultimately concluded that it was dominated by randomizing from the population of eligible nonparticipants.

In addition to variations in overall study design, several variations in the range of outcome measures were considered. To conserve resources, collection of information on food use and/or individual dietary intake could be eliminated. We assume, however, that even limited information on impacts in these domains is of high interest to FNS. An even more expensive study could include two 24-hour recalls on all sampled individuals, as is currently the practice in NHANES.

Finally, a variety of policy variations were considered. We ultimately concluded that all acceptable policy variations should be offered simultaneously to the T's.

Chapter Nine: Study 5—Nonexperimental Replication of the Randomized Experiment

Study 5 will implement and analyze a nonexperimental replication of the randomized experiment conducted in Study 4 to assess FSP impacts net of selection bias. Information about selection acquired from Studies 1 through 3 (assuming they are implemented), will be incorporated in the designs of the nonexperimental replication studies.

Research Question

- Can nonexperimental methods produce valid impact estimates?

Objectives

The main objective of Study 5 is to determine whether nonexperimental methods can yield valid estimates of FSP impacts. This study will be carried out in parallel with the experimental impact study (Study 4), with a population drawn from the same sampling frame in the same three sites. Impacts estimated in this nonexperimental study will be compared to the experimental results. If the experimental and nonexperimental estimates are sufficiently close, this will increase confidence in applying the nonexperimental method in a full-scale impact evaluation (Study 11).

Note that the objective of this study is only methodological. If the nonexperimental results differ from the results of the experiment, the experimental results are to be preferred.

Rationale

Before investing millions of dollars in a national impact study, FNS will want assurance that that study's results will be accepted by the research community at large. This study is intended to represent FNS's best shot at achieving the gold standard set by randomized experiments. It will incorporate best practices for nonexperimental study design, use appropriate statistical techniques to calculate program impacts, and ideally, will be informed by Studies 1 through 3 with regard to important factors to consider in accounting for potential selection bias.

As a trial run for a national impact study, this study should use the same methodological approach that the national study might use, but should cost substantially less. To conserve resources, the only outcomes that are included are *individual intake* and *food security*.

It is of course possible that the Study 4 experiments will detect no effects on these outcomes, while impacts are detected on food expenditures and quality of the household food supply. We do not view this as a reason to collect food use data instead. As counterarguments, we note that:

1. The purpose of this study is to test an approach, and replicating an estimate of zero is as useful as replicating a different value.

2. It is possible that selection bias first becomes a major problem when looking at individual intake; but it is not plausible that selection bias is a problem for household food use and then vanishes for outcomes which depend causally on household food use. Hence a conservative approach is to confirm that selection bias is manageable for the most distal outcome, individual intake, rather than considering only the intermediate outcomes.

3. The cost of obtaining food use data is very large, as it requires one short and one lengthy in-person interview per household per wave. These resources, if available, could be better spent increasing the sample sizes.

4. There is no guarantee that the experiment will find or fail to find impacts in one domain rather than another.

If results from this study are deemed similar those generated by the experimental trials, the study will have provided evidence that nonexperimental methods are viable. If the results are substantively different, FNS will be hard pressed to justify going ahead with a national impact study.

Sample

The sample members for this study will be drawn at the same time and from the same frames as the sample members for Study 4. They will comprise FSP participants, plus the eligible nonparticipants who were randomized into the control group for Study 4. They will thus be fully representative of households eligible for FSP in those sites at the time of the RDD screening.

We have assumed that we wish to be able to detect a 5 percentage point impact on the percent with "poor quality" diets based on HEI scores, and a 4 percentage point difference between the Study 4 and Study 5 estimates. This requires starting with 824 participants and 824 eligible nonparticipants in Wave 1. For purposes of economy, the eligible nonparticipants will be a random subset of those drawn for Study 4. Thus, their first two waves of data are "free".

Data Collection

This study requires that both the comparison group and the control group be administered a baseline interview and three followup interviews that collect data on FSP participation, selection factors, individual intake, and food security. The multiwave data collection supports estimation of random effects models to address selection bias.

Data collection for this study interlocks with and extends data collection for Study 4. Much (but not all) of the information collected by Study 4 for eligible nonparticipants in the treatment group (T's) and in the control group (C's) must be collected identically for the added group of participant (P's). Two extensions for Study 5 are that data collection continues into Waves 3 and 4, and that additional information on selection factors is collected for P's and C's in Waves 1, 2, 3, and 4. The overall scheme is described below, and illustrated in Exhibit 9.1.

1. **Household descriptors:** collected for Study 4 for T's and C's in Waves 1 and 2. Study 5 requires that this same information be collected for P's in Waves 1, 2, 3, and 4, and for C's in Waves 3 and 4.

2. **Selection factors I:** information on potential trigger events. Collected for Study 5 only, for C's and P's in Waves 1, 2, 3, and 4.

3. **Selection factors II:** stable selection factors, e.g., attitudes towards government programs. Collected for Study 5 only, for C's and P's in Wave 1.

4. **Outcomes I:** individual intake and food security. Collected for Study 4 for T's and C's in Waves 1 and 2. Study 5 requires that this same information be collected for P's in Waves 1, 2, 3, and 4, and for C's in Waves 3 and 4.

5. **Outcomes II:** food expenditures and quality of the household food supply. Collected for Study 4 only, for T's and C's in Waves 1 and 2.

Exhibit 9.1

Data Requirements for Studies 4 and 5

	Eligible nonparticipants		Participants
	Treatment	**Control**	
Wave 1 interview			
Descriptors	Study 4	Studies 4 and 5	Study 5
Selection factors I		Study 5	Study 5
Selection factors II		Study 5	Study 5
Outcomes I	Study 4	Studies 4 and 5	Study 5
Outcomes II	Study 4	Study 4	
Wave 2 interview			
Descriptors	Study 4	Studies 4 and 5	Study 5
Selection factors I		Study 5	Study 5
Outcomes I	Study 4	Studies 4 and 5	Study 5
Outcomes II	Study 4	Study 4	
Wave 3 interview			
Descriptors		Study 5	Study 5
Selection factors I		Study 5	Study 5
Outcomes I		Study 5	Study 5
Wave 4 interview			
Descriptors		Study 5	Study 5
Selection factors I		Study 5	Study 5
Outcomes I		Study 5	Study 5

Analysis

This study will use three tools to attempt to overcome selection bias: rich baseline data on factors affecting participation, multiple outcome observations, and econometric modeling. The parameters to be estimated are the impacts of FSP participation on food expenditures and food security, and the estimates are to be based on nonexperimental comparisons of these outcomes between participants and eligible nonparticipants in the pooled comparison group and control group samples. The data will support several analysis approaches. For example, random effects models could be estimated of the form:

$y_{jt} = \alpha_j + \beta X_{jt} + \gamma P_{jt} + \varepsilon_{jt}$, where
y_{jt} is the outcome for household j in period t,
α_j is a random effect for household j,
X_{jt} is a vector of descriptors of household j in period t,
P_{jt} is an indicator of FSP participation by household j in period t, and
ε_{jt}, is the residual.

Alternatively, an explicit participation model could be estimated, also with random household effects, and the model could be used to relate outcomes to predicted participation, or to conduct propensity score analysis (matching participants and nonparticipants on their predicted likelihood of participating).

The *raison d'être* of this study is the comparison between the parameter estimates obtained with the corresponding estimates from Study 4 in each site. This test will be based on the respective point estimates and their standard errors.

Timeline

This study will run a little over five years. The design phase is combined with that of Study 4, happening in the first year. Likewise, the first and second surveys occur in tandem with those of Study 4. The third and fourth waves occur in Year 3, and the analysis of the complete data occurs in Years 4 and 5.

The tasks for this study are:

1. Design phase (8–10 months)
   Combined with Study 4; only addition is incorporation of survey items related to selection

2. Field phase (32–34 months)
   a. Same as Study 4, for pre-field procedures, baseline (Wave 1) and first followup (Wave 2) interview; for purposes of this study, participants will be added to both of these data collections.
   b. Wave 3 survey
   c. Wave 4 survey

3. Analysis and reporting phase (34–36 months; overlaps with field phase, starting in month 26–28)
   a. Preliminary analysis (after baseline)
   b. Preliminary report
   c. Final analysis
   d. Final report
   e. Final briefing

Estimated Cost

Since Study 5 builds on Study 4, the only relevant costs are the incremental ones. Excluding the surveys, this study costs between $900,000 and $1,100,000, nearly all of which is associated with analysis and reporting.

The survey costs are much less than for Study 4, because the RDD screening is already done, and the expensive collection of food use data is not required. The incremental survey costs are between $1,300,000 and $1,500,000. Hence the total cost of conducting this study, conditional on Study 4 also occurring, is between $2,200,000 and $2,600,000.

Limitations and Risks

For this study to achieve its goal, it is essential that it represent FNS's best attempt to replicate experimental results nonexperimentally. Therefore, it should use all means that FNS is willing to implement in a national study—i.e., extensive baseline data collection and multiple waves of followup. Conversely, it should not use any tool that FNS could not implement nationally. Also, the sample sizes need to be large enough for the replications to be informative.

One major source of concern is the statistical power of the comparisons. The estimated resource cost of this study and Study 4 were sufficiently great that it did not seem useful to estimate costs for studies that achieved higher levels of precision. A selection bias skeptic might however fail to be convinced by general similarity of the results from Studies 4 and 5, given that each of these estimates impacts with a confidence interval of $\pm$ 3 or 4 percentage points.

Conversely, the comparison of Studies 4 and 5 might show that a selection bias problem exists when there is none. Recall that Study 5 is attempting to estimate the impact of FSP participation on *participants*, while Study 4 is estimating the impact of FSP participation on *households that normally would not participate, but are induced to do so by an intervention.* Thus, the parameter that is estimated without bias in Study 4 may be smaller by an unknown amount than the parameter than Study 5 is seeking to estimate.

The other major source of concern is that it is risky to generalize about an approach from a single successful replication. The very great cost of Study 4, however, precludes multiple tests.

Variations

The number of replication tests could be multiplied by comparing experimental and nonexperimental results for independent subsamples, e.g., for households with only elderly members, for households with children, and so on. These additional tests would however require proportionately larger samples in each subgroup for the comparisons to be informative.

Chapter Ten: Study 6—Study of Outcomes in Extant Data

This study will use data from two national surveys to explore relationships between food expenditures and the quality of the household food supply. The analyses will shed light on the primary pathway through which FSP participation could result in improved diet quality for individual FS recipients.

Research Questions

- Among low income households, is an increase in food expenditures associated with
  - an increase in household nutrient availability?
  - an increase in the nutritional quality of foods used at home?

- What factors or household characteristics mediate these relationships?
  - Is the relationship between food expenditures and the quality of the household food supply different for FSP participants than for other low-income individuals?

Objectives

The primary objective of Study 6 is to learn about the relationships between outcomes in the hypothesized causal chain shown in Exhibit 2.1. Ideally, we would like to learn about all of the links in the chain. Even in principle, however, extant data can only support exploration of the links between (a) food expenditures and household food use, using data from the NFSPS and/or the Consumer Expenditure Survey (CES); (b) overweight/obesity, using data from the National Longitudinal Survey of Youth (NSLY); (c) individual dietary intake, overweight/obesity, and food security, using the National Health and Nutrition Examination Survey (NHANES); and (d) food expenditures and household food security, using data from the Current Population Survey-Food Security Supplement (CPS-FSS) (see Appendix A).

In our judgment, the first option is most worth pursuing. Reasons for this recommendation are discussed in the final section of this chapter ("Variations"). We further recommend that assessment of the relationship between food expenditures and quality of the household food supply include analysis of both NFSPS and CES data. The NFSPS is attractive because it includes nutrient data and the food codes included in the data can support detailed food-level analyses. Limitations of the NFSPS include the fact that it is limited to FSP participants, and the data, collected in 1996, are almost 10 years old. The CES does not include nutrient data, and the lack of nutrition-oriented detail in its food coding system limits the food-level analyses that can be performed. However, the CES offers more recent information as well as data for both participants and nonparticipants. The latter feature makes it possible to assess whether the relationship between food expenditures and quality of the household food supply differs for FSP participants and other low-income (FSP-eligible) households. If a more favorable relationship exists among FSP participant households, it would be suggestive evidence that (a) the earmarking of funds for food, (b) the restrictions placed on the use of food stamps, and/or (c) the nutrition education provided by the FSP have beneficial effects above and beyond the increase of households' food purchasing power.

A secondary objective for Study 6 is to provide information on the strength of the relationship between food expenditures and the quality of the household food supply. This information will be important for designing subsequent studies. There is reason to believe that the marginal propensity to spend on food out of food stamps is in the range of 0.17 to 0.47 (Fraker, 1990, Burstein *et al.*, 2004a). Average food stamp benefits are around $90 per person per month. Thus, we might expect participants to spend about $27 more per person per month (0.3 × $90) as a result of the program. This is a significant fraction of the amount that low-income individuals spend on food—according to the 2003 CES, households in the lowest 20 percent of the income distribution spend, on average, about $98 per person per month for food at home.[32] Translating this dollar value into an expected impact on, say, individuals' usual dietary intakes or the likelihood of consuming at least five servings of fruits and vegetables per day would have implications for the sample sizes that would be necessary to detect a program effect on these outcomes.

Rationale

There is substantial evidence that FSP participation leads to increased food expenditures (Burstein *et al.*, 2004a). If analysis of the NFSPS and CES demonstrates a positive relationship between food expenditures and the quality of the household food supply, then this link in the hypothesized causal chain will have empirical support. This would suggest the desirability of further research to explore the subsequent link in the chain (improved household food supply → improvement in individual dietary intake) and ultimately to research that might look directly at the relationship between FSP participation and individual intake.

If the analysis fails to demonstrate this relationship, there would be reason to question the value of research focused on FSP impacts on individual dietary intake. Subsequent exploratory analyses may help explain the lack of the anticipated relationship. For example, increases in food expenditures may be due to purchasing higher-priced versions of the same foods or foods that are of lower nutritional quality (e.g., foods that are less nutrient-dense or foods that are higher in fat, added sugars, or sodium).

This study will also yield important information on the factors or household characteristics that mediate relationships between food expenditures, quality of the household food supply, and household food security. One likely mediating factor is baseline food expenditures. For example, a $50 per month difference in household food expenditures between two households with very low expenditures might have a much stronger relationship to nutritional quality than would a $50 per month difference between two households that both have higher monthly food expenditures. Understanding the nature of the functional form of the relationships is both an end in itself, and a basis for sampling and model specification in subsequent studies. Information on another likely set of mediating factors, nutrition-related knowledge and attitudes, is available in the NFSPS but not in the CES. Therefore, the study can explore how these factors might affect the relationship of interest, but

---

[32]   Source: Quintiles of income before taxes: Shares of average annual expenditures and sources of income, Consumer Expenditure Survey, 2003. Total annual household expenditures for food at home = $2,127; mean number of persons per household = 1.8.

it will not be able to assess how these factors are related to FSP participation, as either causes or effects.

Food Expenditures and Quality of the Household Food Supply

Sample

The NFSPS collected data from both participants and nonparticipants. However, the data of interest for this study—data on household food use—were collected only from FSP participants. Therefore, the sample for the NFSPS analysis comprises 957 FSP households.

The CES is comprised of two independent surveys: the quarterly Interview Survey and the Diary Survey. The Diary Survey provides detailed data on expenditures for frequently purchased items including food. The most recent publicly available CES data is from the year 2003. This data set includes more than 7,500 consumer units (households). If we assume that roughly 12 percent of households will be classified as low-income, then we would expect that the analysis sample will include about 900 low-income households. Some will be FSP participants, others will be income-eligible nonparticipants.

Data Elements: NFSPS

The NFSPS includes detailed data on household food use for a 7-day period.[33] The available measures are described below.

Expenditures for Food Used at Home
The NFSPS includes two measures of the dollar value of food used from household food supplies (Zambrowski and Ohls, 1999). One of these (ALLPAID) includes the value of non-purchased food (e.g., gifts, food obtained from WIC, and home-grown food). The other measure (BSTPAID) includes only the value of purchased food. For the purposes of this study, BSTPAID is the measure of interest. For analysis purposes, it will be desirable to normalize this variable to express the data in terms of expenditures per equivalent nutritional units (ENUs), using energy as the nutrient.[34]

Nutrient-Level Measures
The nutrient-level measures shown in Exhibit 10.1 are measures of nutrient availability per ENU that were used in the NFSPS "nutrient availability" report (Cohen *et al.*, 1999) and are readily available in the NFSPS dataset.

---

[33] The NFSPS also includes survey items that ask about food expenditures. These items are essentially the same as those used in the CPS-FSS. As noted in Chapter Two, participants in the TWG meeting concluded that such recall-based data on food expenditures are neither reliable nor detailed enough to meet the needs of the research agenda.

[34] Available nutrients in the household food supply are normalized by ENUs to account for differences among households in (a) nutritional needs as determined by household composition and (b) percent of meals eaten from home food supplies by each household member.

Exhibit 10.1

Nutrient Variables Available in NFSPS

Constructed Variable[a]	Description
PRDCALS	% of RDA calories
PRDAVITA	% of RDA vitamin A
PRDAVITC	% of RDA vitamin C
PRDAB6	% of RDA B6
PRDAFOLA	% of RDA Folate
PRDACALC	% of RDA Calcium
PRDAIRON	% of RDA Iron
PRDAZINC	% of RDA Zinc
YS100B6 (B6) YS100CLC (Calcium) YS100FOL (Folate) YS100IRN (Iron) YS100PRT (Protein) YS100VTA (Vitamin A) YS100VTC (Vitamin C) YS100ZNC (Zinc)	Binary variables indicating availability of 100 percent or more of daily RDA (1 = nutrient availability >= 100 percent of the nutrient)
YS75B6 (B6) YS75CLC (Calcium) YS75FOL (Folate) YS75IRN (Iron) YS75PRT (Protein) YS75VTA (Vitamin A) YS75VTC (Vitamin C) YS75ZNC (Zinc)	Binary variables indicating availability of 75 percent or more of daily RDA

[a]Variables were calculated as total amount of nutrient available in the week divided by [ENU × daily RDA for adult male × 7]

Source: Cohen et al., (1999) Tables V.2-V.5; Zambrowski and Ohls (1999) Table A.3

In addition to these nutrients, the NFSPS dataset includes data for many other nutrients, including other macronutrients (carbohydrate, total fat, saturated fat), fiber, cholesterol, sodium, other vitamins (thiamin, niacin, riboflavin, Vitamin B12, vitamin E), and other minerals (copper, magnesium, niacin, phosphorus, and potassium) (Zambrowski and Ohls, 1999). Our review of the documentation suggests that the data file includes information on the total amount of these nutrients used from the household food supply, as well as information on the amount that came from purchased *versus* non-purchased foods. To be used in conjunction with the variables listed above, these data would have to be normalized to be expressed per ENU. This would be straightforward for variables that had a 1989 RDA. For nutrients without RDAs, other benchmarks could be used, for example, the *Dietary Guidelines for Americans* and National Research Council standards of no more than 30 percent of energy from fat, less than 10 percent of energy from saturated fat, no more than 300 mg cholesterol, and no more than 2400 mg of sodium. Since these standards do not apply to children under the age of 2, adjustments would have to be made for households that included members younger than 2. Our costing assumes that the analysis will include, in addition to the nutrients used in Cohen *et al.*, (1999),

the following nutrients: carbohydrate, total fat, saturated fat, cholesterol, sodium, fiber, vitamin B_{12}, magnesium, and potassium. These are nutrients that are known to be of concern for the US population and/or were cited as nutrients of concern in a recent analysis of the nutrient intakes of WIC participants (Institute of Medicine (IOM), 2004b).

The available data can also be used to assess nutrient density, or nutrients available per unit of energy (Nutrient per ENU ÷ Energy per ENU). The NFSPS documentation suggests that nutrient density was computed for some nutrients (Zambrowski and Ohls, 1999). There are several "Density [nutrient X]" variables on the file (for all the nutrients shown in Exhibit 10.1); however, these variables were not used in the "nutrient availability" report (Cohen *et al.*, 1999).

Another option that was mentioned at the TWG meeting was development of a household-level HEI. While this is an attractive idea, such methodological work is probably outside the scope of this research agenda. We discuss this option further in the "Variations" section.

An obvious question is whether the NFSPS data should be updated to incorporate the most recent version of the RDAs, as defined in the DRIs. While this is theoretically possible, we believe that this would not be the best use of resources in this multi-faceted and complex research agenda. Our rationale for this opinion is discussed in the "Variations" section.

Food-Level Measures
Exhibit 10.2 shows food group variables reported in Cohen et al (1999). Each of these variables can be linked to an expenditure amount. For Study 6, these expenditure amounts will be converted to expenditure dollar amounts per ENU.

To get at the concept of "quality" of foods used at the household level, foods could be categorized based on nutrient characteristics and recommended consumption. For example, fruits, vegetables, and whole grains could be considered foods that should be consumed "frequently," while fats and oils, and sugar and sweets could be considered foods that should be consumed "infrequently." Summary measures could be created for the dollar amounts per ENU spent on foods in each group. In addition to the food groups shown in Exhibit 10.2, the NFSPS dataset includes a detailed food-level file that includes information on all foods used during the 7-day-food-use period. Each food-level record includes a description and food code. For purposes of Study 6, it would be useful to use these data to develop additional food groups that would permit a more detailed assessment of the characteristics of foods used, e.g., fresh fruits and vegetables, *versus* canned or frozen, low-fat or nonfat milk and cheese, *versus* regular varieties, low- and high-fat meats, etc. Our costing assumes that some additional, more finely-tuned, food groups will be created.

Exhibit 10.2

Food Group Variables Available in NFSPS

Food Group/Variable Name	Description
FRUITS AND VEGETABLES	
CFRT	1 = (vitamin-C rich fruit)
OFRT	1 = (other fruit)
HVEG	1 = (high nutrient vegetables)
OVEG	1 = (other vegetables)
POT	1 = (potatoes)
VMIX	1 = (mixtures, mostly vegetables, condiments)
GRAIN PRODUCTS	
WCRL	1 = (whole grain cereal, includes oatmeal)
OCRL	1 = (other cereal, includes cream of wheat)
WFLOUR	1 = (whole grain flour, oat, barley)
FLOUR	1 = (flour, meal, rice, pasta, includes corn)
WBRD	1 = (whole grain bread)
OBRD	1 = (other bread, inc. white)
SBKRY	1 = (bakery products, not bread)
GMIX	1 = (grain mixtures)
MEAT AND ALTERNATIVES	
BACON	1 = (bacon, sausage, luncheon meats)
BEANS	1 = (dry beans, peas, lentils)
EGGS	1 = (eggs)
FISH	1 = (fish, shellfish)
LMT	1 = (lower cost red meats, variety meats)
HMT	1 = (higher cost red meats, variety meats)
MMIX	1 = (mixtures, mostly meat, poultry, fish, egg, legume)
NUTS	1 = (nuts, peanut butter)
PLTY	1 = (poultry)
MILK, CHEESE, CREAM	
CHES	1 = (cheese)
CREAM	1 = (cream mixtures, mostly milk)
MILK	1 = (milk, yogurt)
OTHER FOODS	
ALC	1 = (alcohol)
BEV	1 = (soft drinks, punches, ades)
FATS	1 = (fats, oils)
HERBS	1 = (seasonings)
SUGARS	1 = (sugar, sweets)
TEA	1 = (tea, coffee)

Source: Cohen et al., (1999) Table V.6; Zambrowski and Ohls (1999) Table A.3

Data Elements: CES

As noted previously, the CES has information on expenditures only (i.e., it does not have nutrient data). The CES data file includes information about expenditures for broad food groups as well as for specific types of food. Exhibit 10.3 shows the major food groups and the individual foods included in each group. Expenditure amounts are available at both levels. Averaging the expenditure amounts across the two one-week food diaries completed by a household will produce measures expressed in terms of dollars per week spent on a particular food group or item. Furthermore, using the data on household characteristics, these measures can be expressed in terms of dollars per week per adult male equivalent on a particular food group or item. The CES data cannot be expressed in terms of ENUs because there are no items on the data set that measure the proportion of meals that are eaten from home food supplies.

The only option for using the CES data to assess the quality of the household food supply is to categorize foods into groups based on nutrient characteristics, as discussed in the preceding section on the NFSPS. These groupings will have to be more broad than those that could be developed using the NFSPS data because the CES data and codes are commodity-oriented rather than nutrient-oriented. For example, one cannot distinguish between milks, other dairy products, meats, or salad dressings on the basis of fat, or distinguish breads, cereals, and other grain products based on whether or not they were whole grain.

Exhibit 10.3

Measures of Food Expenditures in CES

FOOD GROUP	ITEM
FOODTOT	Food, total
	FOODHOME + FOODAWAY
FOODHOME	Food at home, total
	CEREAL + BAKERY + BEEF + PORK + OTHMEAT +POULTRY + FISHSEA + EGGS + MILKCRM + OTHDAIRY +FRSHFRUT + FRSHVEG + PROCVEG + SWEETS +NOALCBEV + FATOILS + MISCFOOD
CEREAL	Cereal and cereal products
	010110 Flour
	010120 Prepared flour mixes
	010210 Cereal
	010310 Rice
	010320 Pasta, cornmeal, other cereal products
BAKERY	Bakery products
	020110 White bread
	020210 Bread other than white
	020310 Fresh biscuits, rolls, muffins
	020410 Cakes and cupcakes, fresh and other, excluding frozen
	020510 Cookies, excluding refrigerated dough

Exhibit 10.3

Measures of Food Expenditures in CES

FOOD GROUP	ITEM
	020610 Crackers, excluding crumbs
	020620 Bread and cracker products
	020710 Doughnuts, sweet rolls, coffeecakes, fresh and other, excluding frozen
	020810 Frozen refrigerated and canned bakery products, such as biscuits, rolls, muffins, cakes, cupcakes, doughnuts, pies, tarts, turnovers, and miscellaneous products, including dough, and batter
	020820 Pies, tarts, turnovers, fresh and other, excluding frozen
BEEF	Beef
	030110 Ground beef, excluding canned
	030210 Chuck roast, excluding canned
	030310 Round roast, excluding canned
	030410 Other beef roast, excluding canned
	030510 Round steak, excluding canned
	030610 Sirloin steak, excluding canned
	030710 Other steak, excluding canned
	030810 Other beef, excluding canned
PORK	Pork
	040110 Bacon
	040210 Pork chops
	040310 Ham, excluding canned
	040410 Other pork, excluding canned
	040510 Pork sausage, excluding canned
	040610 Canned ham
OTHMEAT	Other meats
	050110 050210 050310 050410 050900
	050110 Frankfurters, excluding canned
	050210 Bologna, liverwurst, salami, excluding canned
	050310 Other lunchmeat
	050410 Lamb and organ meats, excluding canned
	050900 Mutton, goat, game
POULTRY	Poultry
	060110 Fresh and frozen whole chicken
	060210 Fresh or frozen chicken parts
	060310 Other poultry
FISHSEA	Fish and seafood
	070110 Canned fish, seafood and shellfish
	070230 Fresh fish and shellfish
	070240 Frozen fish and shellfish

Exhibit 10.3

Measures of Food Expenditures in CES

FOOD GROUP	ITEM
EGGS	Eggs
	080110 Eggs
MILKCRM	Fresh milk and cream
	090110 Fresh milk all types
	090210 Cream
OTHDAIRY	Other dairy products
	100110 Butter
	100210 Cheese
	100410 Ice cream and related products, including frozen yogurt
	100510 Other dairy products, including powdered milk, and fresh, canned and non-frozen yogurt
FRSHFRUT	Fresh fruits
	110110 Apples
	110210 Bananas
	110310 Oranges
	110410 Other fresh fruits
	110510 Citrus fruits excluding oranges
FRSHVEG	Fresh vegetables
	120110 Potatoes
	120210 Lettuce
	120310 Tomatoes
	120410 Other fresh vegetables
PROCFRUT	Processed fruits
	130110 Frozen orange juice
	130121 Frozen fruits
	130122 Frozen fruit juices
	130211 Fresh fruit juices
	130212 Canned/bottled fruit juices
	130310 Canned fruits
	130320 Dried fruits
PROCVEG	Processed vegetables
	140110 Frozen vegetables
	140210 Canned beans
	140220 Canned corn
	140230 Miscellaneous canned vegetables, not collected in a separate UCC
	140310 Other processed dried vegetables, such as squash, not collected in a separate UCC
	140320 Dried peas
	140330 Dried beans
	140340 Dried carrots, onions, leafy greens, and

Exhibit 10.3

Measures of Food Expenditures in CES

FOOD GROUP	ITEM
	cabbage
	140410 Frozen vegetable juices
	140420 Fresh/canned vegetable juices
SWEETS	Sugar and other sweets
	150110 Candy and chewing gum
	150211 Sugar
	150212 Artificial sweeteners
	150310 Jams, jellies, preserves and other sweets
NONALBEV	Nonalcoholic beverages
	170110 Cola drinks
	170210 Other carbonated drinks
	170310 Coffee, roasted
	170410 Coffee, instant or freeze dried
	170510 Noncarbonated fruit flavored drinks, including lemonade-non frozen
	170520 Tea
	170530 Other noncarbonated beverages and ice, excluding coffee and tea
FATSOILS	Fats and oils
	160110 Margarine
	160211 Fats and oils
	160212 Salad dressings
	160310 Non-dairy cream substitutes
	160320 Peanut butter
MISCFOOD	Miscellaneous foods
	180110 Soup
	180210 Frozen meals
	180220 Frozen prepared food other than meals
	180310 Potato chips and other snacks
	180320 Nuts
	180410 Salt, other seasonings & spices
	180420 Olives, pickles, relishes
	180510 Sauces and gravies
	180520 Other condiments
	180611 Prepared salads
	180612 Prepared desserts
	180620 Baby food
	180710 Miscellaneous prepared foods including items such as canned meats (see UCC's 030110 - 030810, 040410 - 040510, 050110, 050310 - 050410, 060110 - 060310), fresh and canned ethnic foods, fresh and canned pizza

Exhibit 10.3

Measures of Food Expenditures in CES

FOOD GROUP	ITEM
	180720 Vitamin supplements
FOODAWAY	Food away from home
	190111 Lunch at Fast Food
	190112 Lunch at Full Service
	190113 Lunch at Vending Machine
	190114 Lunch at Employer
	190115 Lunch at Board
	190116 Lunch at Catered Affairs
	190211 Dinner at Fast Food
	190212 Dinner at Full Service
	190213 Dinner at Vending Machine
	190214 Dinner at Employer
	190215 Dinner at Board
	190216 Dinner at Catered Affairs
	190311 Snacks at Fast Food
	190312 Snacks at Full Service
	190313 Snacks at Vend Machine
	190314 Snacks at Employer
	190315 Snacks at Board
	190316 Snacks at Catered Affairs
	190321 Breakfast at Fast Food
	190322 Breakfast at Full Service
	190323 Breakfast at Vending Machine
	190324 Breakfast at Employer

Analysis Approach

Multivariate regression models will be used to examine the relationships between food expenditures and quality of the household food supply in the two extant data sets. Specialized statistical software (e.g., SUDAAN, SAS PROC SURVEYREG) should be used to take account of the complex sampling designs of the NFSPS and CES in calculating standard errors.

For the NFSPS data, we assume that the dependent variables will include:

- nutrients expressed as percent of RDA or in comparison with other reference standards (nutrients listed in Exhibit 10.1, plus carbohydrate, total fat, saturated fat, fiber, cholesterol, sodium, vitamin B12, magnesium, and potassium);

- expenditures on "frequently consume" *versus* "infrequently consume" foods (constructed from food groups shown in Exhibit 10.2 as well as expanded food groups created from food codes and descriptions available in the food-level data file);

- probability of available nutrients satisfying a reasonable threshold nutrient density.

These outcomes will be related to expenditures for food at home per ENU.

With regard to the threshold variables, the NFSPS used cutoffs of 100 percent and 75 percent of the RDA. These analyses could be repeated. However, it may be worth exploring the possibility of developing a separate threshold for each household based on the proportion of meals each household member consumed from the household food supply. If all individuals consumed all meals from the household supplies, then the threshold would be the maximum threshold (e.g., 100 or 75 percent of the RDA). If, as is often the case, some household members consumed meals obtained from other sources, then the threshold for that household would be adjusted downward.

The CES data will be used for two analyses. The first will relate expenditures on food at home to total expenditures on food; the second will relate expenditures for "frequently consume" and "infrequently consume" foods to expenditures on food at home. All expenditure values will be expressed per AME. The sample will be limited to low-income (FSP-eligible) households.

To determine what factors mediate the relationships of expenditures to outcomes, models will be fit with interaction effects between the mediating factors and the expenditure variable. The NFSPS and the CES include information on the following potential mediators:

- household composition
- race/ethnicity
- urban/suburban/rural
- region of the country
- income relative to poverty
- education
- dietary knowledge, attitudes, behaviors (NFSPS only)
- FSP participation (CES only)

Neither data set has information on sample members' immigrant status or English proficiency, mental or physical health conditions, or overweight/obesity.

Timeline

This study will have two phases, lasting 15 to 18 months in total. In the design phase, the analytic files will be constructed and the draft and final analysis plans will be prepared. In the analysis and reporting phase, the analyses will be carried out and the results will be presented. The study tasks are as follows:

1. Design (5–6 months)
   a. Initial meeting with FNS and followup memorandum
   b. Analysis plan

2. Analysis and reporting (10–12 months)
    a. Construction of analysis files
    b. NFSPS analysis
    c. CES analysis
    d. Draft and final report
    e. Briefing

3. Management
    a. Prepare monthly progress reports

Estimated Cost

This study is estimated to cost between $350,000 and $400,000.

Limitations and Risks

The major limitation is that this study can only look at the relationship between food expenditures and the quality of the household food supply (measured on the basis of household food use in the NFSPS and household food expenditures in the CES). Even if a positive association is found, it cannot be assumed that this translates into improvement in the next outcomes on the causal chain (individual dietary intake and food security). Conversely, failure to find a relationship does not necessarily prove that the FSP is powerless to improve dietary intake (although it would provide evidence in that direction). Alternative explanations could be that (a) the sample sizes are too small to detect an effect of food expenditures on food use; (b) that the measures in these data sets are inexact; (c) that increased food expenditures go for better food consumption outside the home; and/or (d) that the current FSP, with its increased focus on nutrition education, could make a difference in households' food purchasing decisions.

Variations

There are several potential variations to the basic analysis we have described above. Some involve expansion of the NFSPS analysis, and others involve use of other extant datasets.

We see three potential variations for expansion of the planned NFSPS analysis. Two of these would considerably increase the resources required for this study. The first possibility is development of a household-level HEI. The NFSPS dataset includes the data necessary to look at the nutrient components of the HEI score (total fat, saturated fat, cholesterol, and sodium). However, the data required to look at the food-level components of the HEI—servings of fruits, vegetables, grains, dairy, and meats, plus a measure of overall dietary variety—are not included. In theory, one could derive this information by linking the food codes reported in the NFSPS to the USDA pyramid servings database and using the gram weights reported in NFSPS to translate information on the total amount of food used into the total number of servings available. A 100 percent match in food codes is unlikely, given that the NFSPS data date back to 1996 and the pyramid database currently includes

codes reported in more recent rounds of NHANES. Thus, time and nutrition expertise would be required to make decisions about how to handle foods that do not merge. In addition, the basic HEI algorithm for food group servings would have to be adjusted to account for household composition (the recommended number of servings varies by age group and there currently are no recommendations for children under 2) and for the fact that some meals are consumed away from home. While we find this possibility intriguing, we believe that methodological work to create a new measure is beyond the scope of this research agenda. If FNS were to fund such research through another mechanism and a validated household-level HEI (or comparable measure) was available when Study 6 is implemented, it should certainly be included in the analysis.

The second possibility is updating the reference RDAs with the most recent RDAs. For most nutrients, this could potentially be a straightforward task. However, we question the value of comparing nutrient data from 1996 to nutrient standards developed in 1999-2001 for several reasons. First, the comparison has little face validity, given that food composition has changed over time: the 1999-2001 versions of foods consumed in 1996 may have different nutrient profiles. In addition, for some nutrients of interest, namely folate and vitamin A, the unit of measure has changed. This means that the nutrient values in the NFSPS file cannot be compared with the most recent RDAs. Finally, if the nutrient standards are updated, one could argue that the food expenditures data should be similarly updated. In our estimation, updating the RDAs is not necessary and may, in fact, compromise the analysis. Given that the main objective of this analysis is to look at relationships between food expenditures and diet quality, rather than estimate current diet quality of FSP households, it makes sense to avoid manipulating either side of this equation because of temporal concerns.

The final variation on NFSPS analysis would be to assess the relationship between food expenditures and food security. This does not seem to be a reasonable path to pursue, however, because of the substantial mismatch in time periods: the food security module used in the NFSPS used a 12-month reference period.

With regard to existing data sets other than the NFSPS and CES, an examination of the diet-related outcomes included in other national surveys (Exhibit A.1) suggests three possibilities. We do not see any of these as a top priority for this research agenda.

The first possibility is to use data from the National Longitudinal Survey of Youth (NLSY) to explore the relationship between food expenditures and overweight/obesity. The 1979 Cohort (NLSY79) and the Children's Cohort (NLSY-C) offer longitudinal data on both of these outcomes, spanning several years. The hypothesis to be tested would be whether individuals in households that spend more on food are less likely to become or remain overweight or obese, because of better quality diets and/or increased food security. It would be hard to draw any persuasive conclusions from such an analysis, however. For one thing, the NLSY lacks information on the intervening variables (diet quality, food security). Second, it also lacks data on physical activity, an important predictor of weight status. Finally, the measure of food expenditures is extremely crude. Respondents are asked simply how much they spend on food used at home and away from home "in an average week".

The second non-NFSPS possibility is to use NHANES data to look at the relationship between individual intake, overweight/obesity, and food security. We also view this analysis as low priority for the research agenda that is the subject of this report. Until it is established that increased

household food expenditures improve individual diet quality and/or food security, establishing the further relationships among these variables or the even more distal outcome of overweight/obesity is of limited policy interest. Even for exploring these relationships NHANES is far from ideal. Its cross-sectional nature makes causal inference highly questionable.

The final possibility is to conduct further analysis of the relationship between food expenditures and food security using the CPS-FSS. The annual Food Security Supplement of the CPS contains not only a food security battery (referring to the past 30 days), but also information on household food expenditures (referring to the past week) including:

- total amounts spent (including food stamps) at
  - supermarkets and grocery stores
  - meat markets, produce stands, bakeries, warehouse clubs, and convenience stores
  - restaurants, fast food places, cafeterias, and vending machines
  - other places where food was bought

- amounts spent at these places on "non-food items such as pet food, paper products, detergents, or cleaning supplies"

- "usual" weekly expenditures for food.

We view this analysis as a lower priority for two reasons. First, some of the desired research has already been done or is in progress at ERS. The cross-sectional data from this module are analyzed and tabulated by ERS on a regular basis, including a calculation of median weekly food expenditures by food security status (see for example Nord *et al*. 2005). These published analyses have established a strong positive relationship between food expenditures and food security. Research is also proceeding at ERS to exploit the possibility of linking successive years of the CPS data in order to explore the dynamic relationships and control for unobserved household characteristics.[35]

The second reason is that the chief concern for Studies 6 through 9 is improving our understanding the relationships between expenditures, household food supply, and individual intake before undertaking a costly national study of the FSP. Understanding the relationships to food security is of lesser importance in the preliminary studies because this outcome is straightforward and relatively inexpensive to measure. Nonetheless, if resources permitted, analyses of the CPS data that focused on the issues pertinent to this research agenda, including the role of food assistance programs in mediating the relationship between food expenditures and food security, could be a worthwhile addition.

---

[35] The CPS sampling structure provides fifty percent overlap from one year to the next. It must be noted, however, that the CPS interview unit is the address, not the people living there. If a household moves away, the survey replaces the respondents with the new people living at that address. Hence the overlapping sample excludes movers, and is in that way nonrepresentative.

Chapter Eleven: Study 7—Qualitative Study of Diet-Related Implications of Changing Food Expenditures

This study uses in-depth retrospective interviewing with relatively small numbers of households to explore the process by which changes in food expenditures affect (or do not affect) dietary patterns and food security.

Research Questions

- How and why do households change their food purchasing patterns and diet when they increase or reduce their food expenditures?

- What factors influence the household in changing its level of food expenditures, and how are differing *reasons* for change associated with different *consequences* of the change?

Objectives

The objective of this study is to describe the food expenditure and other food acquisition practices of low-income households, and to learn if there is an association between those practices and broad dietary patterns. Topics to be pursued include:

- When a low-income household increases its food expenditures, what is the household attempting to buy with the increased expenditure? For example,
  - larger servings of everything
  - better tasting food
  - more nutritious food
  - food that is easier or quicker to prepare
  - treats such as snacks or expensive cuts of meat
  - entertainment from eating out
  - social value from feeding non-household members
  - independence from government and community programs
  - lesser reliance on friends and relatives

- Conversely, when a household reduces its food expenditures, on what dimensions does it make the cuts?

- What are the differences in types and amounts of food eaten by household members as expenditure patterns change?

- Do changes in expenditure levels or patterns seem associated with changes in the attitudes and behaviors related to food security (e.g., skipping meals or worrying about food money running out)?

The reasons that a family changes its food expenditures may be related to the results of the change. For example, expenditures may increase (a) even in the absence of change in available resources, because of greater perceived needs, e.g., as children grow, or (b) because more resources are available for food purchasing, from an increase in income or reduced competing needs. One might hypothesize that the first situation would lead to an increase in nutrient intake, while the second might result in more purchases of foods that are preferred but more costly.

Rationale

Existing research provides limited insight into the question of whether an increase in food expenditures of the size associated with FSP participation (about $27 per person per month; see discussion in Chapter 10) can be expected to alter diet quality visibly at either the household or individual level. Some research, based on cross-sectional data, has examined income-related differences in diet or food purchasing among low-income households, suggesting, for example, that incremental income is more likely to be used to buy meat and bakery products as opposed to fruits and vegetables (Wilde *et al.,* 2000, Blisard *et al.,* 2004). Economizing on food purchases (e.g., buying foods on sale, or buying store brands) is more prevalent among low-income than higher-income households (Leibtag and Kaufman 2003), and increments in income may lead to less economizing.

In addition to the nature of the food purchased, differing food expenditure levels may reflect differences in a variety of other behaviors. Low-income households may obtain food from sources other than purchases in stores, including home production, food pantries and soup kitchens, eating with friends or relatives, eating at restaurants. Changes in the use of these sources will clearly be related to food expenditures, but how they might affect diet quality is unclear. Food expenditures may also support food consumption by non-household members, either for socialization purposes or to help out needy persons, and such expenditures will make no direct contribution to household members' diets. Existing research provides little information on either the extent of such behaviors or the extent to which they account for changes in food expenditures.

Sample

A purposive sample of low-income households will be recruited for the study. To the extent possible, the sample will consist of households that have experienced a recent change in income, such as getting or losing a job, or beginning or ending TANF or FSP participation. At least half of the households will not be current or recent FSP participants, because it is particularly important to learn about the patterns that would be expected in the absence of the program. Community organizations, employers, and local food stamp offices may be used in recruiting subjects for the study.

As in the qualitative study of FSP participation (Study 2), a total of 60 households will be selected, with 20 in each of three groups: households consisting of elderly persons, single-adult households with children, and multiple-adult "working poor" households. Samples will be drawn in three different locations for geographic diversity, including two urban and one small town or rural location. Interviews within each major group will be roughly evenly divided among the three locations. Within each group in each location, interviewers will seek a racial/ethnic balance roughly similar to that of the FSP caseload in that site.

Data Collection

Persons selected for study will be interviewed in person two to four times during a 4-month period, and will be telephoned weekly for a brief status check. The first interview will occur as soon as possible after recruitment into the study. After the initial in-person interview, the interviewer will make a brief weekly telephone call to determine whether any major changes in food expenditures or dietary patterns have occurred. If major changes are reported, an in-person interview will be scheduled (up to a maximum of four in-person interviews for a household). If a household reports no major changes during the four months, a second in-interview will occur at the end of the period. We estimate that an average of three interviews per household will be conducted.

The initial interview will first explore the household's current food acquisition and eating patterns, where the reference period is roughly the week preceding the interview. The interviewer will not attempt to obtain detailed data on expenditure amounts, items purchased, or items eaten. Rather, the focus will be on general patterns of food shopping, factors considered when making food purchasing decisions, types and numbers of meals eaten at home and away, eating of non-purchased foods, and serving food to non-household members. After establishing current patterns, the interviewer will ask the respondent to think back to the last time when expenditure and dietary patterns changed. The interviewer will explore the nature of and reasons for the changes, with particular attention to the respondent's own rationale, intentions, and attitudes.

Subsequent in-person interviews will generally occur after the respondent has indicated that a change has occurred. The interview will obtain the respondent's account of what changed and why, a new picture of current patterns, and a description of patterns existing just before the change.

Interviews will have varying duration, but first interviews are generally expected to last two to three hours. Subsequent in-person interviews may be a bit shorter. Weekly telephone contacts will be about 10 minutes. Respondents will be offered an incentive payment. We have assumed an incentive amount of $50 for the initial interview and follow-up calls, and $25 for each subsequent in-person interview.

Analysis

Two reports will be prepared, as with the qualitative study of FSP participation (Study 2). The first report will be organized by household, presenting a narrative story for each household studied. The narrative will begin with a general characterization of the household's circumstances and its typical food acquisition and eating behavior. The typical pattern may reflect the situation at the first interview, at some other point, or an amalgam. The narrative will go on to describe significant changes that the household made in expenditures and/or diet, along with the respondent's report of the reasons for and consequences of those changes.

The second report will use data from all households to develop a comprehensive list of types of changes in individual households' patterns of food acquisition and eating patterns, and differences in patterns across households. This report will include an interpretative component in which the

researchers develop hypotheses about the ways that changes in food expenditures are linked to changes (or absence of changes) in dietary patterns.

Timeline

The overall study is expected to take 20 to 26 months, in three major phases. The design phase will require six to eight months, the field phase will take eight to 10 months, and six to eight months will be needed for the analysis and reporting phase. Key tasks within the phases are listed below.

1. Design phase (6–8 months)
   a. Initial meeting with FNS and follow-up memorandum
   b. Develop research protocols
   c. Develop procedures and obtain permissions for recruiting sample
   d. Prepare OMB submission

2. Field phase (8–10 months)
   a. Train interviewers
   b. Recruit sample and schedule initial interview
   c. Conduct initial interviews
   d. Transcribe interviews, write up interview notes
   e. Weekly contacts
   f. Conduct interview 2-4
   g. Transcribe interviews, write up interview notes

3. Analysis and reporting phase (6–8 months)
   a. Household stories report
   b. Synthesis report
   c. Final briefing

Estimated Cost

This study is estimated to cost between $900,000 and $1,000,000.

Limitations and Risks

The principal risk for this study lies in the breadth and depth of the topics to be studied. Food acquisition and eating patterns potentially involve enormous amounts of detail, as evidenced by the complicated and burdensome survey instruments that are used to measure food expenditures and dietary intake. The research design and the researchers in the field will have to be carefully and sensitively targeted to the relevant patterns of household reasoning and behavior to obtain useful information without being swamped in unusable details.

Variations

The suggested design includes multiple interviews to make it possible to observe multiple changes in the households' food expenditure and dietary patterns over time. It is conceivable that sufficient information could be obtained from a single interview, avoiding the substantial cost of the follow-up telephone calls and interviews. If this alternative is seriously considered, a thorough pre-test should be conducted to make sure that the retrospective component of the qualitative interview can capture information that is sufficiently comprehensive and credible.

Chapter Twelve: Study 8—Survey of Food Expenditures and Diet-Related Outcomes

Study 8 uses a panel survey with two waves of data collection, separated by one year and conducted in a limited number of sites, to examine the relationship between food expenditures, diet quality, and food security.

Research Question

- To what extent are greater food expenditures among low-income households associated with better diet quality?

- If food expenditures are not closely related to diet quality, what are the results of higher expenditures?

- To what extent are greater food expenditures among low-income households associated with greater food security?

Objectives

Study 8 aims to provide a better understanding of the connections between food expenditures, diet quality, and food security for low-income households in general. The premise is that better information on these relationships will support the design of an impact study of the FSP (Study 11) that can provide more conclusive information on the program's effects. For example, if Study 8 finds that an increase in food expenditures of $1 per person per day is associated with a specific increase in an individual's mean score on the Healthy Eating Index (HEI) or in the likelihood of a household being food secure, the study will provide information on the sample sizes necessary to detect those effects at a 5 percent significance level with 80 percent power.

Alternatively, this study might find no clear relationship between food expenditures and diet quality or food security. As Exhibit 12.1 suggests, there are many scenarios in which an increase or decrease in food expenditures would not result in an equivalent change in diet quality.[36] This study will also seek to determine the extent to which greater food expenditures are used to achieve other household goals, including convenience, luxury, and taste.

The design of Study 8 will incorporate relevant findings from Studies 6 and 7 about linkages between food expenditures and aspects of diet quality, and the household characteristics that mediate these relationships.

---

[36] To avoid complicating still further an already complicated diagram, some details have been omitted. The arrows that go to the "key predictors" box should be understood as pointing to both expenditures at home and expenditures away from home. Furthermore, the determinants of food expenditures have direct effects on the "consequences" (shopping and food serving patterns, household food use, etc.) as well as indirect effects on these outcomes through their effects on food expenditures.

Exhibit 12.1

Determinants and Consequences of Food Expenditures
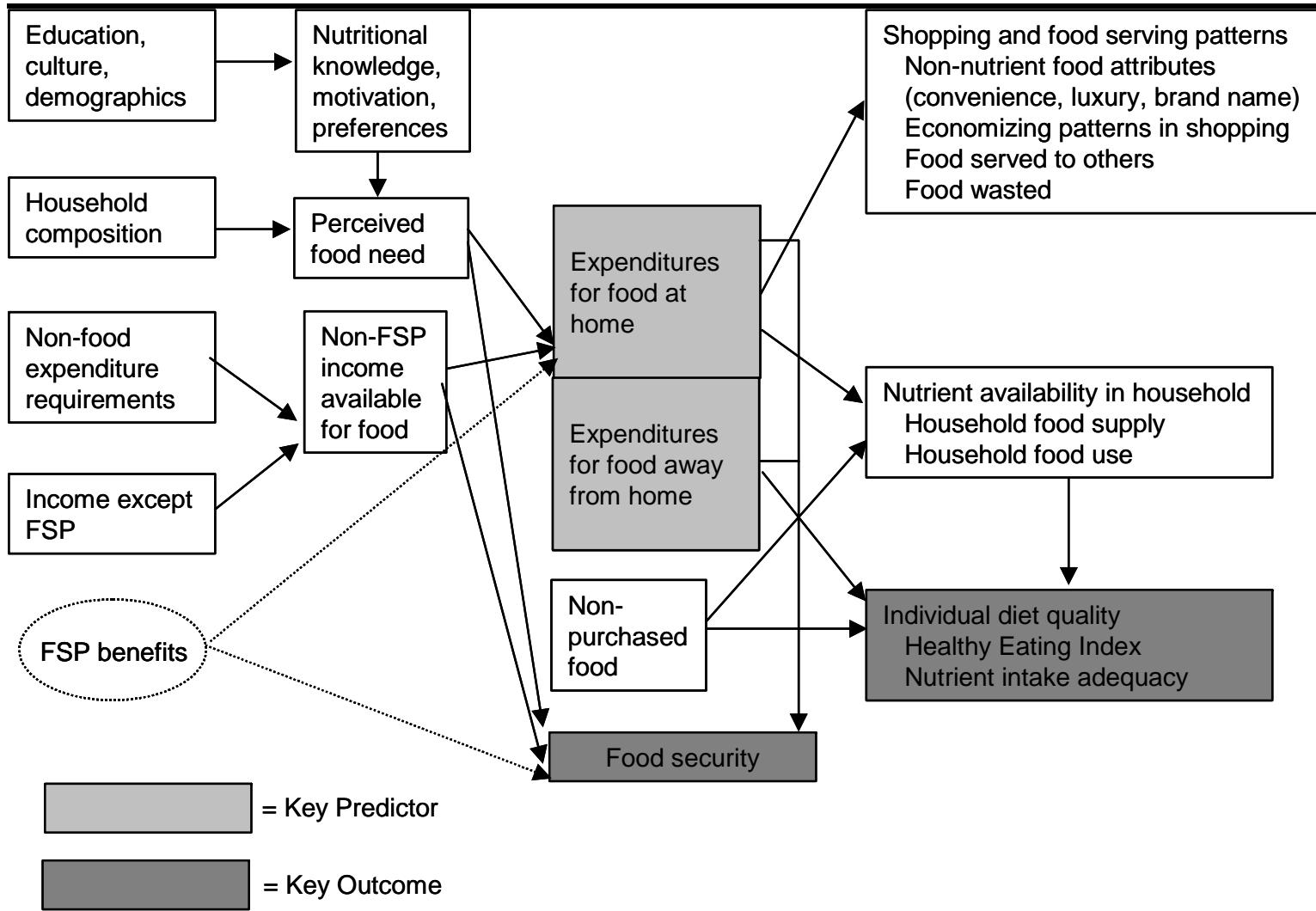
Rationale

As noted previously, existing literature provides quite convincing evidence that participation in the FSP leads to increased household food expenditures, and that the increases in food expenditures associated with FSP benefits is greater than the increase associated with an equal dollar value of ordinary income (Burstein et al., 2004a). The existing literature is much less conclusive regarding outcomes that might be expected to follow from the FSP-induced increase in food expenditures. The lack of conclusive results could stem from either of two sources. One is limitations of the research, such as selection bias, insufficient sample sizes, or imprecise measurement of outcomes. The other possibility is that increased food expenditures by low-income households do not necessarily lead to improvements in diet quality or food security.

Total food expenditures are divided between expenditures for food at home and away from home (Exhibit 12.1). Buying more food away from home could increase total food expenditures without improving diet quality. Indeed, research has suggested that eating more food away from home may have a negative effect on diet quality because away-from-home foods tend to be higher in fat and saturated fat and lower in key nutrients, including fiber, calcium, and iron, compared with at-home foods (Lin, Guthrie, and Blaylock, 1996). On the other hand, non-purchased food—such as free school meals, home-produced food, meals eaten at the home of friends or relatives, or food acquired from food pantries—may add to household members' diets without adding to expenditures.

Even when the focus is limited to expenditures for food prepared and consumed at home, many factors other than the nutrient content of the food affect the food expenditures. Households may shop at corner grocery stores or discount supermarkets; they may buy convenience foods or raw ingredients; they may choose among similar products on dimensions such as reliability and quality (e.g., brand names *versus* generics) or luxury (e.g., steak *versus* ground beef); they may maximize their use of products that happen to be discounted; or they may buy in bulk or small quantities. Once the food is in the home, some proportion may be served to persons who are not household members, and some may be wasted.

Obtaining information on the non-dietary components of food expenditures by low-income households may be important for designing the FSP impact study (Study 11), and could also provide important information for guiding food stamp policy. For example, if low-income households generally do not improve their diet quality when they increase their food expenditures, one would not expect food stamp benefits to improve diet quality unless the program provides educational or other interventions designed to influence participants' food expenditure decisions.

The food security picture is similarly complex. Food security is measured by survey items that have two components. One component specifies a behavior related to inadequacy of the household food supply or eating patterns (e.g., skipping meals). The other component links the behavior to a limitation in financial resources (e.g., "because you didn't have enough money to buy food"). Measured food insecurity therefore depends on the perceived sufficiency of the household's food purchasing power, which is affected by the household's resources and by the competing demands for expenditures, and on household eating patterns, which are affected by food expenditures and the other factors related to dietary intake.

This study will examine the relationships among household food expenditures, quality of the household food supply, individual dietary intakes, and household food security. One unique feature of the study is that it will be the first to provide detailed information in all four of these domains. A second key feature is that households will be interviewed at two points in time, so that many or most will be observed at different levels of food expenditure. This will allow the analysis to include a respondent-level random effect to account for unmeasured differences (e.g., food preferences) that may affect the relationship between food expenditures and the other outcomes. The second survey will occur 12 months after the first in order to avoid possible seasonal effects on expenditure-diet quality relationships.

This study involves a number of thorny design issues, which are discussed in the sections below. These issues do not have an obvious "right answer." The discussion therefore seeks to identify the alternative approaches, characterize their advantages and disadvantages, and indicate which approach seems most promising on the basis of the currently available information. These issues should be reviewed before proceeding with the study, as new information from the extant data analysis (Study 6), the qualitative research (Study 7), or elsewhere may lead to different choices.

Sample

The sample will be drawn from about three locations in the US. Because in-person interviews are envisioned, keeping the number of locations as small as possible will yield some economies of scale. The number and diversity of the locations needs to be sufficient to give some confidence that the findings will be applicable to the national population. The locations should be in different geographic regions, with at least one large urban area and one with small town and rural characteristics, and varying racial/ethnic groups should be included.

An important design issue is whether FSP participants should be included in the sample. The study is intended to focus on low-income households in general, and is explicitly not designed as an FSP impact study. It will not include features that would be needed in an impact study to counter selection bias, such as covariates related to FSP participation. On the other hand, to exclude FSP participants is to exclude a very large segment of the low-income population—the vast majority in the case of subgroups with high participation rates, such as single-parent households receiving TANF.

We see four possible approaches to dealing with this issue, enumerated below. We recommend the fourth approach, if it is deemed feasible:

1. Exclude those households identified as FSP participants at baseline, and do not ask about FSP participation in the followup interview. This will be a valid sample of low-income nonparticipants, but not of all low-income households.

2. Include all low-income households without ever asking about FSP participation. This provides a valid sample of all low-income households, but omits a covariate (FSP participation) that might help reduce standard errors of the estimated relationships.

3. Include all low-income households and measure FSP participation at both baseline and followup. This is a good design for this study, but it opens the possibility that subsequent

users of the public use files could produce FSP impact estimates that would be accepted as valid by people who did not understand the limitations of the study design.

4. Include all households and measure participation, but omit participation measures from the public use data files. This is our preferred design. The analysis might use a data reduction strategy to combine FSP participation with other variables, and the combined measure could be included in public use files. For example, all means-tested benefits could be combined in a single reported measure.

The definition of "low-income" for sampling purposes will use a threshold of 150 percent of the federal poverty guideline. Including some households with incomes above the FSP eligibility cutoff may help in estimating the relationship between food expenditures and other outcomes, particularly because that relationship may well be curvilinear.

An RDD survey will be used to identify households for the sample and recruit sample members into the study. Most of the interview will be conducted with the household member who is most knowledgeable about household food expenditures; typically, this will be the person who does most of the food shopping. Individual dietary intake will be measured for a randomly selected member of the household, using a 24-hour recall. As noted in Chapter Two, some subset of the individuals selected for the 24-hour recall will be asked to complete a second recall. We recommend that infants and children under 2 years of age be excluded from the individual intake sample, on the grounds that their intakes may change, in some cases dramatically, over the course of a year, for developmental reasons.[37]

We assume that the sample will be drawn as a general population sample, with no stratification. Although a variety of subpopulations are of interest to various stakeholders, obtaining separate estimates would add substantially to the resource cost. It is possible that the qualitative research in Study 7 will indicate important differences across subpopulations in the ways that expenditures translate into other outcomes. Absent that kind of information, there is no need for separate estimates for subpopulations.

Among the many relationships that the study will estimate, probably the most important is the relationship between food expenditures and individual diet quality. The sample size must have sufficient power to detect a relatively small effect on one of the central measures of diet quality. For planning purposes, we have assumed that the sample needs to be large enough to detect a 5-percentage-point change in the percentage of individuals with "poor quality" diets based on HEI scores associated with a difference in income that would have a similar impact on food expenditures as participating in the FSP. The required sample size therefore depends on the variability of income, the relationship between income and food expenditures in this population at follow-up, and the statistical power contributed by covariates other than income.

---

[37] The same caution might be appropriate for pregnant and lactating women. In contrast to the unidirectional development of infants and young children, however, women will be observed both entering and exiting the states of pregnancy and lactation between the two waves, which may provide sufficient analytic leverage to allow their inclusion in the study.

The 2001 panel of the SIPP suggests that the standard deviation of *monthly per capita income* in households that were under 130 percent of FPL 12 months previously is about $1000. We assumed that the standard deviation in *monthly per capita food expenditures* in these households was about a tenth as great, or $100, and that other covariates would explain 25 percent of the variance. The implied required final wave number of interviews is 3,339. The initial sample size, allowing for 15 percent attrition, is 3,928.

Data Elements

Study 8 will assess all the key outcomes shown in Exhibit 12.1. As summarized below, the measurement approaches described in Chapter Two will be used.

- Food expenditures and quality of the household food supply: 7-day record of household food use, supplemented with questions about foods purchased and consumed away from home.

- Individual diet quality: one 24-hour recall for all sample members and a second 24-hour recall, collected approximately a week to 10 days later and covering a different day of the week, for a subsample.

- Food security: 30-day reference period, using only the 10 adult items in the standard battery.

Some alternative measurement approaches are mentioned at the end of this chapter as study variations.

In addition to outcomes, the study must measure other variables that may bear on the relationship between food expenditures and the diet-related outcomes. Some of these factors are suggested in Figure 12.1, as discussed earlier, and others may be identified in the qualitative research.

The covariates must include a number of standard items, such as income, household composition, and socio-demographic characteristics. Two areas that will require special attention are the household's non-food expenditures and nutrition-related knowledge, attitudes, and practices. Options for assessing nutrition-related knowledge, attitudes, and practices are discussed in Chapter Two. The approach to measuring non-food expenditures will be to ask about expenditures on a small number of items rather than collect detailed expenditure records. Non-food expenditures are largely a function of income and household composition, which are measured directly. The expenditure items should therefore capture aspects of the household situation that might cause household expenditures to deviate, on either a short-term or long-term basis, from what would be predicted from household composition and income alone. Examples would be average monthly housing costs and work-related expenses, and current-month events such as health crises or vehicle or appliance breakdowns. The qualitative research may identify additional events that cause major short-term adjustments to household spending patterns.

Two of the concepts identified in Figure 12.1—perceived food need and non-food stamp income available for food—will not be measured directly. These concepts are assumed to be predicted adequately by factors that are measured directly.

Data Collection

Households recruited through an RDD screening will be asked to participate. Both baseline and followup interviews will collect information on household food expenditures, quality of the household food supply, food security, and individual dietary intake, as described under "Data Elements." The burden associated with this data collection will be substantial. Incentive payments to participating households, for both baseline and followup data collections, will be essential. For costing purposes, we have assumed a payment of $20 per wave.

Analysis Approach

The study will follow three main lines of analysis: examining, in turn, the relationship between food expenditures and (a) quality of the household food supply, (b) individual diet quality, and (c) food security.

In all three analyses, the outcome measure (e.g., household nutrient availability, individual HEI score, food security status) will be regressed on food expenditures, controlling for covariates that are logically causally prior to both expenditures and outcomes (e.g., income, demographics, nutrition-related knowledge and attitudes). Because multiple observations are available for households, and households are clustered in locations, household-level random effects and location-level fixed effects will be included in the models.

It is quite possible that this initial analysis will leave some questions unanswered, especially if it indicates a weak relationship between expenditures and the other outcomes. The question will then become: If greater food expenditures are not associated with commensurate differences in diet quality, why not?

Figure 12.1 suggests that a number of factors may affect the relationship between food expenditures and diet quality, such as the purchase of food away from home, the acquisition of non-purchased food, and a variety of food shopping and serving practices. The direction of causality among these factors is not necessarily clear (e.g., does a higher level of food expenditure allow more purchases of convenience foods, or does a need to purchase convenience foods cause higher expenditures?). While it is to be hoped that many of the links shown in Exhibit 12.1 could be demonstrated, it remains to be seen whether the data can support an elaborate analysis. One useful approach may be the technique known as seemingly unrelated regression, in which all of the relevant outcomes are simultaneously modeled as outcomes of expenditures and the covariates.[38] This sort of reduced-form estimation could provide the needed information without necessarily establishing unidirectional causation.

---

[38] The name of this technique derives from the feature that the dependent variables each appear only once, so that the equations do not appear to be related. Because all of the same variables appear on the right hand side, however, and the residuals are likely to be correlated, greater efficiency can be obtained by joint estimation.

Timeline

The overall study would be expected to take between four and five years, in three major phases. The design phase will require 8-10 months, and the field phase will take about three years. There will be two rounds of analysis and reporting, a set of cross-sectional analyses based on the first wave of data collection, and a final analysis using both waves of data. Key tasks within the phases are listed below.

1. Design phase (8–10 months)
   a. Initial meeting with FNS and follow-up memorandum
   b. Development and cognitive testing of new survey items
   c. Design, field test, and revise survey instrument and procedures
   d. Prepare OMB submission and make required revisions

2. Field phase (34–36 months)
   a. CATI/CAPI programming
   b. Train interviewers
   c. RDD survey to recruit sample
   d. Baseline survey
   e. Followup survey (12 months later)

3. Analysis and report phase (22–24 months; overlaps with field phase, starting in month 31–33)
   a. Preliminary analysis (baseline data)
   b. Preliminary report
   c. Final analysis
   d. Final report
   e. Final briefing

Estimated Cost

The cost of this study excluding the survey is between $900,000 and $1,100,000. The RDD screening is estimated to cost between $1,000,000 and $1,200,000. The first wave of the survey will cost between $2,500,000 and $3,000,000, and the second wave between $2,000,000 and $2,500,000. Hence the entire study will cost between $7,000,000 and $7,500,000.

Limitations and Risks

As noted in Chapter Two, measuring dietary outcomes as proposed here incurs risks from two sources: the mismatch of time periods for food expenditures, diet expenditures, and food security; and the need for multiple, burdensome interviews at each survey wave.

Variations

In an effort to mitigate these risks described above, numerous variations on the design were considered. Two examples are described below:

Measuring food expenditures, diet quality, and food security for a common 30-day period. This variation is intended to address the limitations associated with mismatched time periods. New or modified methods for measuring food expenditures and individual diet quality would be required, and it might be advisable to drop household-level estimates of diet quality. Food expenditures might be measured through a modified version of the current CES method, which already spans two weeks. (Note that this would assess food *purchases* rather than food *use*, but the use of a 30-day rather than 7-day time period avoids the major objection to 7-day food purchase data—variations in FSP participants' shopping practices over the month). If an acceptable 30-day measure of individual diet quality were available, it might yield greater precision and less respondent burden at lower cost.

Measuring food expenditures and diet quality for one day, and food security for 30 days. This approach would use a 24-hour recall to measure individual intake, and would ask additional questions about each food item that would allow the cost of the food to be estimated. The additional questions would include, for example, the type of store at which the food was purchased, the form in which the food was purchased (e.g., fresh or frozen), and the quantity purchased. External data sources would be used to impute food prices. This approach might yield more precise estimates of the relationship between the cost of food eaten and diet quality. However, because the cost of food eaten in a day differs substantially from the food expenditure measures used in previous food stamp research, the estimated relationships might have less bearing on the expected effects of the FSP.

Several other variations on the design were considered that might allow the study to provide more information but that would entail even greater cost and respondent burden. These included stratifying the sample by age of household member or by other demographic characteristics, sampling two persons within the household, and conducting a larger number of survey waves at shorter time intervals.

Chapter Thirteen: Study 9—Food Expenditures Experiment

Study 9 is a randomized experiment, conducted in several sites, to estimate the impacts of increased food purchasing power on actual food expenditures, the quality of the household food supply, individual diet quality, and food security.

Research Questions

- What is the impact of increased food purchasing power on
    - food expenditures
    - quality of the household food supply
    - individual diet quality
    - food security?

- How do these impacts differ in the presence of intensive nutrition education?

Objectives

This study will generate internally valid estimates of the causal relationship between food purchasing power and the outcomes of interest. If these relationships are found to be weak or nonexistent, there will be little reason to seek to measure FSP impacts on these outcomes in a national study. Knowing whether FSP impacts are markedly stronger in the presence of intensive nutrition education will provide valuable information to policy makers. To address this question, a variation of this experiment offers intensive nutrition education as well as food vouchers to a subset of the treatment group.

Rationale

An experiment that randomly increases the food purchasing power of some households and, hence, indirectly increases their food expenditures, makes it possible to assess the links between food expenditures and later links in the causal chain. It is, of course, not possible to assign households randomly to spend more on food. By giving them vouchers that can be used only for food, however, one can indirectly induce them to do so. Presumably households' incremental spending on food will be about 30 percent of the value of the vouchers (based on estimates of the marginal propensity to spend on food out of food stamps).

Logistically, the simplest way to increase the food purchasing power of FSP participants is to increase the FSP allotments. We have concluded that this approach, which notably requires cooperation from the State agency issuing benefits, is the most feasible strategy for this experiment. The approach does have drawbacks, however. Most importantly, households will already be receiving an FSP benefit that is intended to give them sufficient food purchasing power to achieve food security and a nutritious diet. Therefore, it could be argued that impacts of additional food purchasing power on these outcomes would be small (and would not provide any direct information about the effects of the

FSP). We note, however, that 19 percent of households receiving food stamps in a month are food insecure with hunger in that month (Nord *et al.* 2005), which suggests that additional food purchasing power could have an effect in that domain. Similarly, analysis of NHANES-III data showed that many FSP participants are not consuming nutritious diets: Only about 6 percent of all FSP participants had HEI scores that indicated consumption of a "good" diet and mean HEI scores for FSP participants were significantly lower than scores for both income-eligible and higher-income nonparticipants (Fox and Cole, 2004).

The size of the voucher needs to be sufficient to have an appreciable impact on the outcomes observed. We suggest $30 per household member per month as a plausible value. In addition, the intervention has to continue long enough for households to adjust their food purchasing and consumption patterns. We suggest a minimum of four months.

Sample

The sample members will be households currently receiving food stamps in three purposively selected sites. Households can receive the intervention only as long as they continue as FSP participants. This suggests sampling households at the time of recertification, and excluding households with certification periods of less than six months. We note that this solution is not perfect because some households do experience interim closures, and households in the control group would have less reason than those in the treatment group to remain on the rolls.

Sampling households as they come up for recertification is preferable to sampling newly certified households for at least two reasons. First, newly certified households will be in the process of adjusting to the basic FSP benefit, and possibly to a new TANF benefit as well. This would make it hard to identify the effect of an incremental benefit. Ongoing recipients, in contrast, can be assumed to be in some sort of equilibrium with regard to their expenditure patterns. Second, it is logistically easier to sample households from administrative lists of cases due for recertification and to recruit them for the study than to identify and recruit households that are just entering the FSP rolls.

In selecting sites, it would be valuable to consider variations not only with regard to region of the country and urban versus rural, but also with regard to cost-of-living. The FSP benefit schedule is constant in nominal terms within the continental United States, and consequently FSP benefits vary across the country in terms of purchasing power. In this sense, the FSP benefit is greater for a family living in Mississippi than for a similar family living in New York. This variation can provide analytic leverage: It would be expected that a food stamp recipient in New York would have greater unmet needs than one living in Mississippi, and that an increment to food purchasing power could therefore have a greater effect in New York.

In this two-wave study, we wish to have 80 percent power to detect a 2 percentage point effect on the proportion of treatment group members with poor HEI scores, as a result of an increment in food stamp benefits of $30 per person (about a third of the average FSP benefit). We assume that the covariates, including baseline outcome measures, will explain 25 percent of the variance. The required samples sizes are 650 final wave interviews. The initial sample size, assuming 15 percent attrition between waves, is 765 interviews per site.

Data Collection

The baseline interview will collect information on household composition, demographics, employment, unearned income sources, and the follow-up interview will collect information on changes in these domains. Both the baseline and follow-up interviews will collect outcomes data: household food expenditures, quality of the household food supply, food security, and individual dietary intake. Data collection procedures for each wave will mirror those used in the Study 4 randomized experiment and the Study 8 survey of outcomes. Each wave of interviews will require two in-person contacts and one telephone contact per households. Households that include a person sampled for the second 24-hour recall will require a second telephone contact.

The baseline outcomes information must be collected not only prior to receipt of the first voucher, but also prior to random assignment. The follow-up interviews will be conducted four months later.

Analysis Approach

The impact estimates will be based on a regression model that includes baseline outcome measures as covariates. Models would be of the form:

$$y_{j1} = \quad b_0 + b_1\, T_j + X_j\, b_2 + b_3\, y_{j0} + u_j$$

where y_{j1} is the outcome for household j in the followup period (e.g., food security, diet quality of sampled individual),

y_{j0} is the outcome for household j in the baseline period,

X_j is a vector of (other) baseline characteristics of household j, and

T_j is an indicator that household j is in the treatment group.

The impact of the intervention on the outcome is estimated as b_1.

Timeline

This study runs between three and four years. The design phase, falling entirely in the first year, includes site recruitment and selection, development of instruments and procedures, and the OMB submission. The field phase, beginning at the end of Year 1 and running through Year 2, includes an RDD survey to recruit the sample, the baseline survey and random assignment, and the follow-up survey. The analysis and reporting phase begins when the baseline survey is complete, and runs into Year 4. The tasks for the study are:

1. Design phase (8–10 months)
    a. Initial meeting with FNS and follow-up memorandum
    b. Site recruitment and selection
    c. Development of sampling and analysis plans
    d. Design, revise, and field test survey instruments and procedures
    e. Prepare OMB submission and make required revisions

2. Field phase (17–19 months)
   a. CATI/CAPI programming
   b. Train interviewers
   c. Obtain administrative lists and draw samples
   d. Sample recruitment and baseline interviews
   e. Random assignment
   f. Implementation of intervention
   g. Followup survey

3. Analysis and reporting phase (22–24 months; overlaps with field phase, starting in month 17–19)
   a. Analysis
   b. Final report
   c. Final briefing
   d. Public use data set

Estimated Cost

The estimated cost of this study excluding the survey is between $1,000,000 and $1,400,000. A little over $100,000 of this is the cost of the intervention, increased FSP benefits for the treatment group members. Survey costs are estimated as $1,500,000 to $1,700,000 for the two waves. Total study costs are between $2,700,000 and $3,100,000.

Limitations and Risks

An important risk of this study is differential sample attrition among the controls. The burden of data collection on sample members is high, and control group members will have less reason to cooperate than those who are receiving the intervention. Furthermore, paying cash incentives at the time of data collection runs the risk of altering sample members' food purchasing behavior.

One way to address this risk would be to offer extra incentives to control group members that would be paid when the demonstration is complete. Another possibility would be to provide controls with an increment to their food stamps as well, but less than that received by treatment group members. The fact that random assignment will not occur until after collection of baseline data provides some protection against the risk of differential attrition. A substantial amount of information will be available on attriters; these data can be used to support analytic adjustments (e.g., reweighting).

Another concern is seasonality. Dietary patterns may differ substantially between the baseline and follow-up periods, and it is possible that the effects of additional food purchasing power could also vary across these periods. One solution would be to spread out sample intake over a 12-month period, so that every month would appear as both a baseline and a follow-up month.

Ethical acceptability may also be a problem. Some may argue that the study is not ethical because low-income households may suffer a material hardship after they lose the additional food purchasing power they have had for some months.

Still another risk is that this study comprises a single experiment. Conducting independent experiments in several sites would provide substantially more information, at substantially greater cost.

This experiment determines impacts of food purchasing power above and beyond that supplied by the FSP. If no impact is found, it might be because the FSP benefit itself had done all that was necessary. On the other hand, if FSP participants in the control group are found to have poor levels of dietary quality and food security, then the failure of the intervention to have an effect would be indirect evidence against the FSP having had any.

Variations

Expanding the experiment to include a second treatment group—one that would receive both increased food stamp benefits and intensive nutrition education—could provide FNS with information on a crucial policy question. Suppose that both this study and the randomized experiment in Study 4 fail to find impacts on diet quality. Must it be concluded that the FSP is unable to improve participants' diets? Comparing outcomes for two different treatment groups would show whether households' use of food purchasing power can be improved through nutrition education.

We note that adding a second treatment group would increase sample size requirements by 50 percent, and would have a nearly proportional effect on the survey costs. It would also be important that the nutrition education intervention represent something that could realistically be implemented in the FSP environment.

Another variation for consideration is to conduct the experiment on eligible nonparticipants. An advantage of this approach is that, unlike FSP participants, these households would have insufficient food purchasing power by program criteria. Hence the effect of an increment in their food purchasing power could have an effect more like that of the FSP. Countering this are two disadvantages. First, sample recruitment and retention would be more difficult than for FSP participants. An RDD survey would be required to identify sample members, substantially increasing the cost of the study. Second, some way would need to be devised to give treatment group members vouchers that could be used only for food. Creating a financial instrument that would be acceptable to all food retailers and could not be used for any other purpose would present a serious challenge.

Chapter Fourteen: Study 10—National Study of FSP Experiences, Satisfaction, and Outcomes

Study 10 is a national study of experiences, satisfaction, and outcomes of FSP participants and FSP eligible nonparticipants. It will be implemented in lieu of Study 11 if it is determined from the preliminary studies that program impact estimates should not be attempted.

Research Questions

- How do FSP participants compare with eligible nonparticipants in terms of household food expenditures, diet quality, food security, and household well-being?

- What are FSP-eligible households' experiences and satisfaction with the FSP, WIC, and school meals programs?

Objectives

The objectives of this study are to provide a description of the FSP recipient population, in terms of their diet quality and their experiences and satisfaction with the FSP and other nutrition assistance programs; and to similarly describe the eligible nonparticipant population.

While Study 10 will resemble the 1996 NFSPS in its general scope and approach, there will be some important differences (Exhibit 14.1). First, the samples of participants and nonparticipants will be drawn from the same frame. Second, the objectives of Study 10 are broader than its predecessor in that they include measurement of outcomes for both participants and nonparticipants. These data will allow for comparisons without inferences of causality, e.g., do participants have higher or lower levels of food security than nonparticipants? A third difference between this study and the NFSPS is that information will be collected on an expanded set of outcomes, including food expenditures (data on food expenditures were collected in the NFSPS, but the data were never fully analyzed), individual dietary intake, household well-being, and nutrition knowledge, attitudes, and practices.

Rationale

In the event that the results of the preliminary studies do not support the decision to conduct a national impact study, FNS will still want a descriptive study of the experiences and outcomes of FSP participants and eligible nonparticipants. Study 10 will provide national estimates of diet quality, food security, and other measures of household well-being for both of these groups. This information will provide policy makers with information about how the diets of FSP participants compare with those of income-eligible nonparticipants, and how the diets of both groups compare with public health recommendations. The data may also be useful in identifying subgroups of nonparticipants where the need for outreach is the most acute.

Exhibit 14.1

Comparison of the 1996 National Food Stamp Program Study and Study 10

Characteristic	1996 NFSPS	Study 10
Sampling	Participants from FSP lists and RDD Nonparticipants from RDD	Both participants and nonparticipants from RDD

Outcomes measured	**Participants[a]**
	  Household nutrient availability
	  Food security
	Nonparticipants
	  Food security
		Participants and nonparticipants
		  Food expenditures
		  Quality of the household food supply
		  Individual dietary intake
		  Food security
		  Household well-being
		  Nutrition knowledge, attitudes, and practices

Measurement of participant experiences and satisfaction	**Participants only**
	  Experiences and satisfaction with FSP
		Participants and nonparticipants
		  Participation in FSP and other FNS programs
		  Experiences and satisfaction with FSP, WIC, NSLP, and SPB

[a] Data on food expenditures were collected but not fully analyzed.

Sample

A nationally representative sample of FSP-eligible households will be recruited from an RDD sampling scheme. A geographically clustered sampling design is anticipated to allow for in-person interviewing.

The statistics that will be reported from Study 10 will be univariate estimates of population means or percentages, such as the proportions of FSP participant and eligible nonparticipant households, respectively, that are food insecure.

We wish to do a comparison of means between participants and eligible nonparticipants to detect a difference of 5 percentage points in the proportion with poor HEI scores. This is simply a descriptive comparison of two populations, so no covariates are used. We assume 100 Primary Sampling Units, and a design effect associated with cluster sampling of 1.8. The required sample size is 3,020 completed interviews.

Data Collection

Key variables that will be created in Study 10 include an indicator for current program participation, variables describing household composition and demographics, the outcomes listed in Exhibit 14.1, and measures of participant experiences and satisfaction. We also suggest, given the expanding focus

on nutrition education in the FSP, that this study assess nutrition-related knowledge, attitudes, and practices. Essential household descriptors will include household composition (age, sex, relationships of household members), demographics of household head or primary food preparer (race/ethnicity, education), household resources (sources and amounts of income, assets), other factors affecting program eligibility (such as school attendance, and pregnancy or postpartum status), region, and urbanicity. Survey items that focus on experiences will be asked of both FSP participants and nonparticipants and will cover their history and experiences with FNS programs (FSP, WIC, NSLP, and SBP). Survey items that address satisfaction will cover three major dimensions for each FNS program: accessibility, costs of participation, and service.

Accurate identification of current FSP participants using survey techniques is essential for the success of this study. Although the samples of participants and eligible nonparticipants will be drawn based on a telephone screening, both participation status and eligibility will be determined more exactly during initial in-person interviews. If Study 10 were the entire research agenda, we would recommend that it include a component of cognitive testing to develop better items to measure FSP participation. We assume, however, that that task will have been accomplished as part of the participation study.

Data collection procedures will mirror those described for each wave of data collection in Studies 4, 8, and 9. Two in-person contacts and one telephone contact will be required per household. Households that include a person sampled for the second 24-hour recall will require a second telephone contact. The incentives previously proposed will also be important for this study to promote cooperation and adequate response rates.

Analysis Approach

Analyses will mostly involve calculating univariate statistics describing population means and proportions. Appropriate statistical software should be used to calculate standard errors of means and proportions, taking sampling weights and the sample design into account. Separate estimates will be calculated for FSP participants and nonparticipants. Similar to the reports from the NFSPS, estimates will also be produced for specific subgroups of interest, such as:

- Households with elderly
- Households with children
- Households at various income levels
- Households of various sizes
- Households with various benefits eligibility levels

Timeline

This study is estimated to take about four years. Design activities, including site selection and refining the instruments and analysis plan, occur in the first year. The survey is fielded in the second year, and the data analysis and reporting starting in the second year. Tasks are listed below.

1. Design phase (8–10 months)
   a. Initial meeting with FNS and follow-up memorandum

b. Development of sampling and analysis plans
c. Design, revise, and field test survey instruments and procedures
d. Prepare OMB submission and make required revisions

2. Field phase (18–20 months)
    a. CATI/CAPI programming
    b. Train interviewers
    c. RDD survey to recruit sample
    d. Conduct survey

3. Analysis and reporting phase (26–30 months; overlaps with field phase, starting in month 19–21)
    a. Analysis
    b. Final report
    c. Final briefing
    d. Public use data set

Estimated Cost

The estimated cost of this study excluding survey costs is between $1,000,000 and $1,300,000. Costs of the RDD screener are estimated as $900,000 to $950,000, and the survey itself between $2,400,000 and $2,800,000. The entire study is estimated to cost between $4,500,000 and $5,000,000.

Limitations and Risks

The major limitation of Study 10 is that it is not an impact study but will be presenting data for both participants and nonparticipants. There is a risk that the descriptive findings regarding outcomes of FSP participants and nonparticipants could be misused as indicators of impact. Because study reports will include side-by-side presentations of data for participants and nonparticipants, we can not get around this issue as easily as we might for Study 8, which also collects information on dietary outcomes for both participants and nonparticipants. Study reports will need to include strong caveats about the study limitations and appropriate interpretation of the data.

Variations

We see three broad sets of variations of Study 10, pertaining to drawing the sample, measuring outcomes, and incorporating information on local office practices.

Sampling Variations

As described above, Study 10 uses an RDD sampling approach for both participants and eligible nonparticipants. NFSPS, in contrast, relied primarily on administrative lists for sampling participants and used an RDD to sample nonparticipants. The primary advantage of RDD over list-based sampling for participants is consistency when comparing their outcomes and experiences with those of eligible nonparticipants. If RDD is used for both groups, two sources of total survey error— sampling frame non-coverage and unit non-response—would be similar for the two groups.

While this argument would be decisive in a study that included estimates of program impacts (i.e., in Study 11), one could argue that, for a descriptive study of characteristics and experiences, the completeness and accuracy of a list-based sample of participants may be more valuable than comparability between the participant and nonparticipant samples. Therefore a variation on the design for Study 10 is to use a list-based sample for FSP participants, or to augment the RDD participant sample with a list-based sample. The challenges of obtaining timely lists from local offices, including contact information, should not, however, be underestimated.

Even if RDD is the chosen sampling method for participants, several variations on the sampling scheme should be considered. The potential challenges of obtaining participant and eligible nonparticipant samples via an RDD screening survey include the following:

- Although high-quality RDD surveys use a wide range of survey methods to obtain high response rates—advance letters, refusal aversion, monetary incentives, multiple call attempts, etc.—response rates in RDD surveys have declined over time (Curtin *et al.*, 2005).

- Survey responses about program participation are not completely accurate. We have recommended work on this issue in the context of Study 3.

- A sizeable segment of program participants and eligible nonparticipants fall outside the RDD sample frame.

For a national study, the third issue is especially salient. Some low-income households have no land-line telephone—only a cell phone, or telephone access through a neighbor or some other arrangement (Blumberg *et al.*, 2004). Cell phone numbers are typically excluded from RDD studies because the respondents have to pay for the incoming call. Although cash reimbursement for the call might solve that problem, safety and ethical issues would remain—e.g., that the respondent may be driving in traffic, or may be a child or teenager for whom an offer of reimbursement for the call may be inappropriate or ambiguous.

An idea of how widespread this problem is currently can be gleaned from the 2003 National Health Interview Survey (Blumberg *et al.*, 2004). Overall, 2.2 percent of families in the U.S. did not have any type of telephone service at the time of the survey and 4.2 percent had cell phone service only. The situation is worse, however, among the low-income population. Among families that reported receiving food stamps, 9.1 percent had no phone service and 5.9 percent had only a cell phone. Thus an estimated 15 percent of FSP participants could not be reached by a single-frame RDD approach.

Work is currently being done on how to include cell phone-only households in RDD samples, based on dual-frame sampling methods, and Abt staff attended a conference on that subject in February. We learned that some survey organizations have undertaken dual-frame designs, in which telephone numbers are drawn from land-line telephone exchanges, cellular telephone exchanges, and "mixed-use" telephone exchanges. For cellular numbers an offer of $10 "to pay for the call" has been used. The interview must determine how many land-line telephone numbers and cellular telephone numbers a household has in order to develop dual-frame weights for use in estimation. We plan to track

developments in this field, as it will probably be desirable to use a dual-frame design for this study. The cell phone-only population will likely continue to grow in size.

While use of dual-frame sampling addresses concerns about cell phone-only households, it does not solve the problem of households that have neither land-line or cellular phone service. These households are among the neediest of the populations eligible for nutrition assistance programs, and that their program experiences and satisfaction may differ from those of other participants in significant ways. While the cost of designing the entire study around this subgroup would be prohibitive, an RDD approach might be supplemented to address this potential gap. For some number of geographic areas or sites, *an area probability sample* of households without land-line or cellular telephone service could be selected (at a very high cost). A sample would be drawn in particular geographic areas by doing door-to-door interviews—randomly selecting blocks, buildings, and apartments to locate these households. A comparison of the results from this on-the-ground survey with a parallel RDD survey in the same areas would reveal how much information is lost by relying on RDD alone. For the nonhousehold population (e.g., homeless persons), specialized sampling techniques are required. The estimates based on the RDD sample or the dual-frame RDD/cellular design could be compared with the estimates from the single-frame RDD sample and the area probability samples combined, with appropriate weighting.

A final consideration in drawing the samples is whether some population subgroups should be oversampled. If estimates are desired for particular subpopulations and if those estimates need to meet specific criteria for precision (i.e., confidence intervals that are no wider than a specific criterion), then the overall sample size for the study may need to be increased. Many subgroups have been suggested as being of special interest, including:

- Types of FSP households: e.g., elderly persons only, households with children and with young children;
- Types of individuals defined by age and sex: e.g., young children, teen-aged girls, women of childbearing years, elderly;
- Currently participating/not participating in other food assistance programs;
- Households with earnings;
- Individuals by race/ethnicity;
- Very poor households;
- Urban/suburban/rural;
- Region of the country;
- Recent immigrants;
- Individuals with physical health conditions such as obesity or diabetes;
- Individuals with mental health conditions, drug abuse; and
- Homeless individuals.

The consensus of the attendees at the TWG meeting was that overall program estimates are the highest priority. It appears that there is no identifiable subgroup for which some advocacy group would not want separate estimates. The most sensible way to proceed seems to be to design the research so as to support separate estimates for *unspecified* subgroups of a given size, e.g., 20 percent of the population. If, at the time the study is fielded, it is deemed likely that smaller subgroups will be of interest, either the sample size can be increased, or the smaller subgroups can be specifically

oversampled. While special provisions can be made to reach subgroups that are difficult or impossible to interview over the telephone (e.g., those with limited English proficiency, those with mental health conditions, and the homeless), the costs of doing so need to be taken into account in the context of overall project goals and resources.

Variations Related to Outcomes

An option for reducing cost is to collect outcomes data only for FSP participants, as was done in the NFSPS. Arguably, outcomes data on nonparticipants could be dropped without major consequence, given that two other national surveys (CES and NHANES) routinely collect information on food expenditures, food security, and individual dietary intake. On the other hand, CES and NHANES do not collect information on household food use; FSP participation and eligibility may not be measured sufficiently accurately for FNS's purposes; and neither the timing of these studies nor access to data is in the control of FNS.

Outcomes data on eligible nonparticipants are potentially valuable to FNS even absent an attempt to estimate impacts. Information on the prevalence of poor diet quality and food insecurity can show where the need for the program is greatest, and suggest likely subpopulations for outreach. If, on the other hand, the great majority of eligible nonparticipants appear to have acceptable diets and are food secure, then FNS can conclude that current outreach programs are sufficient to serve low-income households in nutritional need. Given the arguments on both sides of this issue, we conclude that the decision of whether to include nonparticipant outcomes data in Study 10 should be held in abeyance.

Another variation on the current design would be to include height and weight among the outcomes. These data could be used to describe the prevalence of overweight and obesity among FSP participants and nonparticipants. If there is interest in this outcome, we would recommend that height and weight be measured, at least for children, as discussed in Chapter Two. If necessary, self-reports of height and weight for adolescents and adults are acceptable.

Variation Related to Local Office Practices

A final potential variation is to collect information on office practices in the selected sites. Data on program access (e.g., office hours, requirements for in-person certifications, verification requirements) could provide valuable insights when linked to households' reports of program experiences and satisfaction.

Chapter Fifteen: Study 11—National Study of FSP Experiences and Impacts

Study 11 will use data from a multiwave national survey of FSP participants and nonparticipants to estimate FSP impacts on expenditures, diet quality, food security, and household well-being. This study will also measure experiences and satisfaction with the FSP and other FNS programs. Study 11 will be conducted only if results from the preliminary studies provide convincing evidence that (a) the FSP could have meaningful impacts on these outcomes, and (b) Study 11 would be able to produce internally consistent estimates of these impacts.

Research Questions

- What is the impact of the FSP on household food expenditures, diet quality, food security, and household well-being?
- What are FSP-eligible households' experiences and satisfaction with the FSP, WIC, and school meals programs?

Objectives

The objectives of Study 11 are to obtain nationally representative and internally valid estimates of FSP impacts on dietary quality and food security, and to obtain nationally representative data on households' experiences and satisfaction with FNS programs.

Rationale

FNS would like information on the impacts of the FSP but acknowledges that it may not be feasible to obtain it. Study 11 will be implemented only if both of the following results have been obtained from the preliminary studies:

1. There is evidence that *higher food expenditures lead to better nutritional outcomes for low-income families*. This evidence could come from Studies 6 through 9. From this we would infer that the FSP could potentially have meaningful effects on outcomes beyond food expenditures.

2. There is evidence that *the impact of the FSP can be measured accurately using nonexperimental methods*. This evidence could come from Studies 1 through 5.

The strongest nonexperimental approach for measuring impacts is to compare outcomes between participants and nonparticipants in a "repeated measures" design. This approach reflects recent thinking on quasi-experimental designs, which holds that a robust design must include multiple measures of outcomes, including multiple measures before participation begins and multiple measures during the period of participation. This adds stability to the measures and reduces the potential bias related to regression to the mean. The validity of this approach will have been tested previously in Studies 4 and 5.

Sample

While Study 11 is an alternative to Study 10, it retains most of the characteristics shown in Exhibit 14.1 and discussed for Study 10. Instead of a single round of interviews, however, Study 11 includes four waves of data collection on outcomes. Program experiences are measured for FSP participants and eligible nonparticipants in a single wave, and compared descriptively. Impacts of FSP participation are measured by comparing outcomes between participants and nonparticipants, gaining analytic leverage from households that are observed both on and off the program.

The multi-wave design for Study 11 starts with baseline data collection for an RDD sample of low-income households (Wave 1), followed by subsequent waves of data collection at regular intervals. We assume a total of four waves, occurring at four-month intervals. Data collection for each wave will follow the pattern outlined in for the three preceding studies, all of which collect survey data on dietary outcomes.

A key requirement of this design is that the Wave 1 (baseline) sample include a sufficient number of nonparticipant households who enter the FSP during the course of the study, as well as a sufficient number of FSP participant households who eventually exit the program. We estimate that only 7 to 8 percent of eligible nonparticipants in Wave 1 would enter the FSP by Wave 4.[39] Conversely, we estimate that about a quarter of participants in Wave 1 will exit the program by Wave 4 while remaining eligible.[40] Oversampling households based on characteristics that are associated with changes in participation status could help ensure a sufficient number of entries and exits. However, doing this raises the risk that the sample may not be representative. In such a case, widely varying sample weights would be needed to yield valid national estimates.

As noted in the discussion of Study 10, accurate identification of participants and nonparticipants is crucial. Survey techniques are known to result in mis-classifications. We expect that work will have been done on this issue (as part of Study 3) prior to the launch of Study 11.

This four-wave study repeats Study 5, in 100 Primary Sampling Units. Our criterion for precision is the ability to detect a 5 percentage point effect on the proportion with "poor quality" diets. We assume a design effect of 1.77, and wave-to-wave attrition rates of 15 percent. The required sample size is 1,791 final wave interviews. The initial sample size is 2,917 completed interviews.

Data Collection

Data collection for Study 11 includes features described in previous chapters for Studies 4, 5, 8, 9, and 10. In particular,

---

[39] These estimates are based on rough calculations from the 2001 SIPP panel. Of FSP income-eligible nonparticipants in the last month of Wave 1, only 3.5 percent received food stamps in the last month of Wave 2, and 7.4 percent did so in the last month of Wave 2, 3, and/or 4.

[40] These estimates are similarly based on the 2001 SIPP panel. Of FSP participants in the last month of Wave 1, 9.5 percent were income-eligible nonparticipants in the last month of Wave 2, and 24 .6 percent were income-eligible nonparticipants in the last month of Wave 2, 3, or 4.

- Collection of outcomes data will follow the model used in Studies 4, 8, 9, and 10, including:
  - household food use data and nutrition knowledge, attitudes, and practices of the food manager;
  - individual intake data for the randomly sampled household member;
  - food security and household well-being from the primary respondent.

- Explanatory variables will be collected as in Study 5, including information on household composition, income sources, eligibility program participation, and selection factors from the primary respondent.

- Experiences and satisfaction data will be collected in the first wave as in Study 11, from the primary respondent.

Analysis Approach

The analysis approach for Study 11 will be heavily dependent on the results and lessons learned from the nonexperimental replications (Study 5) and other preliminary studies. It is expected that observations will be grouped by household and the impact estimation will be carried out using hierarchical linear modeling or a conceptually equivalent household-level random effects model. Study 5 may also indicate the need to conduct propensity score matching that may potentially result in the omission of households from one group (e.g., participants) that do not have a good match from the other group (eligible nonparticipants). These would be households with very high or very low probabilities of participation. The analysis may also include specification tests. These issues were discussed in Chapter Two.

Timeline

This study is estimated to last four to five years. Design activities, including site selection and refining the instruments and analysis plan, occur in the first year. The first two waves of the survey are fielded in the second year, and the third and fourth waves in the third year. Data analysis and reporting starting in the second year (to describe the Wave 1 data on FSP experiences), and continue through part of Year 5. Tasks are listed below.

1. Design phase (8–10 months)
   a. Initial meeting with FNS and follow-up memorandum
   b. Development of sampling and analysis plans
   c. Design, revise, and field test survey instruments and procedures
   d. Prepare OMB submission and make required revisions

2. Field phase (25–27 months)
   a. CATI/CAPI programming
   b. Train interviewers
   c. RDD survey to recruit sample
   d. Wave 1 interviews
   e. Wave 2 interviews

f. Wave 3 interviews

g. Wave 4 interviews

3. Analysis and reporting phase (30–32 months, overlaps with field phase, starting in month 24–26)

   a. Preliminary analysis following Wave 1

   b. Interim report (program experiences and satisfaction, sample characteristics)

   c. Analysis following Wave 4

   d. Final report (impacts)

   e. Final briefing

   f. Public use data set

Estimated Cost

Exclusive of the surveys, this study is estimated to cost between $1,300,000 and $1,700,000. The RDD screening is estimated to cost between $900,000 and $1,000,000. The first survey wave costs between $2,200,000 and $2,600,000. The other three waves, which collect less extensive data, decrease in cost because of sample attrition. The final survey wave would cost between $1,500,000 and $1,700,000. The total study cost is estimated as $9,000,000 to $11,000,000.

Limitations and Risks

We assume that Study 11 would not be undertaken unless FNS were convinced that it could obtain valid impact estimates. Even if the preliminary studies provide good support for nonexperimental impact estimation, however, some members of the research community might still view the impact estimates with skepticism.

An RDD-based survey may omit a substantial portion of the eligible population. This issue and possible approaches to reducing this threat to generalizability were discussed in the preceding chapter (Study 10) and will not be repeated here.

Variations

The base specification for Study 11 assumes that survey questions will be used to determine whether respondents participate in the FSP, and, if so, when they began participating. An alternative is to ask survey respondents for information that can be matched with administrative data to determine participation more precisely. This greater precision entails additional costs to conduct the match and to deal with households' reluctance to divulge this information.

Results of Studies 1 through 5 are likely to result in new insights that may result in additional options for Study 11. For example, the preliminary studies may indicate the need for additional key variables relevant to the participation decision that have not been anticipated. Or Study 5 could suggest that more or fewer waves of data collection are necessary. Another possibility is that preliminary studies could suggest larger FSP impacts for particular subgroups, which may suggest oversampling of particular subgroups for Study 11.

Other variations mentioned for Study 10 are also relevant here. In particular, area probability samples of households without telephone service could be drawn to supplement the RDD screening samples or height and weight could be added to the list of outcomes.

Chapter Sixteen: Issues in Evaluating Impacts of Other FNS Programs

The preceding chapters described a research agenda that will provide FNS with the information it needs to determine whether it is advisable to field a national study to assess diet-related impacts of the FSP. FNS is also interested in determining the impacts of other nutrition assistance programs (NAPs). Several of the observations made in designing a research agenda for the FSP also apply to other FNS programs:

1. Selection bias is a serious concern for all FNS programs. Determinants of participation among eligibles are only partially understood, and unmeasured factors may also affect outcomes.

2. Virtually no randomized studies have been conducted to assess impacts of any FNS programs. Much could be learned from small-scale experimental studies, both about program impacts and about researchers' ability to correct for selection bias.

3. The hypothesized relationships between proximal outcomes (on which the programs have direct leverage) and distal outcomes (which reflect the ultimate goals of the programs) need to be specified and then demonstrated.[41]

4. Only when researchers have established that (a) selection bias can be overcome in estimating effects, and (b) the distal outcomes of interest can be affected by the proximal outcomes directly affected by the programs, is it worthwhile to conduct national studies of program impacts.

As noted in Chapter One, this project was originally intended to result in a research plan for assessing impacts of several FNS programs in a single study, with highest priority given to the larger ones. At the TWG meeting, however, a consensus emerged that no one study could do justice to obtaining impact estimates for multiple programs with varying eligibility criteria and modes of intervention.

Explication of a full research agenda for any of these other NAPs is beyond the scope of the present contract and this report. As a compromise, this chapter provides an overview of general issues that need to be considered in designing *impact* evaluations of WIC, SBP, and NSLP, and national *descriptive* studies of CACFP, SFSP, CSFP, and FDPIR. The characteristics of these programs are summarized in Exhibit 16.1 (which includes FSP for completeness).

---

[41] For the FSP, for example, food expenditures is the proximal outcome and diet quality and food security are distal outcomes. For the NSLP and SBP the proximal outcome could be the nutritional quality of meals offered to students and distal outcomes could extend from the nutritional quality of meals actual consumed and the quality of students' overall diets to household food security.

Exhibit 16.1

Characteristics of Selected FNS Programs

Program	Target population	FY 2002 participation	Income-eligibility requirement (percent of Federal poverty guideline)	Benefits provided	Availability	Participant lists
Food Stamp Program	Low-income households	19,099,524 participants in 8,194,090 households per month	≤ 130%[a]	Electronic benefits for use in purchasing food for home consumption Nutrition education may be offered	Universal	Household/participant lists available at State level
WIC Program	Low-income pregnant, breastfeeding, and postpartum women; infants; children ages 1-4	7,490,841 participants per month	≤ 185%[b]	Supplemental foods, nutrition education, and referrals to health care and social services	Universal	Participant lists available at State level
Child Nutrition Programs						
National School Lunch Program	School-age children	28,006,873 lunches (children) per day	≤ 130% receive free meals/snacks 131-185% receive reduced-price meals/snacks > 185% may participate but pay full price for meals	Lunches that meet specific nutrition standards[c] Afterschool snacks	Universal	District lists available at State level Student lists available at district or school level
School Breakfast Program	School-age children	8,144,384 breakfasts (children) per day	≤ 130% receive free meals/snacks 131-185% receive reduced-price meals/snacks > 185% may participate but pay full price for meals	Breakfasts that meet specific nutrition standards[c]	Approximately 70% of all NSLP schools	District lists available at State level Student lists available at district or school level

Exhibit 16.1

Characteristics of Selected FNS Programs

Program	Target population	FY 2002 participation	Income-eligibility requirement (percent of Federal poverty guideline)	Benefits provided	Availability	Participant lists
Child and Adult Care Food Program	Children and adults attending licensed, non-residential day care facilities, homeless shelters, and afterschool programs[d]	2,852,110 participants per day 41% meals and snacks served in child care homes, 57% in child care centers, and 3% in adult day care centers. Some participants receive more than one meal/snack per day.	Any child or adult in participating center or home may participate. Reimbursements to providers are based on relative poverty status of populations they serve[e]	Meals and snacks that meet defined meal patterns	Universal	Sponsor lists available at State level Participant lists available at provider level
Summer Food Service Program	Low-income school-age children	Approximately 35,000 feeding sites 1,921,440 children per day	Any child attending an approved feeding site may participate[f]	Free meals and snacks that meet defined meal patterns	Selected feeding sites	Sponsor lists available at State level Feeding site lists available at sponsor level[g] Participant lists available at *enrolled feeding sites*. Participant lists not maintained at open feeding sites[h]
Food Distribution Programs						
Commodity Supplemental Food Program	Low-income pregnant and postpartum women, infants, children up to their 6th birthday, and adults 60 and older	427,444 participants per month 18% women, infants, and children and 82% elderly	≤ 185% for women, infants, and children ≤ 130% for adults 60 and older	Commodity foods, nutrition education, referrals to health care and social services	May change annually In 2003, 32 States, District of Columbia, and 2 Indian Tribal Organizations	Participant lists available at State/ITO level

Exhibit 16.1

Characteristics of Selected FNS Programs

Program	Target population	FY 2002 participation	Income-eligibility requirement (percent of Federal poverty guideline)	Benefits provided	Availability	Participant lists
Food Distribution Program on Indian Reservations	Low-income American Indian households or non-Indian households living on reservations[i]	110,122 participants per month	≤ 130%	Commodity foods (alternate to the FSP)	In 2003, 98 Indian Tribal Organizations and 5 State Agencies Includes 22 States geographically, and 247 tribes.	Household/participant lists available at the ITO/State agency level

Notes:

a   Must also meet certain resource, work-related, and categorical requirements.

b   Must also be certified, by a recognized health care professional, to have a nutritional risk. Participation is not guaranteed. Local programs can serve only as many participants as their funding will allow. Priority system is used to fill slots when funding is tight.

c   Participating schools receive cash subsidies for each meal served (and donated commodities for each lunch served), including those served to students who pay full price. Reimbursement rates are higher for meals served to students free or at a reduced price than for meals served at full price.

d   Nonprofit child care centers are eligible to participate in the CACFP, as are for-profit centers in which at least 25 percent of the center's enrollment or licensed capacity receive either Title XX funds or are eligible for free or reduced-price meals.

e   Providers receive cash subsidies for every meal and snack served. Centers are reimbursed based on the financial need of the children and adults they serve, using the income-eligibility and meal reimbursement rates used in the NSLP and SBP. Homes are reimbursed based on the economic need of providers and the children they serve. Homes located in low-income areas or operated by providers with incomes < 185% of poverty are reimbursed at higher rates than other homes.

f   Most feeding sites are located in areas where at least 50 percent of the children are from households with incomes ≤ 185% of poverty (open sites) or in programs where 50 percent of the enrolled children are eligible to receive free or reduced-price meals, using the income-eligibility criteria defined for the NSLP and SBP. Residential summer camps may receive reimbursement for meals and snacks served to enrolled children whose documented household income makes them eligible for free or reduced-price meals.

g   Program operates for a very short period of time over summer months, and lists may not be finalized until shortly before operations begin. For sampling purposes, may need to work with prior year's list (Briefel, 2004)

h   Source: Briefel, 2004.

i   Low-income households that contain at least one member of a federally recognized tribe and reside in approved areas near reservations or in Oklahoma may also participate.

Key points to consider for each program are:

- What is the eligible population?
  - How can participants be identified?
  - How do individuals select into the program? What are the implications for measuring impacts nonexperimentally?

- What are the chief program outcomes? How can they be measured?

- How are the proximal and distal outcomes related to each other?

Issues in Evaluating the WIC Program

WIC is the single most studied NAP (Fox 2004). In addition, studies of the WIC program have addressed many more nutrition- and health-related outcomes than studies of others NAPs. Despite this vast body of research, definitive evidence on the impact of the program is lacking. As with all NAPs, WIC evaluations have generally been quasi-experimental in nature.[42] Selection bias is a particularly thorny issue for assessing impacts of the WIC program.

Identifying the Eligible Population

WIC eligibility has three main elements: life stage (pregnant women, breastfeeding and other postpartum women, infants, and children up to their fifth birthday); economic need (income up to 185 percent of FPL or participation in a means-tested program); and nutritional risk (anthropometric, biological, clinical, or dietary). State agencies maintain complete administrative lists of WIC participants. Alternatively, participants can reasonably be identified through an RDD household screening. Caution must be exercised, however, in determining exactly who is participating. Abt's recent analysis of WIC participation data in NHANES 1999-2002 suggests that some women responded affirmatively to the question about WIC participation when, in fact, the participant may have been their infant or child.

Identification of fully eligible nonparticipants requires information on life stage, income, and nutritional risk. Determination of nutritional risk requires clinical assessment and/or measurements such as weight, height, and hematocrit. Assessing the nutritional risk of otherwise-eligible nonparticipants would be costly and difficult and is probably not feasible. If nonparticipants are defined on the basis of life stage and income alone, which could be done through an RDD screening[43], the participants will, by definition, be at greater nutritional risk than nonparticipants. As discussed below, this poses a non-trivial challenge in assessing program impacts.

---

[42]   The only exception to this rule is a randomized experiment that was conducted early in program's history (in one site) when the demand for participation exceeded available funding (Metcoff *et al.*, 1985).

[43]   We note, however, that a fraction of pregnant women will not be aware they are pregnant.

Selection into the Program

The nutritional risks that confer eligibility are explicit negative predictors of the outcomes the WIC program is hypothesized to affect. This makes disentangling the causes and effects of participation highly problematic. Positive selection bias also occurs in that individuals are more likely to participate if they are willing to eat WIC foods. The program is not as attractive, for example, for pregnant women who do not want to drink milk or for parents of children who demand highly sweetened breakfast cereal.

An important first step in thinking about assessing impacts of the WIC program is to understand facilitators and barriers to participation and, in particular, the timing of participation: why some eligible pregnant women delay program entry, and why the families of some eligible infants and children leave the program. Extant data studies are probably of limited use in this area because available data sets are unlikely to provide sufficient information on pregnancy and breastfeeding status, medical history, and current nutritional risks. In addition, little leverage on selection bias can be gained from multiple observations of individuals in participant and nonparticipant status, because (except for children) the period of eligibility is so short, and the developmental processes that occur during this period are so dramatic.

It would seem, then, that qualitative research like that envisioned in participation Study 2 and linked experiments and nonexperimental replications (Studies 4 and 5) would be key parts of a WIC-focused research agenda. Developing experiments that are ethically and legally permissible, and also useful, will be a challenge. Although WIC is not an entitlement program, it is sufficiently well funded to serve virtually all eligible persons who apply. An experiment in assisted participation for pregnant women conducted 15 years ago did not succeed in raising the participation rate at all (Puma *et al.*1991)[44]. Furthermore, it is reasonable to assume that near-eligible individuals, such as 5-year-old children, nonbreastfeeding women 7 to 12 months postpartum, breastfeeding women 13 to 24 months postpartum, or persons in all categories between 185 and 250 percent of FPL, would experience smaller program effects than participants. Hence, randomly assigning any of these groups to receive WIC benefits, while ethically unobjectionable, is unlikely to yield good estimates of program impacts. Perhaps a better understanding of barriers to participation, as might be gained from qualitative research, would suggest experiments that might be more successful than the previous assisted participation study in influencing participation.

Program Outcomes

WIC provides three benefits to participants: supplemental foods, nutrition education, and referrals to health and social services. Each of these benefits may influence different outcomes. Moreover, the outcomes that WIC aims to improve are qualitatively different for the various participation groups. For all groups, these outcomes include individual dietary intake and could also include food security. Additional potential outcomes include:

---

[44]   The WIC application rates during pregnancy were 60 percent for treatment group members, 61 percent for control group members (Puma *et al.*, 1991, p. V-41).

- For pregnant women: infants' birth weight and neonatal well-being, health care costs, initiation of breastfeeding

- For breastfeeding women: breastfeeding duration, weight status, iron status, general health status

- For postpartum women: weight status, iron status, general health status

- For infants and children: infant feeding patterns, weight status, linear growth, iron status, utilization of preventive health and dental care services, health and dental care costs, immunization status, physical, emotional and cognitive development

In designing a research agenda for the WIC program, careful thought must be given to the causal chains that lead to hypothesized outcomes. Because WIC prescribes specific foods for individuals to consume, it is reasonable to assume that the likelihood of impact on individual diet quality is greater than it is for the FSP. However, because the program generally provides vouchers or checks for the foods to be purchased,[45] rather than providing the foods directly to participants, slippage between the intervention and intake by the targeted individual can occur if foods are shared with other family members. In addition, WIC vouchers can be expected to substitute to some extent for equivalent foods that would have been purchased absent program participation. For example, provision of a full monthly supply of infant formula might increase an infant's consumption of formula (rather than cow's milk or other age-inappropriate foods) by some amount, but (depending on age) the infant might still consume some formula absent the program.

Thus, in understanding WIC impacts on diet quality, one may ask first:

- How much of the *WIC foods* do participants consume? How much would they consume absent the program?

- If WIC participants consume less than the full WIC package, is it because
  - WIC foods are shared with other household members?
  - some of the WIC foods are not acquired, or are acquired and thrown away?

Secondary effects not directly attributable to nutrients provided by WIC foods may be expected on participants' diets, for two reasons: because the WIC package displaces other foods, and because nutrition education is intended to influence the overall diet. This suggests the following questions:

- How much of other *foods* do WIC participants consume, relative to what they would consume absent the program?

- What is the net effect on diet quality?

---

[45] In a limited number of geographic areas, foods are delivered to participants' homes or participants pick up foods at warehouses.

Issues in Evaluating NSLP/SBP

Assessment of the school meals programs presents several unique challenges. Chief among these are the ambiguous nature of participation and the logistics of evaluating a school-based program.

Identifying the Eligible Population

All children attending schools in which the NSLP and SBP are offered are eligible to participate in the programs. Children from households who meet income-eligibility criteria are eligible to receive meals at a reduced price (household income below 185 percent of the FPL) or for free (income below 130 percent of the FPL). In evaluating the impact of these programs, FNS may have special interest in those who are or could be certified for free or reduced-price meals.

Distinguishing between participants and nonparticipants in the school meals programs presents special challenges because participation status may vary from one day to the next. Children may or may not obtain food/meals at school on a given day and, among those who do, the items obtained may or may not constitute a reimbursable NSLP or SBP meal. In general, this is a greater issue for lunch than for breakfast, and for older children than younger children. Older children have the option to refuse components of a reimbursable meal (younger children may or may not have this option) and also tend to be offered more nonreimbursable options.

Students who report receiving meals free or at a reduced price in a survey can reasonably be identified as program participants in a general sense (e.g., "usually"). It is more difficult, however, to identify higher-income participants and nonparticipants. Home-based survey questions must obtain information on whether meals are obtained from school sources (not always a cafeteria) and must attempt to distinguish consumers of reimbursable meals (participants) from consumers of non-reimbursable meals and non-meals (nonparticipants).[46] Alternatively, data on foods that are chosen may be recorded by trained observers stationed in the school cafeteria, generated by an automated data capture at the point of sale, or (for older children) recorded daily on a questionnaire.

A central decision in conducting a study of the NSLP and SBP is whether to draw the study sample based on an RDD screening of the population to find families with school-aged children, or to draw a sample of schools and then sample children within schools. In the existing body of research, all studies that have included primary data collection have used school-based sampling. There are at least two distinct advantages to this approach. The first is that data collection can occur on site, providing an opportunity to observe foods selected by sampled children or, at least, to obtain information about reimbursable and non-reimbursable offerings that can be used, in subsequent analysis, to reliably identify participants and nonparticipants. The second advantage is that

---

[46] For analyses of dietary intake that use 24-hour recall data, it is preferable to identify participants and non-participants based on the characteristics of the meals consumed (e.g., the meal included three of five components required in reimbursable NSLP meals). This has been done in many studies in the past (Devaney *et al.*, 1993 and Gleason and Suitor, 2000 and 2001). However, the increasing flexibility available to schools in defining what constitutes a reimbursable meal makes implementation of this approach more challenging. Ideally, one should have access to information about the menu planning system used in each school and the availability of a la carte (non-reimbursable) options.

---

information about characteristics of the school food service program can be gathered and included in the analysis. A disadvantage to school-based sampling is that it requires school cooperation and the release of students' contact information. Concerns about student privacy and burden raise challenges in gaining school cooperation.

Selection into the Program

For the NSLP/SBP as well as for FSP and WIC, understanding participation patterns seems essential to understanding impacts. Selection bias is complicated by the ability of children to opt out on a daily basis. In many situations, one can obtain leverage on selection bias by observing an individual both as a participant and as a nonparticipant. This requires, however, that the variation in participation be caused by exogenous trigger events that are not correlated with the outcomes. In the school meals programs, however, variations in participation are largely caused by events that directly affect the outcomes, such as what is on the menu. For example, one day a child chooses fries and a Coke and another day takes a reimbursable meal. Although dietary intake is of higher quality on the latter day, the difference is not "caused" by the NSLP, which was equally available on both days. Instead, the child is, for whatever reasons—including the available selections—willing to eat better on the second day, and takes the reimbursable meal.

The NSLP and SBP are both entitlement programs, so there is no realistic possibility of a randomized experiment. However, a recent FNS study made use of randomization in a different way that could prove useful for future research. In the School Breakfast Program Pilot Project, an evaluation of the impacts of a universal-free breakfast program, volunteer elementary schools were randomly assigned to treatment (universal-free breakfast) and control (normal SBP) groups. In the treatment schools, all children, regardless of household income, were given the opportunity to obtain school breakfasts for free (Bernstein *et al.,* 2004). The evaluation assessed the effects of a universal-free school breakfast program (USBP) rather than the SBP per se. The main threat to validity in generalizing to the SBP is that the impact estimates were based solely on outcomes for those who would take the meals only if they were made free for everyone. Thus, they represent the effects of increasing the scope of the program, which might be less than the impacts on current participants. The effects were not, however, restricted entirely to paid-eligible students. Program changes such as serving meals in classrooms instead of the cafeteria increased participation among lower-income students as well. Participation increased four-fold among paid-eligible children, while doubling for other categories.[47]

---

[47] Overall, the demonstration succeeded in raising SBP participation in the intervention schools by 15 to 17 percentage points. It seemed that the school breakfast essentially substituted for what the children would otherwise have eaten, at home or elsewhere. In particular,

- Children in USBP schools, though significantly more likely to have a substantive breakfast, experienced food and nutrient intakes at breakfast and during the day that did not differ systematically on average from those in the control schools.

- Children in USBP schools were more likely to have two substantive breakfasts, but no less likely to skip breakfast.

- Disciplinary incidents were more common in the USBP schools in the first year, but not in subsequent years. Visits to the school nurse were less common in the second of the three years.

- No impacts were found on academic achievement in reading or math, or in tardiness or absenteeism.

Despite this arguable lack of generalizability, it is doubtful that there is a better design for an ethically permissible experiment to test the impacts of the SBP. A similar approach might therefore be worth considering for estimating impacts of NSLP: offering free reimburseable lunches to all children in randomly selected schools (while *à la carte* selections would still require payment). The issue of generalizability could be addressed in part by calculating effects separately for free, reduced-price, and paid-eligible participants.

Program Outcomes

Proximal outcomes of NSLP/SBP participation include whether a student has a substantive breakfast or lunch, nutrient and food intake at breakfast or lunch, and 24-hour nutrient and food intake. Potential distal outcomes include household food security, household food expenditures, school attendance and tardiness, attentiveness in class, disciplinary infractions, cognitive achievement, general health status, linear growth, and weight status.

Program impacts are not straightforward to define because of the daily nature of NSLP/SBP participation. Immediate impacts, e.g., on 24-hour dietary intake, can be related to selection of a reimbursable meal on that day; but longer-run impacts, e.g., on cognitive achievement, would need to be related to more flexible measures of participation using thoughtfully chosen decision rules.

Issues in Conducting National Descriptive Studies of CACFP, SFSP, CSFP, and FDPIR

Each of these programs pose difficulties in building probability samples of participants and eligible nonparticipants for describing program experiences. If such samples could be assembled, one could proceed to national studies that could describe characteristics and relevant outcomes of participants and eligible nonparticipants. Given the dearth of this type of information, this is logically the place to begin.[48]

Eligible Populations

A central issue in designing studies for such programs is whether individuals can reliably report whether or not they (or their children) are participating in the program. If they can, then a household screening sample based on random digit dialing (RDD) is at least feasible as a sampling approach. If not, some alternative way must be found to build probability samples of participants and eligible nonparticipants.

---

[48]   A national descriptive study of the CACFP was published in 1997 (Glantz *et al.*, 1997 and Fox *et al.*, 1997). The study was limited to participating centers and homes (and thus participating children) and the only outcome assessed was nutrient intake while in care. Likewise, the most recent national descriptive study of the SFSP, published in 2003, did not include nonparticipants and did not assess impacts on overall nutrient intake (Gordon and Briefel, 2003). The only national study of the FDPIR (Usher et al., 1990) included neither nonparticipants nor outcomes data. There has never been a national study of the CSFP.

With the caveat that program names may vary locally, respondents can be expected to have direct knowledge of whether they are receiving commodities through *FDPIR* or *CSFP*. Household screening surveys could therefore be used to identify program participants. Identification of eligible nonparticipants requires consideration not only of limited geographic availability, but also of the relationships with other programs. Because these programs are alternatives to FSP and (for some individuals) to WIC, respectively, the eligible nonparticipant populations might be defined to exclude participants in those programs. Alternatively, policy makers might be interested in comparing outcomes between FDPIR and FSP participants, and between (WIC categorically eligible) CSFP and WIC participants. Elderly CSFP participants (now comprising the bulk of that program) and five-year-old child participants are, of course, separate populations.

CACFP and *SFSP* require different approaches for building samples. Parents are unlikely to know if their child care providers are reimbursed by CACFP, or if the meals their children receive at summer programs are provided by SFSP. Furthermore, there are no administrative lists of participants for these programs.

The previous national study of the CACFP drew a sample of participant children through a multi-stage process: sampled states provided lists of sponsors, sampled sponsors provided lists of centers and family child care homes, and the centers and child care homes provided lists of children and families (Glantz *et al.*, 1997). A study that included nonparticipants could reverse this process. Children in nonparental care would be identified through an RDD screening, and contact information would be obtained for their care providers. The providers could then report whether they participated in the CACFP.

Alternatively, and preferably in our view, a list-based approach could be used to generate samples of participating and nonparticipating providers. State licensing lists cover licensed homes (and also include providers that are no longer in business, typically 30 to 40 percent of the total). Many homes that do not participate in the CACFP are, however, license-exempt (or operating illegally), and would thus not be on the state lists. Child care centers may also be license-exempt, for example if operating out of a church. As such, they would also not be on state lists of licensed providers. A question for policy makers is whether eligible children should be defined as those in care with *licensed* child care providers, or those in *any* nonparental care.

Child Care Resource and Referral services (CC R&Rs) may be able to identify license-exempt providers if they are receiving subsidies. The lists of licensed (and perhaps license-exempt) providers could be cross-checked against lists of CACFP providers obtained from those sponsors serving the specific geographic area.

A final consideration is that the three types of programs included in CACFP should be analyzed separately because of their markedly different hours of service and meal patterns: family child care homes, child care centers, and Head Start Centers.[49]

---

[49]   Nonresidential day care facilities for elderly or functionally impaired adults are also included in the program. Because adults comprise less than 3 percent of CACFP participants, however, we assume that they would be excluded from the next national study as from the previous one.

Identifying SFSP participants is even more difficult. First, as in NSLP/SBP, appropriate meals are offered that children may or may not take in full. Information is needed on the foods that children select in order to determine participation.[50] Second, not only are parents unaware of SFSP participation, but SFSP sites themselves may not maintain participant lists. *Enrolled feeding sites* maintain participant lists, but *open feeding sites*, which constitute more than 80 percent of all feeding sites (Gordon and Briefel, 2003), offer meals to all children in attendance.

Drawing comparable samples of SFSP participants and eligible nonparticipants would require an RDD screening to identify children who participate in summer programs that are eligible to offer SFSP, and collecting sufficient identifying information on these programs to determine if they do participate in the SFSP. A question for policy makers is whether children deemed eligible should be limited to those who attend structured programs. The difficulty of determining participation as defined by taking a full SFSP meal could potentially be resolved by reframing the question in terms of the *availability* of SFSP meals.

We suggest two potentially promising approaches to assessing the SFSP that do not require starting with a list of SFSP sites. The first is obtain lists of children approved for free and reduced price meals from schools areas of high *versus* low SFSP coverage, and then following up a random sample of these children over the summer. Such a study would allow for a comparison of outcomes between children who do and do not have access to a summer feeding program, among those who have typically been getting one or two meals a day at school during the school year.

A second approach is to capitalize on the fact that SFSP is not available everywhere. Schools in low-coverage areas could be recruited into an experiment in which they would be randomly assigned to an intervention which would comprise making SFSP available over the summer to their students who are eligible for free or reduced-price meals during the school year.

It must be acknowledged, however, that any approach based on drawing comparable samples of SFSP participants and eligible nonparticipants would miss two important potential contributions of the program. The SFSP motivates providers to give meals, and it also encourages and facilitates children's participation in the concurrent programs.

Outcomes

For the *CACFP* and the *SFSP*, key outcomes include food and nutrient intake while "in care" and over 24 hours. For the *CSFP*, individual dietary intake is a key outcome. For the WIC-eligible populations, distal outcomes similar to those noted above for the WIC program, may also be of interest. The *FDPIR*, like the FSP, is household-based. Because it provides foods directly, rather than food purchasing power, household food use rather than food expenditures is the most proximate outcome. Household food security and individual dietary intake are additional outcomes of interest.

---

[50]   This is less of an issue in the SFSP than in the school meals programs, because there is typically much less choice: many sites serve children complete lunches that are pre-packaged in a bag or a box. The most recent study of the program recommended, however, that this approach to meal delivery be changed to reduce plate waste and increase acceptability of meals (Gordon and Briefel, 2003).

References

Allison, P.D. (1990). Change scores as dependent variables in regression analysis. *Sociological Methodology*, *20*, 93–114.

Ashenfelter, O. (1978) "Estimating the Effects of Training Programs on Earnings." *Review of Economics and Statistics, 60* (1) (February 1978), 47-57.

Ashenfelter, O. and Card, D. (1985) "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs." *Review of Economics and Statistics, 67*(4): 648-660.

Bartlett, S., & Burstein, N. (2004). *Food Stamp Program Access Study: Eligible nonparticipants.* Report prepared for the U.S. Department of Agriculture, Food and Nutrition Service. Cambridge, MA: Abt Associates Inc.

Bartlett, S., Burstein, N., & Hamilton, W. (2004). *Food Stamp Program Access Study: Final report*. Report prepared for the U.S. Department of Agriculture, Economic Research Service. Cambridge, MA: Abt Associates Inc.

Basiotis, P.P., Guthrie, J.F., Bowman, S.A., Welsh, S.O. (1995). Construction and evaluation of a diet status index. *Family Economics and Nutrition Review, 8*(2), 2-13.

Bernstein L. S., McLaughlin , J. E., Crepinsek, M. K., & Daft, L. M. (2004). *Evaluation of the School Breakfast Program Pilot Project: Final report* (Nutrition Assistance Program Report Series, No. CN-04-SBP). Report prepared for the U.S. Department of Agriculture, Food and Nutrition Service, Office of Analysis, Nutrition, and Evaluation. Cambridge, MA: Abt Associates Inc.

Blaylock, J.R., & Smallwood, D. M. (1984). Reasons for nonparticipation in the Food Stamp Program. *Western Journal of Agricultural Economics*, *9*(1), 117-126.

Blisard, N., Stewart, H., & Jolliffe, D. (2004). Low-income households' expenditures on fruits and vegetables (Agricultural Economic Report No. 833). Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Blumburg, S. J., Luke, J. V., & Cynamon, M. L. (2004). *Has cord-cutting cut into random-digit-dialed health surveys? The prevalence and impact of wireless substitution.* Paper presented at the Eighth Conference on Health Survey Research Methods, February, Peachtree City, GA.

Brandt, L., Feino Cunha, J., Weinreich Olsen, A., Chilima, B., Hirsch, P., Appelberg, R., & Andersen, P (2002), Species of environmental mycobacteria block multiplication of BCG and induction of protective immunity to tuberculosis. *Infection and Immunity*, 70, 672-678.

Brener, N. D., McManus, T., Galuska, D. A., Lowry, R. and Wechsler, H. (2003). "Reliability and validity of self-reported height and weight among high school adolescents." *Journal of Adolescent Health, 32*(4): 281-7.

Briefel, R. (2004). Personal communication.

Burstein, N. R. (1993). *Dynamics of Food Stamp Program as reported in the Survey of Income and Program Participation*. Cambridge, MA: Abt Associates Inc.

Burstein, N., Fox, M. K., Hiller, J. B., Kornfeld, R., Lam, K., Price, C., & Rodda, D. T. (2000). *Profile of WIC children*. Cambridge, MA: Abt Associates Inc., March.

Burstein, N., Price, C., Rossi P. H., & Fox M. K. (2004a). Food stamp program. In M. K. Fox, W. Hamilton, & B-H. Lin (eds.), *Effects of food assistance and nutrition programs on nutrition and health: Vol. 3, Literature review* (Food Assistance and Nutrition Research Report No. 19-3). Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Burstein, N., Hamilton, W., Fox, M. K., & Price, C. (2004b). *Assessing the food security and diet quality impacts of FNS program participation: Initial menu of survey options (draft)*. Cambridge, MA: Abt Associates Inc.

Burstein, N., Hamilton, W., Fox, M. K., Price, C. & Battaglia, M. (2005). *Assessing the food security and diet quality impacts of FNS program participation: Final menu of survey options*. Cambridge, MA: Abt Associates Inc.

Burstein, N., & Visher, M. (1989). *The Dynamics of Food Stamp Program Participation.* Cambridge, MA: Abt Associates Inc.

Coalition for Evidence-Based Policy. (2002). *Bringing evidence-driven progress to education: A recommended strategy for the U.S. Department of Education*. Washington, DC, Author.

Cody, S., Gleason, P., Schechter, B., Satake, M., and Sykes, J. (2005). *Food Stamp Program entry and exit: An analysis of participation trends in the 1990s* (Contractor and Cooperator Report No. (CCR8). Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Cody, S. (2004). *Elderly Nutrition Demonstrations: Interim report on elderly participation patterns* (E-FAN-04-009). Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Coe, R. D. (1983) Participation in the Food Stamp Program, 1979. In G. J. Duncan & J. N. Morgan (Eds.), *Five thousand American families—Patterns of economic progress*, *Vol. 10*. Ann Arbor, MI: University of Michigan, Institute for Social Research.

Cohen, B., Ohls, J., Andrews, M., Ponza, M., Moreno, L., Zambrowski, A., & Cohen, R. (1999). *Food stamp participants' food security and nutrient availability: Final report.* Princeton, NJ: Mathematica Policy Research.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Crepinsek, M. K., and Burstein, N. (2004). *Maternal Employment and Children's Nutrition: Final Report, Volume II: Other Nutrition-Related Outcomes*. Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Cunnyngham, K. (2005). *Food Stamp Program participation rates: 2003*. Report prepared for U.S. Department of Agriculture, Food and Nutrition Service. Washington, DC: Mathematica Policy Research.

Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly, 69* (Spring), 87–98.

Dehejia, R. (2004). Practical propensity score matching: A reply to Smith and Todd. *Journal of Econometrics*, *125*(1–2), 355–364.

Dehejia, R.H., & Wahba, S. (2002). Propensity score matching methods for non-experimental causal studies. *Review of Economics and Statistics, 84*(1), 151–161.

Devaney, B., Gordon, A. and Burghardt, J. (1993). *The School Nutrition Dietary Assessment Study: Dietary Intakes of Program Participants and Nonparticipants*. Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service.

Devaney, B., Seigler, P., Pac, S., Karwe, V. and Barr, S. I. (2004). "Nutrient Intakes of Infants and Toddlers." *Journal of the American Dietetic Association*, *104*(1), Supplement 1, January.

Dunifon, R.E. and Kawaleski-Jones, L. (2004). *Exploring the Influence of the National School Lunch Program on Children*. Discussion Paper 1277-04. Institute for Research on Poverty, University of Wisconsin-Madison.

Farrell, M., Fishman, M., and Langley, M. (2003). *The Relationship of Earnings and Income to Food Stamp Participation: A Longitudinal Analysis*. Washington, DC: USDA, ERS. Report No. E-FAN-03-011.

Fox, M. K (2004). The WIC Program. In M. K. Fox, W. Hamilton, & B.-H. Lin (Eds.), *Effects of food assistance and nutrition programs on nutrition and health: Volume 3, Literature review*. (Food Assistance and Nutrition Research Report No. 19-3). Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Fox, M. K. & Cole, N. (2004). *Nutrition and Health Characteristics of Low-Income Populations: Volume 1, Food Stamp Program Participants and Nonparticipants*. Washington, DC: U.S. Department of Agriculture, Economic Research Service, E-FAN report No. 04-014-1.

Fox, M. K., Glantz, F. B., & Geitz, L. (1997). *Early childhood and child care study: Nutritional assessment of the CACFP: Volume II, Final report*. Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service.

Fox, M. K., Hamilton, W., & Lin, B. H. (Eds.) (2004). *Effects of food assistance and nutrition programs on nutrition and health: Vol. 3, Literature review* (Food Assistance and Nutrition Research Report No. 19-3). Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Fraker, T. (1990). *The effects of food stamps on food consumption.* Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service.

Fraker, T., & Maynard, R. (1987). The adequacy of comparison group designs for evaluations of employment-related programs. *Journal of Human Resources, 22*(2): 194–227.

Frankel, M. R., Srinath ,K. P., Hoaglin, D. C., Battaglia, M. P., Smith, P. J., Wright, R. A., & Khare, M. (2003). Adjustments for non-telephone bias in random-digit-dialing surveys. *Statistics in Medicine, 22*, 1611–1626.

Gibson, D. (2003). "Food Stamp Program Participation is Positively Related to Obesity in Low Income Women." *Journal of Nutrition*, 133: 2225-2231.

Gibson, D. (2004). "Long-term food stamp program participation is differentially related to overweight in young girls and boys." *Journal of the American Society for Nutritional Sciences, 134*: 372-379.

Glantz, F.B., Rodda, D., Cutler, M. J., Rhodes, W., & Wrobel, M. 1997. *Early Childhood and Child Care Study: Profile of participants in the CACFP: Volume I, Final report.* Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service.

Glazerman, S., Levy, D. M., & Myers, D. (2003). Nonexperimental *versus* experimental estimates of earnings impacts. *Annals of the American Academy of Political and Social Science, 589*(1), 63–93.

Gleason, P., Rangarajan, A. and Olson, C. (2000). *Dietary Intake and Dietary Attitudes Among Food Stamp Participants and Other Low-Income Individuals.* Alexandria, VA: USDA, FNS.

Gleason, P., & Suitor, C. (2001). *Changes in children's diets: 1989-1991 to 1994-1996 (*CN-01-CD2). Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service, Office of Analysis, Nutrition and Evaluation.

Gleason, P., & Suitor, C. (2000). *Children's diets in the mid-1990s: Dietary intake and its relationship with school meal participation*. Report submitted to the U.S. Department of Agriculture, Food and Nutrition Service. Princeton, NJ: Mathematica Policy Research, Inc.

Gleason, P., Schochet, P. & Moffitt, R. (1998). *The Dynamics of Food Stamp Program participation in the early 1990s*. Report submitted to U.S. Department of Agriculture, Food and Nutrition Service. Washington, DC: Mathematica Policy Research, Inc.

Gordon, A., & Briefel, R. (2003). *Feeding low-income children when school is out: The Summer Food Service Program, Executive summary* (ERS-FANRR-30). Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Haines, P., Siega-Riz, A. and Popkin, B. (1999). "The Diet Quality Index Revised: A Measurement Instrument for Populations." *Journal of the American Dietetic Association, 99*: 697-704.

Hamilton, W. L., & Rossi, P.H. (2004). *Effects of food assistance and nutrition programs on nutrition and health: Vol. 1, Research design* (Food Assistance and Nutrition Research Report No. 19-1)*.* Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Hanley, J.A., & McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, *148*, 839–843.

Harding, D. J. (2003). Counterfactual Models of neighborhood effects: The effect of neighborhood poverty on dropping out and teenage pregnancy." *American Journal of Sociology, 109*, 676–719.

Heckman, J. J. & Smith, J. A. (1995). Assessing the "Case for Social Experiments. *Journal of Economic Perspectives*, *9*(2), 85–110.

Heckman, J., & Hotz, V. J. (1989). Alternative methods for evaluating the impact of training programs. *Journal of the American Statistical Association*, *84*(408), 862–880.

Hisnanick, J.J., & Walker, K.G. (2000). *Food stamp receipt: Those who left versus those who stayed in a time of welfare reform* (Working Paper 236). U.S. Department of Commerce, Census Bureau. Retrieved from http://www.sipp.census.gov/sipp/workpapr/workpapr.html.

Huffman, S., and Jensen, H. (2003). Do food assistance programs improve household food security? Recent evidence from the United States (Working Paper 03-WP 335). Ames, Iowa: Center for Agricultural and Rural Development, Iowa State University.

Human Resources Development Canada [HRDC]. (1998). *Quasi-experimental evaluation and data development*. Retrieved from http://www11.hrdc-drhc.gc.ca/pls/edd/QEE.lhtml.

Institute of Medicine, Food and Nutrition Board [IOM] (1999). *Dietary reference intakes: Calcium, phosphorus, magnesium, vitamin D, and fluoride.* Washington, DC: National Academy Press.

Institute of Medicine, Food and Nutrition Board [IOM] (2000a). *Dietary reference intakes: Thiamin, riboflavin, niacin, vitamin B6, folate, vitamin B12, pantothenic acid, biotin, and choline.* Washington, DC: National Academy Press.

Institute of Medicine, Food and Nutrition Board [IOM] (2000b). *Dietary reference intakes: Vitamin C, vitamin E, selenium, and carotenoids.* Washington, DC: National Academy Press.

Institute of Medicine, Food and Nutrition Board [IOM] (2001). *Dietary reference intakes: Application in dietary assessment.* Washington, DC: National Academy Press.

Institute of Medicine, Food and Nutrition Board [IOM] (2002a). *Dietary reference intakes: Vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium, and zinc.* Washington, DC: National Academy Press.

Institute of Medicine, Food and Nutrition Board [IOM] (2002b). *Dietary reference intakes: Energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein, and amino acids (macronutrients).* Washington, DC: National Academy Press.

Institute of Medicine, Food and Nutrition Board [IOM] (2004a). *Dietary reference intakes for water, potassium, sodium, chloride, and sulfate.* Washington, DC: National Academy Press.

Institute of Medicine, Committee to Review WIC Food Packages [IOM] (2004b). *Proposed criteria for selecting the WIC food packages.* Washington, DC: The National Academy Press.

Jastrzab, J., Giordono, L., Chase, A., Valente, J., Hazlett, A., LaRock, Jr., R., & James, D. (2004) *Serving country and community: A longitudinal study of service in AmeriCorps—Early findings.* Cambridge, MA: Abt Associates Inc.

Jones, S. J., Jahns, L., Laraia, B. A., & Haughton, B. (2003). Lower risk of overweight in school-aged food insecure girls who participate in food assistance. *Archives of Pediatric and Adolescent Medicine, 157*(8), 780-784.

Kant, A.K., Schatzkin, A., Graubard, B., Shairer, C. (2000). A prospective study of diet quality and mortality in women. *Journal of the American Medical Association, 283*(16), 2109-15.

Kennedy, E. T., Ohs, J., Carlson, K. S. and Fleming, K (1995). "The Healthy Eating Index: Design and Applications." *Cereal Food World, 42*: 74-78.

LaLonde, R. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review, 76*, 604-620.

Lin, B.-H., Guthrie, J., & Blaylock, J. (1996). *The diets of America's children: Influences of dining out, household characteristics, and nutrition knowledge (*Report No. AER746). Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Linz, P., Lee, M., & Bell, L. (2005). *Obesity, poverty, and participation in nutrition assistance programs.* Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service.

Leibtag, E.S., & Kaufman, P. R. (2003). Exploring food purchase behavior of low-income households: How do they economize?" *Agriculture Information Bulletin* No. 747-07. Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Logan, C., Fox, M. K. and Lin, B-H (2002). *Effects of Food and Nutrition Assistance Programs on Nutrition and Health: Volume 2, Data Sources.* Washington, DC: USDA, ERS.

McConnell, S. & Nixon, L. (1996). *Reaching the working poor and the poor elderly study: Report on literature review and data analysis*. Washington, DC: Mathematica Policy Research, Inc.

McCullough, M., Feskanich, D., Stampfer, M., Giovannucci, E., Rimm, E. B, Hu, F. B., Spiegelman, D., Hunter, D. J., Colditz, G. A. and Willett, W.C. (2002). "Diet quality and major chronic disease risk in men and women: Moving toward improved dietary guidance." *American Journal of Clinical Nutrition 76*(6): 1261-71.

McKernan, S-M., & Ratcliffe, C. (2003). *Employment factors influencing Food Stamp Program participation: Final report* (E-FAN-03-012). Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Metcoff, J., Costiloe, P., Crosby, W. M., Dutta, S., Sandstead, H. H., Milne, D., Bodwell, C. E., & Majors, S. H. (1985). Effect of food supplementation (WIC) during pregnancy on birth weight. *American Journal of Clinical Nutrition, 41,* 933-47.

Nord, M., Andrews, M., & Carlson, S., (2005). *Household food security in the United States, 2004* (Economic Research Report No. 11). Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Orr, L. L., Bell, S. H., and Kornfeld, R. (2003). *Tests of nonexperimental methods for evaluating the impact of the New Deal for Disabled People (NDDP)*. Bethesda, MD: Abt Associates Inc. and Urban Institute, July (for the Center for Research in Social Policy, Loughborough University, under contract with the U.K. Department for Work and Pensions).

Ouellette, T., Burstein, N., Long, D. and Beecroft, E. (2004). *Measures of Material Hardship*, Abt Associates Inc., Bethesda, MD, 2004. Report prepared for DHHS/ASPE.

Paccaud, F., Wietlisbach, V. and Rickenbach, M. (2001). "Body mass index: Comparing mean values and prevalence rates from telephone and examination surveys." *Review of Epidemiology Santa Publique 49*(1): 33-40.

Ponza, M., Ohls, J., Moreno, L., Zambrowski, A., and Cohen, R. (1999). *Customer service in the Food Stamp Program*. Washington, DC: Mathematica Policy Research, Inc.

Puma, M. J., Pietro, J., Rosenthal, J. Connell, D., Judkins, D., & Fox, M. K. (1991). *Study of the Impact of WIC on the growth and development of children: Field test, Final report: Vol. I, Feasibility assessment*. Bethesda, MD: Abt Associates Inc.

Sherwood, N. E., Jeffrey, R. W., French, S. A., Hannan, P. J., and Murray, D. M. (2000). Predictors of weight gain in the Pound of Prevention study. *International Journal of Obesity, 24,* 395-403.

Singer, E., Hoewyk, J. V., & Maher, M.P. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly, 64,* 171–188.

Smith, J. (2000). "A Critical Survey of Empirical Methods for Evaluating Active Labor Market Policies." Downloaded from http://www.glue.umd.edu/~jsmithz/papers/jeff_papers.htm.

Smith, J., & Todd, P. (2004). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics, 125*(1-2), 305–353.

Smith, J., and Todd, P. (2003). Rejoinder. http://www.glue.umd.edu/~jsmithz/papers/jeff_papers.htm.

Spencer, E., Appleby, P., Davey, G. and Key, T. (2002). "Validity of self-reported height and weight in 4808 EPIC-Oxford participants." *Public Health Nutrition 5*(4): 561-5.

U.S. Department of Agriculture, Food and Nutrition Service (2000). *Food and Nutrition Service (FNS) Strategic Plan, 2000 to 2005.*

U.S. General Accounting Office. (1988). *Food Stamps: Reasons for nonparticipation.* Washington, DC: GAO, December.

Usher, C.L., Shanklin, D, & Wildfire, J. B. (1990). *Evaluation of the Food Distribution Program on Indian reservations (FDPIR:, Volume I, Final report.* Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service.

Vasan , R. S., Pencina, M. J., Cobain, M., Freiberg, M. S., and D'Agostino, R. B. (2005). Estimated risks for developing obesity in the Framingham Heart Study. *Annals of Internal Medicine, 143,* 473-80.

Villanueva, E. V. (2001). "The validity of self-reported weight in US adults: A population based cross-sectional study." *BMC Public Health 1*(1): 11.

Whitmore, D (2004). *Do School Lunches Contribute to Childhood Obesity?* Preliminary and incomplete paper presented at the annual meeting of the American Economic Association, January 2005, Philadelphia.

Wilde, P. E., McNamara, P.E., & Ranney, C.K. (2000). *The effect on dietary quality of participation in the Food Stamp and WI programs* (ERS-FANRR-9). Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Wilde, P.E. & Ranney, C.K. (2000). The monthly food stamp cycle: Shopping frequency and food intake decisions in an endogenous switching regression framework. *American Journal of Agricultural Economics, 82*(1), 200-213.

Zambrowski, A., & Ohls, J. (1999). *File documentation for the National Food Stamp Program Survey Final Report.* Princeton, NJ: Mathematica Policy Research.

Appendix A: Extant Data for Analyzing Food Stamp Program Participation and Program Outcomes

To develop an approach to Studies 1 and 6, we identified 13 existing national surveys that include information on FSP participation and at least one of the outcomes of interest (food expenditures, household food supply, individual dietary intake, height and weight, food security). The data elements available in each survey are shown in Exhibit A.1.

Data Sources for Studying FSP Participation

The best data sources for studying this topic are the Survey of Income and Program Participation (SIPP), the Panel Study of Income Dynamics (PSID), and the Survey of Program Dynamics (SPD). All three are nationally representative longitudinal surveys.

The SIPP collects a wide variety of economic and demographic information on panels of respondents over a period of several years, contacting sample members every four months. Detailed information is available on earned and unearned income, program participation, assets and liabilities, and taxes. Labor force measures include hours worked per week, wage/salary, union membership, employer size and location, occupation, industry, work activities and duties. Topical modules provide information on education and training, migration, recipiency histories, work schedules, work-related expenses, work disabilities, marital history, fertility history, household relationships, expenses (housing, medical, work-related, other living), and functional limitations and disability, child well-being, and adult well-being. The recipiency history and detailed information about program participation and eligibility make the SIPP particularly relevant to studying disadvantaged populations. SIPP data have recently been used to study the influence of employment characteristics (McKernan and Ratcliffe, 2003) and long-run household income (Farrell et al., 2003) on FSP participation.

The SPD, which was initiated in 1997 following welfare reform, was specifically designed to provide information on spells of actual and potential program participation over a 10-year period and to examine the causes of program participation, the effects of national welfare reforms, how these reforms interact with each other, and how they interact with employment, income and family circumstances. It should be noted, however, that the survey has had a very serious problem with low response rates, especially among low-income households.

The PSID is a longitudinal study of income, employment, family composition, and residential location which has been ongoing on an annual basis since 1968. The content of the survey has remained largely unchanged over the decades in order to ensure comparability. The core topics addressed regularly in the PSID include income sources and amounts, poverty status, public assistance, other financial measures (taxes, interhousehold transfers), family structure and transitions (marital events, birth and adoptions, children forming households), labor force participation, housework time, housing, geographic mobility, socioeconomic background (education, ethnicity, religion, military service, parents' education, occupation, poverty status), and health. Other modules have addressed include food and housing expenditures, food sufficiency, health status, and fertility histories. There has been substantial turnover, but the panel still includes individuals who have been

interviewed since 1968. Members have also been added to the panel because of births, other additions to sample households, and establishment of new households by panel members. In 1990 a national sample of 2,000 Latino households was added.

Data Sources for Studying Relationships Between Hypothesized Outcomes

The hypothesized causal chain shown in Exhibit 2.1 suggests numerous links that could be tested if an appropriate extant database exists. Because few databases include information on the household food supply, we include some relationships that ignore this link and move further down the causal chain. The potential relationships to be examined include:

- Food expenditures → household food supply
- Household food supply → individual dietary intake
- Individual dietary intake → overweight/obesity
- Food expenditures → individual dietary intake
- Food expenditures → overweight/obesity
- Household food supply → overweight/obesity
- Food expenditures → food security

Below, we discuss the availability and appropriateness of extant data for addressing each of these relationships.

Food Expenditures and the Household Food Supply

With the exception of the Consumer Expenditure Survey (CES) and the National Food Stamp Program Survey (NFSPS), available food expenditure data are recall-based (past week, past month, past 12 months) and focus on *where* foods were purchased rather than on the specific types of food purchased. This type of data is not helpful in understanding the relationship between food expenditures and the nutritional quality of household food purchases or food use.

The CES provides detailed information on household food purchases (2-week diary). The NFSPS provides detailed information on food used at the household level over a 1-week period, including information on prices paid.[51] The NFSPS expenditure data have never been analyzed, although Jim Ohls reports that the data were cleaned and missing prices were imputed.

Advantages of the NFSPS are that it includes both cost and nutrient data, and that the food codes provide sufficient detail to support food-level as well as nutrient-level analyses. In contrast, the CES does not include nutrient data and the lack of nutrition-oriented detail in the food coding system necessarily limits the analyses that could be performed. On the other hand, the NFSPS is limited to FSP participants. Supplementary analyses on the CES could contrast participants and nonparticipants, as well as incorporate more timely data (from 2003).

---

[51] The data file contains, for each food item used during the week, the total weight used and the price per pound.

Household Food Supply and Individual Dietary Intake

No source has data in both of these domains.

Individual Dietary Intake and Overweight/Obesity

None of the identified longitudinal databases includes information on both usual dietary intake and height and weight.

Researchers have explored the relationship between *FSP participation* and overweight/obesity using the National Longitudinal Survey of Youth (NLSY) (Gibson, 2003 (adults) and 2004 (young children)) and the Panel Study of Income Dynamics—Child Development Supplement (PSID-CDS) (Jones et al., 2003). The only remaining longitudinal database that, to our knowledge, has not yet been used to study FSP participation and overweight/obesity is the Early Childhood Longitudinal Study, Kindergarten Cohort (ECLS-K).[52]

In our estimation, analyzing the ECLS-K data to look at FSP participation and overweight/obesity among young children would not be the optimal use of available project resources. This recommendation is consistent with the conclusion of the expert panel FNS convened to assess the literature in this area and provide recommendations. The panel cautioned against conducting new research on the issue of program participation and obesity without a greater understanding of (a) all the possible impacts of participation in food assistance programs and (b) the complex relationship between poverty, program participation, and obesity (Linz, Lee, and Bell, 2005).

Food Expenditures and Individual Dietary Intake

The CSFII is the only data set with information in both these domains. It does not hold a great deal of promise, however. Linking the measures directly is problematic because the food expenditure data are based on a 3-month recall and dietary intake on two 24-hour recalls. We considered an alternative approach of examining the relationship between *food costs* and individual intake—imputing food costs to the 24-hour recall data—but ultimately rejected the idea because the descriptive detail available in the 24-hour recall is not adequate to support such an analysis. For example, the food codes used in these data sets do not differentiate between expensive and inexpensive cuts of meat, prepared foods made at home *versus* store-bought, and single-serve packaging. Moreover, imputed

---

[52] Some investigators have begun to use this data set to study the relationship between participation in the school meals programs and obesity (Dunifon and Kowaleski-Jones, 2004; Whitmore, 2004).

food costs could not capture many of the behaviors low-income families use to stretch their food dollar, including coupon use, purchasing items on sale, and buying in bulk.[53]

Food Expenditures and Overweight/Obesity

Four extant data sets include information on food expenditures and height and weight—CSFII, NLSY, NLSY-C, and PSID-CDS. Two of these can be eliminated from consideration. The CSFII is a cross-sectional database, so the multiple measures of height and weight needed to adequately assess the relationship between any variable and the prevalence of overweight or obesity are not available. The PSID-CDS, though longitudinal, has only crude data on food expenditures, using a 12-month recall period.

The other two data sets, however, could potentially support such an analysis, as they are longitudinal and include one-month measures of food expenditures. Arguably, households that consistently (over many years) spend more on food can achieve higher levels of food security and diet quality, which can lead to lower rates of overweight/obesity. The data collected for the NLSY79 and NLSY-C might be used to explore this hypothesis.

Household Food Supply and Overweight/Obesity

No database has information in both of these domains.

Food Expenditures and Food Security

Longitudinal data are essential to advancing our understanding of the relationship between FSP participation and food security. Four of the longitudinal databases include the full 18-item food security module and could be used to study this relationship. One author has already used data from the Survey of Program Dynamics (Huffman and Jensen, 2003). It is not clear that additional analyses of these data or analysis of the other data sets (ECLS-K, PSID, and PSID-CDS) would contribute to a meaningful advancement of our knowledge in this area. In particular, the 12-month recall period for the food security measure makes it poorly suited for use in an impact analysis.[54]

Good cross-sectional data in both domains are available in the CPS-FSS. The annual Food Security Supplement of the CPS contains both a food security battery that refers to the past 30 days and information on household food expenditures that refers to the past week. Households are first queried

---

[53] We also considered an analysis that would relate FSP participation and individual dietary intake. However, this analysis has already been completed for the most recent data from the Continuing Survey of Food Intakes by Individuals (CSFII) (Gleason et al., 2000), and an ongoing Abt project will be exploring this relationship using the latest data from the National Health and Nutrition Examination Survey (NHANES).

[54] We note that Mark Nord has suggested an analysis of CPS data to *relate total household income including food stamp benefits to food security*. The purpose of this is to provide a lower-bound estimate of the impact of FSP participation on food security. We are hesitant to recommend this, however, as it seems plausible that absent the Food Stamp Program households would make other adjustments that would affect their food security, such as increased reliance on friends and family, increased labor supply, and so on.

---

as to whether they bought any food at each of several venues (supermarkets and grocery stores; meat markets, produce stands, bakeries, warehouse clubs, and convenience stores; restaurants, fast food places, cafeterias, and vending machines; and "other places"). For each affirmative response, they are asked how much they spent at the venue; and for purchases from stores, how much was spent on "non-food items such as pet food, paper products, detergents, or cleaning supplies". They are also asked their "usual" weekly expenditures for food.

These data are analyzed and tabulated by ERS on a regular basis (see for example Nord *et al*. 2005). The publications include tabulations of median weekly food expenditures by food security status. In addition, research is proceeding at ERS to exploit the possibility of linking successive years of the CPS data in order to explore the dynamic relationships and control for unobserved household characteristics.[55]

Further analysis of these data seems a lower priority than analysis of data from other sources because of the work already done and in progress at ERS. Nonetheless, if resources permitted, new analyses that focused explicitly on the role of food assistance programs in mediating the relationship between food expenditures and food security could be a valuable undertaking.

---

[55] The CPS sampling structure provides fifty percent overlap from one year to the next. It must be noted, however, that the CPS interview unit is the address, not the people living there. If a household moves away, the survey replaces the respondents with the new people living at that address. Hence the overlapping sample excludes movers, and is in that way nonrepresentative.

Exhibit A.1: Summary of Extant National Surveys that Include Data on FSP Participation and Outcomes of Interest

Survey	Most Recent Data	Data on FSP Participation	Participation in Other Programs	Other Relevant Data				
				Food Expenditures	Household Food Use	Individual Dietary Intake	Height and Weight	Food Security
Consumer Expenditure Survey (CES)	2003	Past 12 months (Y/N); number of months in past 12; receipt date and benefit level for each of past 12 months (including current); current month (Y/N)	NSLP or SBP (Does not differentiate)	2-week diary of food and beverage purchases, including both at home and away				
Continuing Survey of Food Intakes by Individuals (CSFII)	Survey no longer done (has been combined with NHANES). Most recent data for 1994-96 and 98. 98 data includes only children 9 and under.	Past 12 months (Y/N); current month (Y/N); receipt date and benefit level for current/most recent month	WIC, NSLP, SBP	Reported weekly food expenditures over preceding 3-month period (grocery stores, specialty stores, fast-food and carryout places; food bought and eaten away from home)		2 non-consecutive 24-hour recalls	Self-reported	Questions about food sufficiency (not FSS module)
Current Population Survey-Food Security Supplement (CPS-FSS)	2004	Last month (Y/N); benefit level for current/most recent month	WIC, NSLP, SBP	Reported food expenditures for previous week (supermarkets and grocery stores; specialty stores, warehouse clubs, and convenience stores; restaurants, fast food restaurants, cafeterias, and vending machines; and any other locatio.)				Full FSS module
Early Childhood Longitudinal Study, Kindergarten Cohort (ECLS-K)[a]	2004 Separate databases for K, 1st, and 3rd grade years as well as longitudinal database with data for all 3 time points.	Past 12 months (Y/N); number of months in past 12 months; since birth of sample child (Y/N)	WIC, NSLP, SBP				Measured	Full FSS module
National Food Stamp Program Survey (NFSPS)	1996–97	Current month (Y/N); last month (Y/N); receipt date and benefit level for current/most recent month; length of last spell	WIC, NSLP, SBP	Reported weekly food expenditures over preceding 3-month period (grocery stores, specialty stores, fast-food and carryout places; food bought and eaten away from home)	Detailed 1-week records of household food use			Full FSS module
National Health and Nutrition Examination Survey (NHANES 99-00 and 01–02)	Two 2-year waves: 1999–2000 and 2001–02. Can be analyzed separately or together, but not by single year.	Past 12 months (Y/N); number of months in past 12 months; current month (Y/N)	WIC, NSLP, SBP			Single 24-hour recall; variance components needed to estimate usual nutrient intake will be released by NCHS	Measured	Full FSS module

Exhibit A.1: Summary of Extant National Surveys that Include Data on FSP Participation and Outcomes of Interest

Survey	Most Recent Data	Data on FSP Participation	Other Relevant Data					
			Participation in Other Programs	Food Expenditures	Household Food Use	Individual Dietary Intake	Height and Weight	Food Security
National Longitudinal Survey of Youth 1979 (NLSY79)	2000 Data available back through 1979, when respondents were 14-21 years old.	Past 12 months (Y/N); number of months during past 12; benefit level for current/most recent month	WIC	Reported food expenditures for previous month (food used at home; food delivered to door; eating out)		Questions about dietary practices during pregnancy and infant feeding practices	Self-reported	
National Longitudinal Survey of Youth—Children (NLSY-C)	2000 Sample is children born to women in NLSY79. Data available back through 1986, when children were 0 to 9 (now 0 to 14)	Past 12 months (Y/N); number of months during past 12; benefit level for current/most recent month	WIC	Reported food expenditures for previous month (food used at home; food delivered to door; eating out)		Questions about infant feeding practices	Measured and self-reported	
National Longitudinal Survey of Youth—Young Adults (NLSY-YA)	2000 Sample is children born to women in NLSY79 who are now older than 14 (ages 15-22). Data available back through 1994, when children were 15 and older.	Current calendar year (Y/N); lifetime (Y/N); number of months in calendar year; benefit level for current/most recent month; average monthly benefit for each year	WIC			Questions about dietary practices during pregnancy and infant feeding practices	Self-reported	
Panel Study of Income Dynamics (PSID)	2003	Past 12 months (Y/N); number of months in last 12; last month (Y/N); prenatally (Y/N); benefit level for current/most recent month	WIC, NSLP, SBP	Reported expenditures over past 12 months for food used at home, food delivered to home, and food eaten out				Full FSS module
Panel Study of Income Dynamics—Child Development Supplement (PSID-CDS)	2003	Past 12 months (Y/N); number of months in last 12; last month (Y/N); prenatally (Y/N); benefit level for current/most recent month	WIC, NSLP, SBP	Reported expenditures over past 12 months for food used at home, food delivered to home, and food eaten out		Questions about whether child usually eats breakfast	Measured height and parent-reported weight	Full FSS module
Survey of Income and Program Participation (SIPP)	1996 (est.)	Start and length of first spell; no. spells, participated in past 4 months, monthly detail on benefit level	WIC, NSLP, SBP					Limited questions on food sufficiency and hunger
Survey of Program Dynamics (SPD)	1998 (est.)	Past 4 months (Y/N); number of months in past 4; benefit level for each month	WIC, NSLP	Reported expenditures over past 12 months for groceries and for meals from restaurants, fast food, cafeterias, vending				Full FSS module

Exhibit A.1: Summary of Extant National Surveys that Include Data on FSP Participation and Outcomes of Interest

Survey	Most Recent Data	Data on FSP Participation	Participation in Other Programs	Food Expenditures	Household Food Use	Individual Dietary Intake	Height and Weight	Food Security
				Other Relevant Data				

NOTES:

Table includes all national surveys identified in Logan, Fox, and Lin (2002) that include data on FSP participation and at least one of the outcome measures of interest.

[a] A related survey built around a cohort of children born in 2001 (ECLS-B (Birth Cohort)) is also ongoing. Data have been released for the 9-month measurement point. Next measurement point is 2 years of age.