



United States
Department of
Agriculture

Food and
Nutrition
Service

Office of
Analysis,
Nutrition, and
Evaluation

Feasibility Study of Capturing Food Data at Checkout

FINAL REPORT

September 1999

Authors:

John Kirlin

Nancy Cole, Abt Associates

William Adam

Charles Pappas, Consumer Card Marketing, Inc.

Submitted by:

Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 02138

Project Director: John Kirlin

Submitted to:

Office of Analysis, Nutrition, and Evaluation
USDA Food and Nutrition Service
3101 Park Center Drive, Rm. 214
Alexandria, VA 22302

Project Officer: Ken Offerman

This study was conducted under Contract No. 53-3198-4-022 with the Food and Nutrition Service, United States Department of Agriculture, under the authority of the Food Stamp Act of 1977, as amended. Points of view or opinions stated in this report do not necessarily represent the official position of the Food and Nutrition Service.

"The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, gender, religion, age, disability, political beliefs, sexual orientation, or marital or family status. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD).

To file a complaint of discrimination, write USDA, Director, Office of Civil Rights, Room 326-W, Whitten Building, 14th Street and Independence Avenue, SW, Washington, DC 20250-9410 or call (202) 720-5964 (voice and TDD). USDA is an equal opportunity provider and employer."

Acknowledgments

I wish to thank a number of individuals for their assistance and support during the course of this project. First, the hard work and diligence of the staff of our subcontractor, Consumer Card Marketing, Inc. (CCMI), cannot be overlooked. Mr. William Adam, CCMI project director, participated fully in the study from its inception through to this final report. I would also like to thank Ms. Ann Raider, senior vice president at CCMI, and Mr. Charles Pappas, CCMI's research director, for their assistance in recruiting retailers for the study and in working out the details of data collection from the participating supermarkets. Special thanks are reserved for Ms. Amy Claus Grozier, who traveled frequently to participating non-scanning stores in Georgetown County to create their store-specific price files, set up their data collection scanning units, train the retailers and their staff, and upload collected data on a weekly basis.

Mr. Adam joins me in extending special gratitude to the eleven retail stores in Georgetown County, South Carolina that participated in the study. In order to preserve the confidentiality of these stores and the data collected, we cannot thank these retailers by name. Nevertheless, our gratitude to these individuals is not diminished by their anonymity. Without their support, this project could not have gone forward.

We would also like to thank the South Carolina Department of Social Services. Mr. Paul Brawley was the EBT director in South Carolina when this project was initiated, and he and his staff were most helpful to us as we sought to identify potential study sites within the state. Mr. Jim Apple of his staff was particularly helpful in joining project members during a visit to Georgetown County and introducing them to retailers in the county.

At FNS, I wish to thank the many individuals who contributed to this feasibility study. As project officer, Mr. Ken Offerman of the Office of Analysis, Nutrition, and Evaluation (OANE) provided ongoing support and guidance to issues of research design, analysis and presentation of results. Similar support was provided by others within OANE, including Steven Carlson, Theodore Macaluso, and Carol Olander. Ms. Debbie McIntosh of the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) provided detailed information on the WIC Program's successful efforts to link scanner data to EBT transactions.

Finally, from Abt Associates, I would like to thank Dr. Nancy Cole, the project's lead analyst, and Ms. Ellen Lee, the project's programmer, for their work in completing the difficult task of assembling the merged analysis files and in exploring the types of analysis that can be conducted on these scanner data. I would also like to thank Dr. William Hamilton, the project's technical reviewer, Ms. Susan Byers Paxson, who handled all report production tasks, and Mr. John Lightfoot, who—as a consultant to the project—helped with the initial stages of project design and contacts within the food retail industry.

John Kirlin
Project Director

Contents

Executive Summary	1
Chapter 1: Introduction	9
1.1 Context for the Study	10
1.2 Associated WIC Efforts	11
1.3 Report Organization	12
Chapter 2: Selecting a Study Site and Retailer Recruitment	13
2.1 Identifying EBT States	13
2.2 Identifying Potential Study Sites	14
2.3 Determining the Retailers' Willingness to Participate in the Study	16
2.4 Final Site Selection and Retailer Recruitment	19
2.5 Retailer Agreements	22
Chapter 3: Equipping Retailers and Collecting Data	25
3.1 Data Collection Scanning Unit	25
3.2 Equipment Purchase and Configuration	27
3.3 Building the Price File, Equipment Installation, and Training	29
3.4 Data Collection at the DCSU Stores	31
3.5 Data Collection at the Supermarket Chains	33
Chapter 4: Processing Scanner Data	35
4.1 Description of Scanned Food Purchase Data	36
Scope of Data	36
File Formats and Media	36
4.2 Initial Processing of the Scanned Data	37
Reformatting the Scanner Data	37
Composition of Reformatted Files	38
Impact of Missing Data in Scanner Files	38
4.3 Construction of the Master Item File	41
Assigning Category Codes	44
Identifying Food Stamp-Eligible Items	46
4.4 Data Processing: Matching Scanner Data to ALERT EBT Log	47
Matching Scanner Data to ALERT	47
4.5 Completeness of Captured Scanner Data	50
4.6 Adjustments to Combination-Tender Transaction Records	50
4.7 Demographic Data	52

Chapter 5: Food Purchasing Patterns of Food Stamp Households	55
5.1 Sample of FSP Retailers and Households	55
Captured Scanner Data, by FSP Household	56
Demographic Characteristics of FSP Households	60
The Sample Period	63
Summary of Sample Definition	63
5.2 Application of Scanner Data to Proposed Analysis	64
Completeness of Scanned Data	64
Food Purchasing Patterns	65
Consumer Behavior	65
5.3 Characteristics of Food Baskets Captured at Checkout	65
Number and Size of Food Baskets	66
Distribution of Captured FSP and Non-FSP Food Baskets, by Time of Month	66
5.4 Characterization of the Content of Food Baskets	72
Distribution of Spending on Items Purchased, by Store Department	72
Percentage of Item Purchases on Store-Brand Items	74
Distribution of Item Purchases by Product Category	74
Distribution of Purchases of Specific Items, by Quantity Size	79
Distribution of Purchases by Nutritional Characteristics	86
5.5 Characteristics of DCSU Data	92
Chapter 6: Lessons Learned	95
6.1 Collecting Data from Retailers with Scanning Systems	95
6.2 Collecting Data from Retailers without Scanning Systems	96
6.3 Matching Scanner Data to Recipient Characteristics	98
6.4 Analyses Supported by the Data	98
6.5 Feasibility of Expanding Data Collection Efforts	99
Retailer Recruitment	100
Data Collection at Non-scanning Stores	101
Cost of Collecting Scanner Data	102
6.6 Potential Research Using Scanner Data	104
Effects of USDA Initiatives on Food Expenditures	104
Monitoring of Long-Term Trends in Food Purchasing Behavior	105
Effects of Workforce Participation on Food Purchasing Behavior	105
Food Stamp Expenditures at Different Store Types	105
More Complete Expenditure Studies	106
Methodological Studies	107
Appendix A: FSP Research on Food Consumption	
Appendix B: Sample Recruitment Letter to Retailers	
Appendix C: Sample Data Collection Agreements	
Appendix D: Installation Guide and End-user Manual for Data Collection Scanning Unit (DCSU)	

List of Exhibits

Exhibit 2-1	States with EBT Systems in December 1996	13
Exhibit 2-2	Potential Study Sites	17
Exhibit 2-3	South Carolina Counties with Georgetown Study Site	20
Exhibit 2-4	Georgetown County, South Carolina	20
Exhibit 2-5	Participating Stores	22
Exhibit 3-1	Data Collection Scanning Unit (DCSU)	26
Exhibit 3-2	Compsee Scanning Unit	28
Exhibit 3-3	DCSU Components and Price	29
Exhibit 3-4	Missing Data at Supermarkets	34
Exhibit 4-1	Characteristics of Collected Data	36
Exhibit 4-2a	Transaction Record Layout: Transaction Summary File	39
Exhibit 4-2b	Item Record Layout: Item-Detail File	39
Exhibit 4-3	Categorization Schemes Provided in the Chain A Host Price File	43
Exhibit 4-4	Re-classification of Items from Broad to Narrow Product Categories	45
Exhibit 4-5	Match of ALERT and Scanner Data	49
Exhibit 4-6	EBT Transactions and Redemption Captured in Scanner Data	51
Exhibit 4-7	Information Obtained from South Carolina Administrative Records	53
Exhibit 5-1	EBT Transactions in Georgetown County South Carolina	57
Exhibit 5-2	Distribution of FSP Households, by Percent of EBT Redemptions at Sampled Stores	58
Exhibit 5-3	Captured and Uncaptured FSP Redemptions for Households EVER Shopping at Sample Stores, September and October 1997	59
Exhibit 5-4	Characteristics of FSP Households and FSP Redemptions, by Percent of Captured Data: Households EVER Shopping at Sample Stores	61
Exhibit 5-5	Characteristics of FSP Redemptions, by Household Type: Households EVER Shopping at Sample Stores	62
Exhibit 5-6	Characteristics of Transaction Baskets Scanned at Checkout, by Tender Type	67
Exhibit 5-7	Number of Food Stamp and Non-Food Stamp Transactions at Sample Stores, by Day: September 1997	68
Exhibit 5-8	Percent of Food Stamp and non-Food Stamp Transactions at Sample Stores, by Day: September 1997	69
Exhibit 5-9	Average Number Items per "Basket," by Day: Food Stamp and Non-Food Stamp Transactions at Sample Stores, September 1997	70
Exhibit 5-10	Average Basket Value (\$) by Day: Food Stamp and Non-Food Stamp Transactions at Sample Stores, September 1997	71
Exhibit 5-11	Distribution of Spending on FSP-Eligible Items in Food Stamp and Non-Food Stamp Purchases, by Department	73
Exhibit 5-12	Distribution of Food Stamp Redemptions, by Supermarket Department and House- hold Type: Redemptions at Seven Supermarkets in Study Site	75
Exhibit 5-13	Percent of FSP and Non-Food Stamp expenditures on Store Brand Items	76
Exhibit 5-14	Store Brand Percent of Food Stamp Redemptions and Non-Food Stamp Purchases, by Supermarket Department	77

Exhibit 5-15	Store Brand Percent of Food Stamp Redemptions, by Household Type and Super-market Department	78
Exhibit 5-16	Distribution of Food Stamp and Non-Food Stamp item Purchases, by Product Category	80
Exhibit 5-17	Distribution of Food Stamp Redemptions, by Product Category and Household Type	81
Exhibit 5-18	Distribution of Milk Purchases by Quantity Size: Food Stamp and Non-Food Stamp Purchases	82
Exhibit 5-19	Distribution of Sugar Purchases by Quantity Size: Food Stamp and Non-Food Stamp Purchases	83
Exhibit 5-20	Distribution of Flour Purchases by Quantity Size: Food Stamp and Non-Food Stamp Purchases	84
Exhibit 5-21	Distribution of Breakfast Cereal Purchases by Quantity Size: Food Stamp and Non-Food Stamp Purchases	85
Exhibit 5-22	Distribution of Milk Purchases by Milk Type: Food Stamp and Non-Food Stamp Purchases	88
Exhibit 5-23	Distribution of Purchased Breakfast Cereal Servings by SUGAR Content: Food Stamp and Non-Food Stamp Purchases	89
Exhibit 5-24	Distribution of Purchased Breakfast Cereal Servings by IRON Content: Food Stamp and Non-Food Stamp Purchases	90
Exhibit 5-25	Distribution of Purchased Breakfast Cereal Servings by FAT Content: Food Stamp and Non-Food Stamp Purchases	91
Exhibit 5-26	Items Purchased in FSP Transactions at DCU Stores	93
Exhibit 6-1	Estimated Scanning Capabilities of FSP-Authorized Stores, by Store Type	97

Executive Summary

The Food and Nutrition Service (FNS) of the U.S. Department of Agriculture is interested in knowing more about the impacts of the Food Stamp Program (FSP) on the diet and nutritional status of program participants. In the past, FNS has relied on a series of large surveys to gather and compare information on food expenditures and food consumption among program participants and non-participants. Studies based on survey data, however, have a number of drawbacks, including the expense of collecting the survey data, sampling error, response bias (resulting from sampled respondents who cannot be located or who refuse to participate in the survey), errors in respondent recall of what was purchased or consumed, and deliberate misinformation about what was purchased or consumed. Furthermore, due to increased difficulties in obtaining cooperation from respondents, large-scale surveys are becoming harder and more expensive to conduct.

Recognizing that another source of information on food expenditures exists, FNS has sponsored a study to examine the feasibility of collecting and analyzing item-level, bar-coded data captured by food retailers' optical scanning systems at the checkout. Specifically, the study tested the feasibility of collecting scanned bar-code data on the Universal Product Codes (UPCs) and price look-up (PLU) codes of items paid for with food stamp benefits. These UPCs and PLUs identify the products being purchased. Additionally, the study examined the feasibility of merging these food purchase data with transaction data from electronic benefits transfer (EBT) systems so that food stamp participants' demographic information could be used to improve FNS' understanding of the shopping behavior of food stamp households.¹ This report presents the findings of the feasibility study.

Study Description

FNS selected Abt Associates Inc. to conduct the feasibility study. Consumer Card Marketing Inc. (CCMI), a firm with extensive experience in collecting and analyzing scanner data from retail stores, is a subcontractor to Abt Associates for this study.

The project began with an in-depth review of scanner technology and how it is applied in the retail food industry. The results of that review indicated that it should be technically feasible to collect bar-code data on products purchased with food stamp benefits and to link those data with demographic information about the food stamp household. The review also assessed the likely cost to collect bar-code data in stores with different types of point-of-sale (POS) systems.

The project then turned to finding a suitable site for testing the feasibility of collecting and analyzing bar-code information. An initial search identified 17 counties in six EBT states as potential candidates. These counties had enough program-authorized retailers (more than 25) to provide an opportunity to collect data from a variety of stores of different size and POS systems, yet they were not too large (fewer

¹ It should be noted that, for the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), a specially-designed EBT system already links detailed information on purchased WIC items to EBT transaction data. The EBT system design used to capture this information and deliver WIC benefits, however, cannot be used to capture similar food stamp data at a reasonable cost.

than 100 retailers) to serve as a test site for a feasibility study—the project planned to collect data from enough stores to account for a substantial portion of all food stamp redemptions by program recipients. After a detailed examination of the FSP-authorized retailers in these 17 counties, FNS selected Georgetown County, South Carolina as the project’s study site. Located along South Carolina’s eastern coastline, Georgetown County had about 53 FSP-authorized retailers and 2,600 food stamp recipients when it was selected.

The project’s planned first phase included recruiting about a dozen retailers to test the feasibility of collecting, merging, and analyzing food stamp scanner data. At the outset, the expected major obstacles included: (1) recruiting retailers for the study; (2) combining scanner files from different retailers together into a consistent format; and (3) merging together information from retailer scanner records, EBT transaction log records, and FSP eligibility file records. In a second, optional phase of the project, as many retailers in the county as possible would be recruited to maximize collected information on the buying patterns of recipients in Georgetown County. For a number of reasons, including financial constraints, the second phase of the study has been dropped from FNS’ research agenda. This current report therefore presents the study’s final results.

Project staff recruited 11 stores to participate in the project’s first phase. Seven of the stores represent two large supermarket chains operating in Georgetown County. The remaining four stores are small grocery or convenience stores that do not use scanners. Special data collection scanning units (DCSUs) were installed in these four stores for the study’s data collection period, which ran during September and October 1997.

Study Findings and Lessons Learned

Linking Food Stamp Purchase Data to Household Characteristics

The main finding from this study is that it is indeed possible to collect food purchase data from stores with scanning systems and to link UPC and PLU data from food stamp purchases to information about the food stamp recipient making the purchase. In one of the two supermarket chains participating in the study, scanner data were matched to over 98 percent of EBT transactions; the match rate exceeded 96 percent in the other chain.

With these high match rates, it is possible to examine relationships between recipient and household characteristics and what types and quantities of food items are being purchased with food stamp benefits. For example, the study divided all food stamp households into three groups (households with elderly persons and no children, households with children, and households with neither children nor elderly members) and examined the following differences in how these three groups allocate their food stamp benefits:

- distribution of items across broad food groupings (e.g., meat, produce, dairy);
- distribution of items across detailed food categories (e.g., red meat, poultry, fish);
- distribution of store-brand food items versus national-brand items;
- distribution of milk, sugar, and cereal purchases by product size; and
- distribution of purchased breakfast cereal servings by sugar, iron and fat content.

Actual findings are not summarized here due to the small number of participating stores and the resulting non-generalizability of the collected data.

Technical Feasibility of Collecting Data in Stores Without Scanning Systems

The study also demonstrated that it is difficult to collect information on what is being purchased in stores that do not already use scanner systems. The difficulty is not mainly technical. Although the DCSUs installed for this study encountered a few hardware and software problems that reduced the amount of usable data, these technical problems could be addressed in future implementations. Of greater consequence was the difficulty and/or unwillingness of store employees to use the DCSUs consistently. The DCSUs were simply not used for many transactions at these stores, often because employees said they were too busy to use both the DCSU and their own cash register.

An estimated 26 percent of all food stamp benefits are spent in stores without scanning systems. FNS has three choices for how to handle these stores in any future research efforts involving food purchase data. The first alternative is to exclude such stores, thereby limiting research to purchases made in large grocery stores and supermarkets and, possibly, a few large convenience store chains using scanners. The second option is to wait until more program-authorized stores decide to install scanner systems on their own (or perhaps offer some incentives for them to install scanners). The third is to again test the concept of collecting data with a portable DCSU, but with a redesigned, easier-to-use system that could be substituted temporarily for the retailer's existing cash register.

It is not known whether efforts to collect bar-coded data in non-scanning stores with a reconfigured and easier-to-use DCSU would be more successful. Nevertheless, because there is a policy interest in learning how these stores serve the buying needs of food stamp recipients, it may be worth applying the lessons learned from this first test of the DCSU concept to a second effort. An important lesson from this study is that, if further efforts to collect scanner data with DCSUs are planned, more time needs to be allocated for on-site training and early monitoring of DCSU use than was done for the current study.

Recruiting Efforts

One of the lessons learned early in the study is that FNS cannot expect universal support from state EBT directors in its efforts to link EBT data and recipient information to scanner records. The study found that a small number of EBT directors are philosophically opposed to using information from their EBT systems to examine what food stamp recipients are purchasing with their FSP benefits. At least one other EBT director who was contacted had promised concerned advocates that the EBT system would not be used in this manner.

FNS will not be able to conduct research using scanner data unless food retailers are willing to provide these data to the agency. Furthermore, such research will not be affordable unless retailers are willing to provide the data "at cost" (i.e., for the actual cost to the retailer of copying POS transaction logs and transferring the data to FNS). Thus, obtaining retailer support is crucial for future research efforts using scanner data. With respect to obtaining this support, the study provides a number of lessons:

- It is difficult to get the attention and support of food retailers for a study of this nature. The fact that this was a USDA-sponsored study made little difference as long as store participation was not mandated.
- Despite recruiting difficulties, it is possible to obtain scanner data from some retailers at a reasonable cost. Not all retailers, however, will be willing to provide scanner data on all transactions. Of the two supermarket chains recruited for this study, one agreed to provide data from all transactions, whereas the other provided scanner data only from food stamp purchases.
- Among store chains, support was greatest within supermarkets and weakest within convenience stores. This difference is due, at least in part, to the different levels of scanning experience and knowledge across the two store groups.
- Some stores are simply unwilling to participate voluntarily in **any** government study.
- Even for stores that eventually said they were willing to participate in the study, it was often difficult to obtain this support. Corporate CEOs usually directed us to a division manager to discuss the proposed project. These managers were often quite busy and, without a strong directive from the CEO, they had little to gain by talking with study staff about the technical details of their POS system and how to transfer data to the study.
- Food purchase data are quite valuable to marketers, and some store personnel were reluctant to provide scanner data at cost. Somewhat surprisingly, this attitude was less often found among CEOs than their senior managers.

Thus, with regard to retailer participation, the two most important lessons for the future are: (1) it should be possible to recruit a number of stores to provide food purchase data, at least for food stamp purchases, but (2) support will not be anywhere near universal. In areas where one or more reluctant retailers dominate the market, there is little hope that sufficient scanner data can be collected for research purposes.

Data Collection, Preparation, and Analysis

The study provided several lessons dealing with data collection, data preparation, and analysis:

- Even when stores with scanning systems agree to provide food purchase data, one should anticipate some loss of data due to telecommunications or POS system problems. During the study's two-month data collection period, several days of supermarket data were lost for these reasons.
- Due both to variation among stores in how they organize their host price files and the size of these files, construction of a combined master item file—with detailed and consistent information on every item in store inventories—will be labor intensive. For instance, nearly 34,000 unique items were scanned at the 11 stores in the study, and the master item file needs to provide sufficient information on these items to allow classification into product categories. Nevertheless, without this initial effort, analysis of buying patterns across stores is impossible.

- Once the master item file is constructed, it is necessary to develop a meaningful taxonomy of food products to facilitate analysis. For this study, all food items were first assigned to one of 243 detailed product categories. These 243 detailed categories were then collapsed into 35 summary categories within six main groupings.
- Research using scanner data should focus on the shopping trip as the unit of analysis, because the data are recorded on a transaction-by-transaction basis. Because most food stamp households shop at multiple stores during the month, it may not be possible to obtain their complete food stamp shopping record.
- Nevertheless, detailed information on what food stamp recipients buy with their benefits can be collected and analyzed. Furthermore, variations in buying behavior across subgroups of the food stamp caseload can be examined after demographic information is merged to the scanner data.
- Finally, it is also possible to compare what is being purchased with food stamp benefits to non-food stamp purchases. This analysis cannot be related to household or purchaser characteristics, however, because no information is available on the characteristics of non-food stamp shoppers. The analysis also is possible only when the store is willing to provide scanner data from all transactions.

Data Collection Costs

Little prior evidence is available for estimating the cost of collecting scanner data. Although several companies collect scanner data for market research purposes, no information is available on their cost structures. This study's experience in Georgetown County, however, provides preliminary information on data collection costs related to stores with and without existing scanning systems.

The total cost to collect scanner data from the two participating supermarket chains was about \$79,000, which covered initial retail contacts, contract negotiations, data collection for two months, and initial processing of the files provided by the two chains. On a per-chain basis, the average cost was \$39,500. In thinking about costs for possible future data collection efforts, one should use a somewhat lower figure—about \$30,000 per chain—to account for greater efficiency and the fact that some tasks (e.g., writing programs for data processing) have already been performed. This estimate assumes that, as in the current study, retailers will be willing to provide copies of their scanner data at cost. If future data collection efforts were to last longer than two months, the expected average cost would be perhaps as high as \$35,000 per chain. Most of the costs are for the up-front tasks of retailer recruitment and working out procedures for transfer of data, so lengthening the data collection period should have only a marginal impact on total costs.

To place the \$35,000 per chain cost estimate in context, it is useful to compare it to the cost of a major survey designed to collect data on food shopping patterns, food expenditures, and household food use. The National Food Stamp Program Survey (NFSPS) was conducted in 1996-97. The total cost of the survey was about \$1.7 million. Thus, one could collect scanner data from approximately 50 store chains for about the same price as the NFSPS. With each chain representing an estimated 43 stores, this means that scanner data could be collected from about 2,150 supermarkets for the same price as the NFSPS. Such a sample would represent an estimated 6 percent of all FSP redemptions.

Data collection costs at stores without scanners were much higher, about \$47,000 per individual store, for several reasons. First, the DCSU itself—which included a small computer-based register with keyboard, display screen, and scanner—had to be configured, programmed and installed. Next, project staff had to build a “price file” for each store. (Stores with scanning systems need a datafile that relates UPC and PLU data to item prices, so that the correct price is registered when the item is scanned.) During the feasibility study, project staff used hand-held devices to scan the UPCs and enter the price and description of all items in each store’s inventory. The four non-scanning stores in the study had a combined inventory of about 9,000 items, with surprisingly little overlap in the inventories of the four stores. Finally, project staff visited the four stores each week to download data from the DCSUs and assist store personnel.

Representativeness of Scanned Food Purchase Data

It was not the goal of this feasibility study to collect a **representative** sample of scanner data for analysis. Nevertheless, based on what was learned about recruiting problems and use of DCSUs in non-scanning stores, it is clear that collecting scanner data representing all food stamp purchases will not be possible in the near future. First, as mentioned, an estimated 26 percent of all food stamp redemptions occur in stores lacking scanning systems, so a substantial segment of all redemptions is lost to analysis unless a workable and cost-effective DCSU solution can be found. Second, given the difficulties of encouraging retailers to voluntarily provide their scanner data at cost, it will be very difficult to create a random sample of food purchase data even for FSP redemptions within scanning stores.

The inability of collecting a nationally representative sample of scanner data poses serious limitations on the use of these data for research purposes. It may be possible, however, to collect reasonably representative samples of scanner data in selected market areas (i.e., those in which most or all stores that scan agree to provide their data).

Future Research Possibilities

Despite the difficulties present in trying to obtain a representative sample of scanner data for food stamp purchases, several possible research uses exist for these data. As described below, these research topics include validation of survey methodologies and evaluation of USDA initiatives designed to change food expenditure patterns.

One problem with food expenditure data collected during household surveys is that the accuracy of the collected information is difficult to assess. Errors may arise from respondent recall problems or deliberate efforts to provide misinformation. With the detailed information available in scanner data, it should be possible to design a study to cross reference scanner data from food stamp purchases with survey expenditure data from a sample of food stamp households, thereby identifying survey errors. With this information, researchers would have a better understanding of the strengths and limitations of survey expenditure data.

Data that are not representative of events at a single point in time can still provide useful information on temporal changes when gathered at multiple points in time. For instance, scanned food purchase data could be collected at two or more points in time to evaluate the effects of USDA initiatives on food expenditures. Examples of such initiatives might include a nutrition education campaign or an effort to

help food stamp recipients become more price conscious when food shopping. When such initiatives are implemented at the local level, scanner data have clear advantages over survey data for evaluating the impacts of the initiative. First, national survey data generally do not provide sufficient coverage within small areas to support estimation of local-area effects, whereas it is generally easier to collect a high percentage of scanner data within a small area than a large area (because fewer retailers need to be recruited). Second, the marginal cost of collecting scanner data at a second point in time (to provide a pre-post comparison of shopping behavior) should be low for scanner data, whereas survey costs should be about the same for each wave of data collection.

The low marginal cost of collecting scanner data at future points in time suggests other research possibilities. In evaluating the effects of an initiative to change food shopping patterns, collecting scanner data at multiple periods would allow estimation of not only the immediate effects of the initiative, but also any decay in treatment effects. That is, even if food stamp recipients respond to an initiative, one does not necessarily know whether the change in shopping patterns will persist. With scanner data collected at a third point in time, the persistence (or decay) of the effect could be estimated. Collecting scanner data at multiple points (e.g., annually over a 10 to 20 year period) could also allow long-term monitoring of underlying changes in the food shopping patterns of food stamp recipients and other households.

Chapter 1

Introduction

The Food and Nutrition Service (FNS) of the U.S. Department of Agriculture (USDA), together with designated state agencies, administers the Food Stamp Program (FSP) and other nutrition assistance programs. The goal of the FSP is to improve the nutritional status of low-income households. The program seeks to achieve this goal by providing to eligible households benefits that are earmarked for the purchase of eligible food items at program-authorized food retail outlets.

In order to understand better the impacts of program participation on diet and nutrition, FNS has relied in part on a series of large surveys. Some of these surveys have measured household expenditures on food, whereas others have focused on food consumption at home, either at the household or individual level.¹ Another possible source of information on food expenditures by food stamp recipients is scanned bar-code information collected by retailers' optical scanning systems. Scanned data on Universal Product Codes (UPCs) and price look-up (PLU) codes, when captured and retained at the individual transaction level, provide detailed information about what was purchased during a grocery shopping trip. To see whether or not scanned food purchase data can be used to help assess the benefit expenditure patterns of food stamp recipients, FNS awarded a contract² to Abt Associates Inc. to conduct a feasibility study with the following major objectives:

- Determine the technical feasibility and cost of collecting scanner data from those FSP-authorized stores using point-of-sale (POS) systems with optical scanners.
- Test the feasibility of using portable scanning devices to collect scanner data from program-authorized stores that do not use optical scanners.
- Determine the feasibility of linking scanner data to electronic benefits transfer (EBT) transactions.
- Recruit a sufficient number of FSP retailers in a selected geographic area to provide a representative sample of FSP store types within that area. Install necessary equipment and procedures to collect scanner data from retailers participating in the study.
- Create multi-store analysis files and describe the food purchasing patterns of FSP participants in the selected geographic area. Examine the differences between FSP food purchases and non-FSP purchases for a sample of purchases.
- Identify and compare the relative advantages and disadvantages, including cost, of using survey data and scanner data to investigate policy questions regarding food consumption among FSP participants.

1 Appendix A presents brief summaries of some of the major surveys providing data on food consumption.

2 "Feasibility Study of Capturing Food Data at Checkout," FNS contract #53-3198-6-029.

Consumer Card Marketing, Inc. (CCMI), a firm with extensive experience in collecting and analyzing scanner data from retail stores, is a subcontractor to Abt Associates for this study.

As originally envisioned, the feasibility study was to include two phases. Approximately one dozen FSP retailers were to be recruited in the first phase to test the basic concept of collecting scanner data, merging in EBT transaction data and household demographic data, and analyzing the data. Additional stores were to be recruited in the second phase to increase the representativeness and validity of the captured scanner data and to gather more information on data collection costs in a variety of retail environments. For a number of reasons, including financial constraints, the second phase of the study has been dropped from FNS' current research agenda. This current report therefore represents the final report of the feasibility study.

1.1 Context for the Study

Prior to the introduction of EBT systems in 1984, FNS had very limited knowledge about how food stamp recipients used their program benefits. The agency knew where recipients shopped through its tracking of stores' Redemption Certificates—forms that stores fill out when depositing food stamp coupons at their financial depository institutions. (Thus, for instance, FNS knew that about 75 percent of all benefits are redeemed at supermarkets.³) The agency also knew something about when recipients spent their benefits, based on the dates that retailers deposited their food stamp coupons. This redemption information was only approximate, however, because not all stores deposit their coupons on a daily basis. Furthermore, it was not feasible to match dates of coupon use to specific households, and thereby to when the coupons were issued.

With the advent of EBT, FNS' knowledge about recipients' benefit use increased dramatically. All food stamp EBT purchases are recorded on the EBT vendor's transaction log, and therefore it is now possible to identify exactly when and where FSP redemption occurs.⁴ Furthermore, information about the exact dollar amount of each food stamp purchase became available for the first time. Most importantly, however, the agency could now match individual food stamp households to individual purchases because each EBT card has a unique card number that is recorded on the transaction log. This matching ability enables research into the purchasing behavior of subgroups of food stamp recipients, as in a study of Maryland food stamp recipients' access to food stores.⁵

Even with this detailed EBT information about every food stamp purchase, however, the agency knew little about the types of foods that clients purchase with their program benefits. Survey data were

3 In September 1986, supermarkets accounted for almost 74 percent of retail redemptions. USDA, FNS, "State Tables of Activity Ranking, Plus," April 1987. More recent information compiled at FNS indicates that this percentage figure has not changed much in recent years.

4 Thus, for example, the evaluation of the demonstration EBT system in Reading, Pennsylvania showed that most food stamp purchases occur between 4:00 and 6:00 pm, a pattern that grocers report is similar to that for all store customers. See Susan H. Bartlett and Margaret M. Hart, "Food Stamp Recipients' Patterns of Benefit Redemption," Cambridge, MA: Abt Associates Inc., May 13, 1987, p. 22.

5 Nancy Cole, "Evaluation of the Expanded EBT Demonstration in Maryland: Food Store Access and Its Impact On the Shopping Behavior of Food Stamp Households," Cambridge, MA: Abt Associates Inc., July 1996.

available, and some of these surveys collected fairly extensive information about food consumption and purchase. Studies based on survey data, however, have a number of drawbacks, including the expense of collecting the survey data, sampling error, response bias (resulting from sampled respondents who cannot be located or who refuse to participate in the survey), errors in respondent recall of what was purchased or consumed, and deliberate misinformation (as when a respondent says he or she has purchased more fruits and vegetables and fewer snacks and cookies than actually purchased).⁶ The agency therefore began looking for a way to capture information at the point of sale about the composition of food baskets purchased with food stamp benefits.

1.2 Associated WIC Efforts

FNS' efforts to link EBT data to detailed information about purchased food items have not been limited to the FSP. In 1991, the State of Wyoming, with support from FNS, pilot tested in one county an EBT system for the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC). In 1993, the Wyoming EBT WIC demonstration expanded to six counties, including one county where FSP benefits are delivered through the EBT system as well.⁷ Wyoming's EBT system, which stores information about WIC and FSP benefits on the card itself, is now expanding statewide. In addition, several other states are preparing to implement EBT systems that can deliver WIC as well as FSP benefits.

The WIC and FSP programs differ in one fundamental way that is important for EBT systems and the ability to identify what participants are purchasing with their program benefits. FSP benefits are issued for a specified dollar amount and may be used only to purchase program-eligible food items, but there is no restriction on which of the tens of thousands of eligible food items participants actually purchase. In contrast, WIC benefits represent a prescription for a specified list and quantity of WIC-approved food items, and the total number of approved food items is relatively small (e.g., about 500 in Wyoming). Due to this programmatic difference, Wyoming and FNS faced a unique problem when designing an EBT system for WIC—the EBT system would need to identify exactly what was being purchased with WIC benefits and compare each item to the participant's food prescription. This was accomplished by storing the food prescription in the memory of the participant's EBT card, a "smartcard" containing an integrated circuit memory chip, and equipping the EBT terminal with a scanner to read the bar codes affixed to food packages.⁸ Thus, the WIC Program has demonstrated that it is indeed possible to link EBT transaction data with detailed information on items purchased.

This pioneering effort by the WIC Program, however, requires a specialized EBT system using smartcards that operates quite differently from most food stamp EBT systems. Although EBT systems

6 A similar problem may exist even when surveys ask respondents to keep register receipts for food purchased during the week. Knowing that his or her food purchase decisions will be examined in a government study, the respondent may modify his or her shopping patterns that week.

7 For a description of the Wyoming demonstration EBT system, see William Hamilton *et al.*, "Evaluation of the Wyoming EBT System for WIC and the Food Stamp Program: Costs and Impacts of the Wyoming Smartcard EBT System," Cambridge, MA., Abt Associates Inc., May 1997.

8 For stores with existing scanning systems, the EBT terminal is connected to the store's system to access an item's bar-code information.

for both programs reimburse retailers for purchased food items in a similar manner, the automatic check of item identifiers is not required by the FSP, and it would be quite difficult and expensive to implement for several reasons. First, whereas the total number of WIC-approved items is relatively small, the number of FSP-eligible items probably exceeds 100,000. Second, there are approximately four times as many FSP-authorized food retailers as WIC retailers. Finally, whereas most WIC retailers are larger grocery stores or supermarkets with existing scanner systems, the majority of FSP-authorized stores do not have scanners. With a technical solution like that used in the WIC Program not feasible for food stamps, FNS is using this current study to examine the feasibility of a different approach to linking scanner and EBT transaction data.

1.3 Report Organization

This report is divided into five chapters, plus several technical appendices. Chapter Two describes the process of selecting a study site and recruiting retailers to participate in the study. The challenging tasks of configuring and installing the DCSUs and collecting data from all eleven participating stores are the subjects of Chapter Three. Chapter Four describes the collected scanner data and the data processing steps required to create final analysis files for the study. Chapter Five presents examples of the types of analyses that are possible with the collected data. Finally, the report ends with a discussion of what has been learned from this study that would facilitate future efforts to collect detailed information on what food stamp recipients are purchasing with their program benefits. Chapter Six also discusses areas of potential research using scanner data.

Chapter 2

Selecting a Study Site and Retailer Recruitment

The project faced a series of “recruitment” tasks during the search for an appropriate study site. We first contacted EBT directors in all states whose EBT systems were in operation by early 1997. We continued site exploration activities in a state only if the EBT director indicated his or her support for the project. Second, while investigating numerous sites, CCMI contacted retailers operating in the sites to ascertain the likelihood of their participation if the selected study site included one or more of their stores. Third, once the likely study site had been identified, CCMI sought the retailers’ final agreement to provide data for the study. As it turned out, none of these tasks proved easy. This chapter documents the difficulties faced in eliciting participation in the study.

2.1 Identifying EBT States

One basic site selection criterion was that the site’s FSP recipients had to be receiving their food stamp benefits through an EBT system. Only in an EBT site could the study ultimately attempt to match recipient characteristics to purchased food items. Consequently, the search for a suitable study site began with an inventory of the EBT status of all states.

When our search began in December 1996, 23 states had EBT systems either in operation or about to be implemented. These states are listed in Exhibit 2-1. We sent an introductory letter to the EBT directors in all of the states except one.¹ The letter explained the study, said we were beginning the process of searching for a suitable study site, and asked for an extract of the state’s food stamp authorization file. The plan was to use information from the file about each state’s food stamp caseload, including which data elements were available for use in later analyses, to help select the final site.

Exhibit 2-1

States with EBT Systems in December 1996

Alabama	Iowa	Missouri	Ohio	Texas
Arkansas	Kansas	New Jersey	Oklahoma	Utah
Colorado	Maryland	New Mexico	Pennsylvania	Wyoming
Connecticut	Massachusetts	New York	South Carolina	
Illinois	Minnesota	North Dakota	South Dakota	

1 Only Montgomery County in Ohio was using EBT at the time, and the county was too large for consideration as a study site. We therefore did not contact the EBT director in Ohio. Furthermore, North and South Dakota had teamed together for implementation of a single EBT system, so we contacted the EBT director for the joint system.

Of all the states contacted, only six were selected for detailed examination of potential study sites within the state. To avoid collecting data in stores where EBT system problems might cause difficulties, we did not want to select a site where EBT had just recently been implemented. This eliminated nine states from consideration. Several other states were eliminated for various, site-specific reasons. For instance, recipient participation in one state's EBT system was voluntary (Iowa), and in three other states (Illinois, Minnesota, and New Jersey) all the counties that had been converted to EBT were too large for the feasibility study.

In several states, however, the EBT directors simply did not want to participate in the study unless participation was mandated by FNS. Reasons varied. Certainly, state efforts to comply with welfare reform was one factor; the EBT directors voiced concern about their staff not having time to work with us to create an extract of the food stamp authorization file. One director refused, however, because he had specifically promised concerned client groups that the new EBT system would not be used to collect information on purchases. Two other EBT directors refused for similar reasons concerning client privacy.

The number of states with potential study sites was therefore limited to six: Colorado, Kansas, Maryland, New Mexico, South Carolina, and Wyoming.

To summarize, the three main reasons why the remaining states were not selected were:

- The state's EBT system would not be implemented in time for the feasibility study, or it would be operating only in counties that would not be suitable for the study.
- The state EBT directors believed that their staff was too busy with welfare reform to provide any assistance to the study.
- Due to either privacy concerns or general philosophy, the EBT directors were opposed to any attempts to conduct research into how food stamp recipients spend their food stamp benefits.

It is worth noting that, if we were to attempt this study again today, the first factor listed above would be much less limiting; a large number of states implemented EBT systems in 1997. The second factor might still be an issue, although probably less of an issue now than before, because states have a better understanding of how welfare reform is being implemented. The third factor of philosophical opposition would still be an issue in at least some states. If FNS were to attempt to collect scanner data on food stamp purchases nationwide, resistance at the state level would certainly be encountered in some states. There is clearly a belief among some state officials (and, presumably, their constituencies) that collecting information on what recipients buy with their benefits is an invasion of privacy.

2.2 Identifying Potential Study Sites

Within the six states, we used several criteria to narrow the number of potential sites for detailed investigation. First and foremost, we were looking for a site that represented a fairly self-contained shopping area. Although it would not hurt the analysis if recipients from outside the site came in to buy groceries (because they could be identified and excluded from analysis), information on what recipients in the site were buying would be lost if they traveled elsewhere to shop.

Second, we wanted a site with a “medium” number of FSP-authorized stores, somewhere in the range of 30 to 70 stores, to provide diversity in retail environments without exceeding the project’s resources for recruiting stores and collecting data. We also did not want a site with a “dominant” retailer, because that would have reduced the diversity we sought.

Third, we wanted a site with at least 2,000 food stamp households so the analysis would have enough data to examine how shopping decisions vary across different demographic subgroups. There was no specified upper limit on caseload size because the analysis could always be conducted on a sample of recipients. Given the limit on the number of stores, however, it turned out that the largest site selected for further investigation had about 5,700 food stamp households.

Fourth, we were looking for a site with both rural and urbanized areas. We did not want a site that was “too” rural, because we wanted a more heterogeneous group of recipients than typically found in rural areas. We knew, however, that a highly-urbanized site would not match the two criteria of a limited number of stores and a self-contained shopping area. A major metropolitan area would be too big. Although we did consider the possibility that a suitable self-contained shopping area within a metropolitan area might exist, we decided against looking for such an area because of problems of identifying such areas and gathering information at this scale.

Fifth, to the extent possible, we wanted a site where the food stamp caseload represented a demographically diverse group of recipients.

Largely because many pertinent data were easily available at the county level, we decided to use counties as our geographic unit when looking for a study site. Although we recognized that we might modify this decision after a site had been selected (e.g., by adding cross-border stores to the study or by excluding stores in a remote area within the county), using counties as the geographic unit greatly eased the problem of collecting information and identifying the most suitable study sites.

For each county using EBT in the six states, we compiled the following information to aid in our identification of potential study sites:

- Information on county population (1990 Census data)
 - number of residents
 - percent of households with income below the federal poverty level
 - percent of households in different race and ethnicity categories
 - percent of population living in “urbanized” census tracts (as defined by the U.S. Census Bureau)
- Information on FSP-authorized retailers (FNS STARS data)²
 - number of retailers, by store type

2 FNS provided what are known as “EBT-IV” files from the Store Tracking and Redemptions Subsystem (STARS) of the Food Stamp Program Integrated Information System. The six files, one for each state, were provided in February 1997.

- concentration of redemptions in large stores
- number and names of supermarket and convenience store chains represented
- percent of FSP benefits redeemed in convenience stores
- Information on FSP households (state authorization data)
 - number of households
 - percent of FSP benefits spent within county³
- Distance to nearest major markets in adjacent counties or states

Based on the above information, we identified 17 counties in five states as potential sites for the study (Exhibit 2-2). Colorado was dropped from the list because EBT had been implemented in only three counties at this point in our search effort. Of the three, one was too small, one was too big, and the third was judged to be too close to a major metropolitan area to serve as a self-contained shopping area. In the other five states, most of the counties that did not make the list were either too small in terms of caseload size or number of retailers, or the county was too close to a major metropolitan area.

2.3 Determining Retailers’ Willingness to Participate in the Study

The next step was to examine the characteristics of the program-authorized stores in the 17 counties and to ascertain the willingness of county retailers to participate in the study. In stores with existing scanning systems, participation meant that the store would be willing to provide the study with confidential scanner data on items purchased by food stamp customers and others. For stores without scanning systems, participation would require that store employees use a stand-alone “data collection scanning unit” (DCSU) to capture bar-code information on items purchased.

Using information included in the FNS STARS files, we listed all program-authorized retailers operating in each of the 17 counties. Beginning in March 1997, we performed three “pre-contact” activities. We informed FNS regional and field office staff in writing that we were planning to contact retailers in the 17 counties to determine their potential interest in the feasibility study. We also contacted the EBT directors in the five states to let them know we were about to begin contacting retailers. Finally, we spoke with FNS field office staff to obtain their perspective on the retailer community in each county and to elicit their ideas about which retail communities would be more or less likely to be interested in participating. We also used this opportunity to determine whether any of the 17 retail communities were going to be subject to intensive law enforcement activities during the planned data collection period. (We wanted to avoid such areas.) Fortunately, none of the sites had to be eliminated due to planned major investigative efforts.

3 This information could be tabulated only for counties in Maryland and South Carolina. The Maryland figures were based on EBT transaction data and recipient data collected during a prior study. The South Carolina figures were based on a merge of the state’s FSP authorization file with a May 1997 extract of FNS’ Anti-Fraud Locator of EBT Retailer Transactions (ALERT) subsystem, which contains EBT transaction data provided by states’ EBT vendors.

Exhibit 2-2**Potential Study Sites**

County, State	Major City	Size of FSP Caseload	Number of Retailers
Shawnee County, KS	Topeka	4,000	74
Allegany County, MD	Cumberland	3,549	64
Cecil County, MD	Elkton	2,282	52
Frederick County, MD	Frederick	2,881	67
Chaves County, NM	Roswell	3,929	50
Curry County, NM	Clovis	2,309	41
Eddy County, NM	Carlsbad	2,468	57
Lea County, NM	Hobbs	2,879	54
Santa Fe County, NM	Santa Fe	2,310	55
Beaufort County, SC	Beaufort	2,677	77
Clarendon County, SC	Manning	2,677	77
Colleton County, SC	Jacksonboro	2,136	47
Georgetown County, SC	Georgetown	2,716	58
Greenwood County, SC	Greenwood	1,910	44
Lancaster County, SC	Lancaster	2,128	71
Laurens County, SC	Clinton	2,304	55
Natrona County, WY	Casper	2,156	40

FSP caseload figures from December 1996 or early 1997. Number of retailers from a February 1997 STARS extract.

Project staff at CCMI then took the lead in contacting retailers. On March 26, CCMI began mailing letters of introduction to the headquarters of chain stores operating in the selected sites.⁴ CCMI started with chain stores for two reasons. First, in many sites, a single chain had multiple stores operating within individual counties. In such cases, it would be difficult to obtain widespread retailer participation in the county without the chain's support. Second, many of the chains operated in more than one of the 17 counties. Thus, discovering that chain's level of support (or non-support) for the study could quickly help identify a "short list" of two or three counties with the best potential to meet the study's objectives.

4 See Appendix B for a sample letter.

Determining a chain's potential level of support turned out to be a laborious and time-consuming task. The initial contact letters were typically mailed to the president or chief executive officer of the corporation. CCMI staff then attempted to follow up a week later with a phone call. In many instances, however, the president was well shielded from incoming phone calls, despite the advance letter. Thus, verbal contact often was made only after repeated call-backs spread out over several weeks. In addition, once contact was made, the chain's president often referred CCMI staff to other executive personnel, especially POS or information technology (IT) managers. Typically, this initiated another time-consuming round of placed calls, unanswered voice mail messages, and call-backs.

Because the effort to obtain definitive responses from the chains was taking so long, CCMI continued mailing initial letters to other stores operating in the 17 selected counties. In general, larger stores were targeted first. Because we wanted to test the stand-beside DCSU in non-scanning stores, however, CCMI also sent letters to convenience stores in the belief (which turned out to be correct) that few convenience stores in the sites would have scanners.

Eventually, we had enough information to start narrowing the list of potential study sites, but not until nearly three months had been spent trying to determine stores' willingness to participate. The main lessons learned from this exercise were:

- It was difficult to get the attention and support of chief executive officers for a study of this nature. The fact that this was a USDA-sponsored study made little difference as long as store participation was not mandated.
- Among store chains, support was greatest within supermarkets and weakest within convenience stores. This difference is due, at least in part, to the different levels of scanning experience and knowledge across the two store groups.
- Some stores were simply unwilling to participate voluntarily in any government study. An extreme effect of this was that all counties in one state had to be dropped from consideration because the owners of a large supermarket chain and a large convenience store chain operating throughout the state refused to participate.⁵
- Even for stores that eventually said they would be willing to participate in the study if their county was selected, it was often difficult to obtain this support. As previously mentioned, CEOs would often direct CCMI to a division manager to discuss the proposed project. These managers were often quite busy and, without a strong directive from the CEO, they had little to gain by spending time talking with CCMI about what type of POS systems they had and the technical details of how to transfer data to the study.
- Particularly in the senior support positions (e.g., manager of POS systems), staff turnover seemed to be quite high when CCMI was contacting retailers. CCMI's efforts to follow through on retailer contacts were often delayed when a newly-hired manager said he or she needed some time to get situated before dealing with an outside request for assistance.

⁵ FNS staff in the local field office also warned us that many retailers in the southern part of the state were very leery of government-related activities, and that we would be unlikely to find strong levels of support there. CCMI staff found that this was indeed the case.

- Scanner data are quite valuable to marketers, and some store personnel were reluctant to consider providing such data at cost.⁶ Somewhat surprisingly, this attitude was more often found among corporate division managers than CEOs.

2.4 Final Site Selection and Retailer Recruitment

By May 1997, the project team had narrowed the number of potential study sites to a handful. A number of sites were dropped from consideration either because they were too close to other shopping opportunities, or major retailers within the site were not interested in participating. CCMI staff visited representatives of two large supermarket chains in May, and they visited with a number of retailers in Georgetown County, South Carolina that same month. CCMI staff then visited stores in Topeka, Kansas (Shawnee County) in June. Much of June and July was spent determining likely retailer cooperation in potential sites in South Carolina and Kansas.

During a conference call on July 22, 1997, and based on information provided by Abt Associates and CCMI, FNS selected Georgetown County, South Carolina as the project's study site. Part of the problem with the Shawnee County site in Kansas was that several major chains there were unwilling to commit their support to the project. As shown in Exhibit 2-3, Georgetown County is located on South Carolina's eastern coast, just south of the Myrtle Beach vacation area (which is located in Horry County). The county's population in 1990 was 46,302. With two small population centers—Georgetown and Pawley's Island—the Census Bureau classifies the county as an urban area with no central place. Exhibit 2-4 shows a map of Georgetown County and surrounding areas.

Census data from 1990 indicate that about 43 percent of Georgetown County's residents are African-American, whereas fewer than 1 percent are Hispanic, American Indian, or Asian. Twenty percent of the county's households live below the federal poverty level. Thirty-five percent of residents live in urban census tracts, mostly along the coast. The FSP caseload in Georgetown County was 2,574 households in July 1997. These households spent approximately 82 percent of their monthly FSP benefits within the county.

Although the number of FSP-authorized food retailers in any location is constantly changing as new stores open, older stores close, and other stores change ownership, there were approximately 53 FSP-authorized food retailers in Georgetown County in March 1997. This retail community included 11 supermarkets, five grocery stores, five specialty stores, 22 convenience stores, and ten stores with other classifications (including six combination grocery/gas stations.)⁷ Together, these stores were redeeming approximately \$430,000 in food stamp benefits per month. Nearly 82 percent of the redemptions were

6 As described later, the study did not have the resources to purchase scanner data at what might be considered their market value. Instead, participating retailers were reimbursed for their expected cost of providing the data to CCMI on a weekly basis throughout the data collection period.

7 Store counts and redemption data are based on EBT transaction data from March 1997, as documented in that month's FNS ALERT file for South Carolina.

Exhibit 2-3

South Carolina Counties with Georgetown Study Site

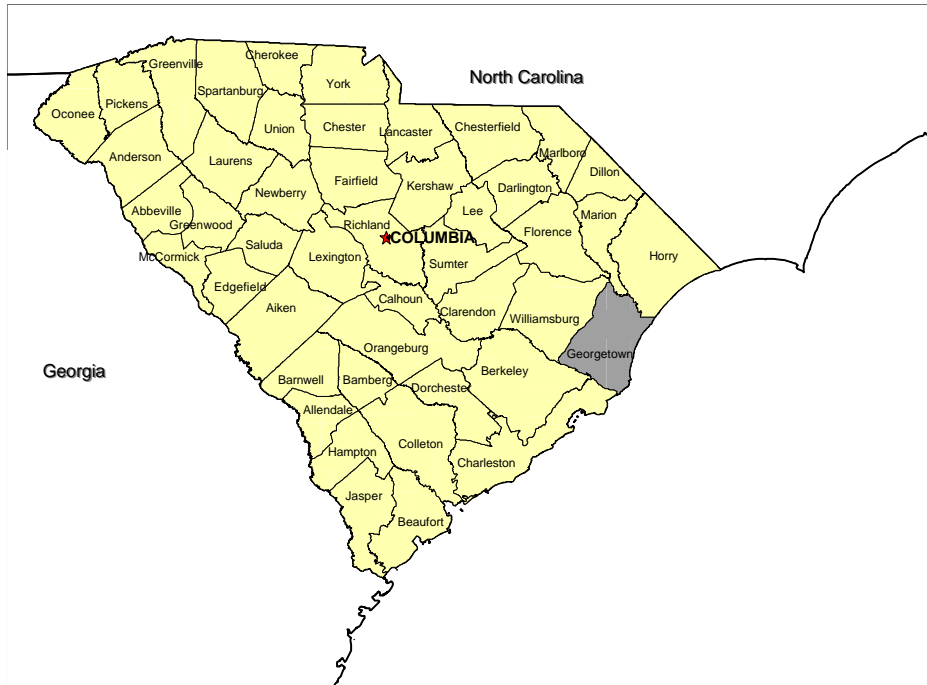
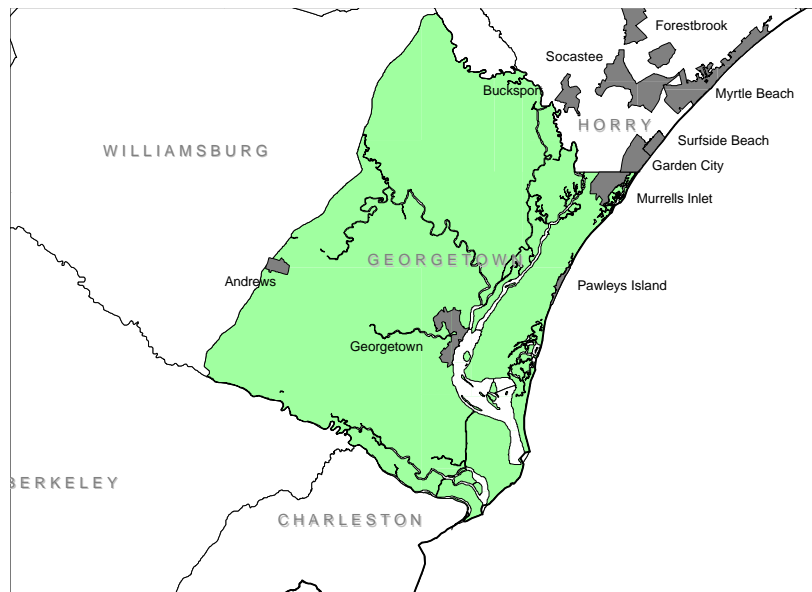


Exhibit 2-4

Georgetown County, South Carolina



occurring in supermarkets, 4.5 percent in grocery stores, 3.5 percent in specialty stores, and 2.9 percent in convenience stores. A number of major supermarket chains are represented in the county. There were also several convenience store chains in the county, but most food stamp households were shopping at independent convenience stores.

Retailer recruitment efforts in Georgetown County began in March 1997, when CCMI sent letters to the corporate headquarters of all the major chains operating in the 17 counties then under consideration. Six major supermarket chains were operating in Georgetown County. Shortly thereafter, letters were sent to those grocery, specialty, and convenience stores in the 17 counties (including Georgetown County) that were redeeming reasonable amounts of food stamp benefits.⁸ CCMI then followed up with phone calls to most of these chains and stores to determine initial levels of support for the study.

In May 1997, CCMI visited with headquarters staff of one of the supermarket chains and obtained their general support for the study. At the end of May, CCMI staff met with a representative of the South Carolina Department of Social Services and visited a number of stores in Georgetown County. At that time, four small non-scanning stores in the county agreed to participate in the study using the DCSUs.

After recruiting these four non-scanning stores, efforts concentrated on the supermarket chains. From the very beginning, two of the six chains operating in Georgetown County said they were not interested in participating in the study. Having already obtained support from one chain, CCMI therefore focused its efforts on the remaining three chains, which had nine stores located in or very near Georgetown County.

Recruiting efforts were not easy. As noted earlier, senior executives often turned over CCMI's request for participation to a POS manager or an information technology (IT) manager without committing the chain to participation. CCMI would then need to gain the confidence of a mid-level manager who was already busy and had little or nothing to gain from participation. Retailers seldom returned calls, so recruiting efforts continued right through the end of August.

By mid- to late-August 1997, eleven stores had agreed to participate in the study. The four non-scanning stores included two convenience stores, one grocery store, and one combination grocery store/gas station. Of the remaining seven stores, two represented one supermarket chain and five represented another. Exhibit 2-5 presents summary information about the stores. Some of the supermarkets were located outside Georgetown County, but near enough to be a shopping destination for food stamp recipients in the county. Together, these eleven stores processed over 6,000 EBT transactions initiated by county recipients in March 1997. These transactions represented 34 percent of all benefits redeemed by Georgetown County recipients during the month.

8 Originally, CCMI and Abt Associates defined "reasonable" in terms of at least 100 expected EBT transactions during a month. (The expected number of transactions was calculated using actual STARS redemption data from March 1997 and results from a prior study in Maryland that computed average EBT transaction amounts by store type.) Once the number of potential sites was narrowed, letters were sent to many stores processing fewer than 100 EBT transactions per month.

Exhibit 2-5

Participating Stores

Chain A	Major regional supermarket chain with almost 1,200 stores throughout the sunbelt states. Total annual sales exceed \$13 billion. Two stores located in or near Georgetown County. The chain stocks nearly 80,000 different items with unique UPC or PLU bar codes.
Chain B	Major regional supermarket chain with more than 1,200 stores in southeast and mid-atlantic states. Total annual sales exceed \$10 billion. Chain B has five stores located in or near Georgetown County.
Store C	This is a 24-hour combination grocery store/gas station, which is often very busy. It is located at the junction of several major highways and is an all-service type of store, with lots of fishing equipment and bait as well as gas, groceries, beer, and ice. The store has two checkout lanes and stocks about 2,300 items with different UPC codes.
Store D	This store is a busy grocery store. It was the only “grocery” in the area, being located 6-7 miles outside of Georgetown proper. The store also has a full delicatessen and gas pumps. The store, which is open until 8:00 p.m., stocks about 3,200 items with UPC codes.
Store E	This is a small convenience store located about 10 miles inland from Georgetown. The store hours are from 8:00 am to 9:00 pm. When the owner/operator needs to go home or elsewhere, she either closes the store temporarily or gets a family member or friend to substitute. The store stocks about 2,050 items with UPC codes.
Store F	This is a small convenience store located close to Store E. The store sells gasoline, and it is the only non-supermarket in the study that sells liquor. The hours of operation are from 8:00 am to 9:00 pm. Although there are two cash registers in this store, only one DCSU was installed. The second cash register serves mainly as a backup unit. The store stocks about 1,960 items with UPC codes.

2.5 Retailer Agreements

From the beginning of the project, the plan was to reimburse retailers for the cost and effort they would incur to participate in the study. To do so, each retailer and Abt Associates needed to enter into a “Data Collection Agreement” outlining what data the retailer would provide and the agreed-upon level of reimbursement. In July, therefore, CCMI and Abt associates began preparing three draft agreements. The first was for stores with scanning capabilities who were capturing and storing bar-code information at the transaction level. No hardware or software upgrades were needed to collect data for the study in what we called these “Category I” stores. The second draft agreement was for “Category II” stores, or stores that were scanning but not capturing bar-code information on a per-transaction basis. Hardware or software upgrades would be needed before such stores could supply data for the study. The final draft

agreement was for “Category IV” stores—those that did not use scanning equipment and had no current plans to invest in scanning equipment.⁹

During CCMI’s trip to Georgetown County in late May, CCMI staff told the non-scanning stores they visited that the project would reimburse them \$2,500 if they participated. Given that these stores would need to use the portable DCSUs for six to seven weeks, during which time store clerks would need to both scan purchased items at the DCSU **and** ring up purchase prices on their regular cash registers, the \$2,500 figure seemed justified. CCMI reported that the \$2,500 reimbursement did seem to induce several stores to participate.

In its conversations with the supermarket chains operating in Georgetown County, CCMI sought information about their expected level of effort to participate. Store personnel were largely unable to provide an estimate of expected costs. For the two chains that agreed to participate, the agreed-upon reimbursement was \$1,000 per chain (not per store). Their effort was largely limited to drawing an extract of their transaction log file once a week and sending the file to CCMI.¹⁰

Samples of the final data collection agreements are included as Appendix C. The agreements contain language assuring the stores that the collected data will be used for research purposes, not compliance efforts, and that the data will remain confidential. Abt Associates and CCMI included this language at the beginning of the process. Generally, the participating stores did not seem to worry too much about specific assurances in the agreement. The major exception is one of the two supermarket chains, which declined to provide data on non-FSP purchases (i.e., those paid for with resources other than EBT food stamp benefits) and on non-FSP eligible items in FSP purchases (which are present in purchases involving both EBT and cash tender). The chain cited issues of confidentiality as the reason for not providing these data.¹¹

9 CCMI had determined at this point that none of the potential participants in the study represented “Category III” stores, those without scanning equipment but actively considering investing substantial resources to install and use a scanning system.

10 CCMI provided blank tapes and pre-paid mailers to the supermarket chains.

11 It is worth noting that, to protect the confidentiality of these data, the supermarket chain had to do extra programming and data processing to pull the create extract files for the study. It would have been easier and cheaper for the chain to provide all its data to CCMI during the data collection period.

Chapter 3

Equipping Retailers and Collecting Data

In order to collect UPC-level data from the four stores without scanning systems, we needed to install a scanning device and a means to associate the item's price with the item. We therefore developed what we called a "data collection scanning unit," or DCSU, to be installed at the checkout counters in the four non-scanning stores. This chapter describes the DCSU, what procedures were followed to install the DCSUs and train the employees, and the process of collecting data from both the non-scanning and scanning stores. The original plan called for the data collection period to extend from September 1 through October 10, 1997.

At the outset of the study, we were prepared to work with one or two stores that were scanning but would need either a hardware or software upgrade to capture UPC data on a transaction-by-transaction basis. None of the eleven stores who participated in the study fell into this category.

3.1 Data Collection Scanning Unit

The planned use of DCSUs posed three unique problems not faced in stores with scanning units, namely:

- (1) The store would not have an existing price look-up file, matching UPC or other bar codes to an item's price.
- (2) Using the DCSU would require the store clerk to perform additional activities at the checkout, so the DCSU had to be easy to use.
- (3) There would be very limited counter space available to place the DCSU, so it had to be small.

Numerous technical "solutions" were considered in an attempt to minimize the impact of these problems. To address the first issue, CCMI staff decided they would build a price look-up file (also called a "UPC master file") for each of the four non-scanning stores prior to the start of the data collection period.

To capture the customer purchase information, the store clerk would be required to scan all items into the DCSU prior to entering each item again into the store's cash register system. For non-UPC items, the clerk would be asked to press an appropriate department key on the DCSU. At the end of the transaction, the clerk would press the appropriate tender key on the DCSU and enter the dollar amount of the transaction. Information on all scanned and key-entered items would be retained in the DCSU database.

To record these transactions and to overcome the counter space limitations, the recommendation for non-scanning stores was to offer to install a small, portable (and hence retrievable) stand-alone POS system. The components of the DCSU, described below, are shown in Exhibit 3-1.

Exhibit 3-1

Data Collection Scanning Unit (DCSU)



CCMI first selected the scanning device to be deployed with the DCSU, a Spectra-Physics VS1000 scanner shown at the right side of the exhibit. Spectra-Physics is a market leader in supplying small-footprint scanning devices. The VS1000 scanner also offered a multi-directional scanning ability, which simplified its ease of use.¹ Other benefits of the VS100 included a small footprint and the ability to scan multiple bar-code formats. In addition, the Spectra-Physics VS1000 was at the low end of the price range for fixed (as opposed to hand-held) scanners.

CCMI then turned to selection of the POS terminal for the DCSU, and selected a PC register (the PowerRegister 5480 system) manufactured by Riva. The criteria for the hardware selection were:

- ease of use
- small footprint
- portability
- compatibility with the Spectra-Physics VS1000
- competitive pricing
- MS-DOS operating system
- high storage capacity

¹ Multi-directional scanning allows the store clerk to scan an item regardless of its orientation to the scanning unit. In contrast, hand-held scanning units offer only a single-scan beam, which will work only in a parallel product orientation. Thus, a multi-directional scanner has a faster product throughput than a single-scan beam (i.e., there is less need for repeated waving of the item's bar code in front of the scanner).

The Riva 5480 system met the above functional criteria and was competitively priced. Characteristics of the Riva that were particularly attractive were its small size, portability, “open architecture,” and expandability.² Because the DCSUs were to be installed for only a short period, CCMI wanted a unit that was easily portable and would not require special wiring or mounting to a stand or table. The Riva 5480 was also expandable to a full POS cash register system. The latter was important because, if the feasibility study demonstrated the ability to collect and analyze UPC data, CCMI wanted the flexibility of totally **replacing** a store’s current cash register system with an expanded DCSU if future data collection efforts were to be pursued. This would eliminate the need for two separate registers at the checkout lane.

Finally, CCMI decided to use a POS software package developed by Multi-Link. The Multi-Link POS software is easy to use and is currently in operation in both grocery and convenience stores. Furthermore, Multi-Link staff agreed to assist CCMI in modifying their software package to meet the needs of the study.

The most efficient method for building a price file for the four non-scanning stores was to program a hand-held scanning device to allow CCMI staff to scan the bar-code label of a store item into the device, and then to key-enter the item description (e.g., Coke) and unit size (e.g., 12-oz can) and the item’s price. The resulting UPC master file could then be downloaded into a laptop computer, where it would be transferred to the retailer’s DCSU.

Hand-held scanning devices by such manufacturers as Compsee, Telxon, Intermac, and Symbol are commonly used in the retail environment. For purposes of this study, CCMI selected the Compsee APEX II scanning unit for the following reasons:

- It could be programmed in BASIC, which was easier than programming in a proprietary language;
- The unit’s database had a high storage capacity;
- The unit was quite durable (the unit survived a four-foot drop to a concrete floor during a demonstration);
- Its database could be transferred easily to a laptop computer; and
- It was in the middle of the price range for similar devices.

The Compsee scanning unit, with its battery-charging stand, is shown in Exhibit 3-2.

3.2 Equipment Purchase and Configuration

Abt Associates purchased one Riva PC register (with monitor) and a Spectra-Physics VS1000 bar-code scanner in March 1997 so CCMI and its software vendor, Multi-Link, could begin modifying and testing software for the unit. The cost for these two pieces of equipment, including an introductory discount on the scanner, was \$3,090.

2 Open architecture means that software and devices from different manufacturers can be added to the unit.

Exhibit 3-2

Compsee Scanning Unit



In late July 1997, once Georgetown County, South Carolina, had been selected and it was known that four stores (with five cash registers in use) would be needing DCSUs, CCMI ordered six more Riva units, monitors, and scanners. With five units to be placed in the field, this provided two backup units, one for software maintenance and one replacement unit in the event of equipment failure.

For this study, CCMI obtained the DCSU system components at the VAR (value-added reseller) pricing level shown in Exhibit 3-3. Total cost per complete DCSU was \$4,027. End-user retail pricing would add between \$800 and \$1,800 depending upon the standard VAR markup ranges of 20 to 45 percent. Such a configuration would retail for about \$4,800 to \$5,800 per lane (at mid-1997 prices).

In addition, CCMI ordered two Compsee APEX II units at a price of \$1,484 per unit. Buying two units enabled CCMI to use a two-person team to initialize each store.

Total cost for all of the above equipment was \$30,220, which does not include shipping.

CCMI had to create a standard set of procedures to set up each DCSU. Each component had its own initialization requirements, and the components had to be integrated within the whole system. The steps needed to achieve this initialization and integration were:

- The Multi-Link POS software had to be modified to meet the needs of the study.
- The modified Multi-Link software had to be loaded into the Riva unit and tested.

Exhibit 3-3**DCSU Components and Price**

Riva 5480 base system	\$1,140
8 mB RAM	50
600 mB disk drive	178
DOS software	37
Color LCD panel	638
Spectra-Physics VS1000 scanner	796
Multi-Link POS software	1,090
Carrying case	98
Total	\$4,027

- The layout of the Riva keyboard had to be modified to meet the needs of the study. In particular, department keys and a “food stamp” key had to be identified and programmed.
- The Spectra-Physics VS-1000 scanner had to be attached to and integrated with the Riva unit.
- The program to load UPC and related data into the Compsee hand-held scanner had to be written, as did the programs to load these data into a laptop and then into the Riva unit itself.

The main modifications needed for the Multi-Link software were an interface to accept price file data from the laptop computer and a change in information to be written to the unit’s transaction log.

3.3 Building the Price File, Equipment Installation, and Training

Original plans called for the DCSUs to be installed in the four participating, non-scanning stores the week of August 18, 1997. One day per store was allocated for building the price file, installing the DCSU(s), and training the employees. CCMI staff would then return the following week to confirm that the units were working and being used properly prior to the planned September 1, 1997 start of the data collection period. The procedure for building the store-specific master price files, however, turned out to be more complicated than expected, as described below. The complication led to a one-week delay in installation, which eliminated the planned one-week trial data collection period prior to September 1.

To build a price file, CCMI intended to use the Compsee units to create a database with one record per unique UPC within a store’s inventory. This would be accomplished by scanning the UPC of a selected item, which would enter the UPC code into the record. CCMI staff would then manually key-enter a short description of the item and its weight or size, using the keyboard on the Compsee unit. Finally, the price of the item would be manually key-entered, completing the record for that item.³ After all items in

3 For items in the inventory without a UPC code, information on item price and department code was manually entered into the DCSU.

the store's inventory had been scanned, the resulting database would be downloaded into a laptop computer. The file would then be transferred to the DCSU.

The above process is clearly very labor intensive, but CCMI planned to shorten the process by pre-loading the Compsee with the UPC codes and item descriptions of the price file from a supermarket. In that way, and to the extent that a non-scanning store's inventory overlapped the supermarket's inventory, item descriptions would not need to be key entered. Instead, CCMI would scan the UPC of a selected item. The Compsee unit would check its database for the same UPC. If found, then the item description would already be on the file. CCMI would then need to enter manually only the price of the item to complete the record. If the scanned UPC was not on the pre-loaded price file, then both the item description and item price would be key entered.

Unfortunately, the Compsee unit's processing speed was too slow to check quickly a large database of stored codes, resulting in a very slow response time. CCMI staff therefore decided to delete the pre-loaded database of stored bar codes and to create a Compsee database just of each store's UPC codes and item prices, thereby avoiding the time-consuming entry of item descriptions. This database would then be downloaded to a laptop and compared to the larger, supermarket bar-code database including item descriptions. For just those items that did not match, CCMI staff would have to go back to the shelves and enter item descriptors.

Not only did this process turn out to be very time-consuming, it revealed problems with the Compsee's scanning software. A number of zero-suppression schemes were encountered that the software could not decode.⁴ CCMI therefore had to postpone the installation process for one week while the Compsee software was fixed.

Installation occurred during the last week of August 1997, when CCMI finished scanning the inventory of the four stores and installed the five DCSUs. Total time to create the four price files was 92 person-hours. The final combined record size of the four non-scanning stores was about 9,000 items, of which fewer than about 5 percent overlapped with the price file of an out-of-state supermarket. There was not even much overlap in inventory from one store to another because the stores used different wholesalers and most items were private label.

Upon installation, CCMI trained store staff in how to use the equipment. Training covered the following functions:

- how to log onto the DCSU
- how to reset the unit after a power interruption
- how to scan bar-coded items
- how to enter department codes for unscannable items
- how to enter price information if not available on the price file
- how to enter payment information (e.g., cash, EBT, other)
- how to reach CCMI for help and assistance.

4 The normal UPC bar code is a 12-digit code that uniquely indicates manufacturer and item. When normal UPC codes contain a string of embedded zeros, manufacturers can shorten the UPC code by using a zero-suppression scheme; this essentially tells the scanner to assign zeros to certain (missing) digits in the printed UPC code.

CCMI staff then watched store clerks as they entered data for several sales, answering any questions that arose.

CCMI decided not to train retailers in how to update the DCSU's stored price file. Instead, retailers were asked to log items for which the scanned price did not match the actual price. CCMI staff would use the log to update the file during weekly visits to the store during the data collection period. CCMI also left a training manual and a toll-free number for retailers to call if they had problems with the equipment.⁵ The toll-free support number connected to CCMI's office in Massachusetts.

3.4 Data Collection at the DCSU Stores

The study planned to collect data during South Carolina's September 1997 disbursement month for food stamp benefits.⁶ With food stamp benefits in South Carolina being issued over the first ten days of each month, this meant that data collection needed to extend from September 1 through October 9 in order to observe at least a full month's shopping for all recipients. As described below, the data collection period ultimately was extended through October 31, 1997 due to data retrieval problems and inconsistent use of the DCSUs by stores in September.

Once data collection began, CCMI staff traveled to Georgetown County on a weekly basis to download scanner data from the DCSUs to a laptop computer. CCMI staff traveled to South Carolina on September 5, 11, and 18 to collect data from the DCSUs. This involved running a "data capture" routine on the DCSU that processed the prior week's raw transaction data and created an output file for transfer to a laptop. At the same time the routine transferred the raw data into a backup file, leaving room for the next week's raw transaction data.

When the DCSU data were captured on September 5, CCMI realized that data in one field were being truncated during processing; only the first three digits (always "DEP") of the seven-digit department code were being retained. CCMI modified the data capture routine the following week and installed the modified software during its September 11 visit to the Georgetown County stores. At this time all raw data for the period September 5 to 11 were transferred to a backup file, overwriting raw data captured prior to September 5.⁷ The modified software, however, contained an error. Although it fixed the department code truncation problem, the new department code information was overwriting the item's UPC information, resulting in a loss of critical data.

This loss of UPC data was not identified until CCMI closely examined the data brought back from the September 18 data capture. CCMI returned to South Carolina on September 25 to copy the DCSU's backup files, which contained raw transaction data for the period September 11 to 18. CCMI then used

5 A copy of the training manual is presented in Appendix D.

6 The data collection period was originally planned to start on August 1, 1997. Due to late responses by the retailers to our inquires about possible participation, the data collection start date was pushed back one month.

7 This overwriting was planned.

these files to modify and test the data capture routine. The revised data capture routine was installed the week of September 29.

These software problems led to the following status of data captured by the DCSUs:

September 1-5	Processed transaction file; department code truncated but recoverable, UPC code overwritten and lost
September 5-11	Processed transaction file; department code ok, UPC code overwritten and lost
September 11-18	Raw backup file, from which department code and UPC code were recovered
September 18-25	Raw transaction file, from which department code and UPC were recovered
Sept. 25-Oct. 9	Processed transaction file; department code OK, UPC code OK

By the end of September, it was also clear that the four DCSU stores were having problems using the DCSUs. One store manager was clearly computer “phobic;” UPC data in her store were being collected only when a part-time sales cashier was in the store. Other problems, described below, were more technical. For instance, the DCSUs were initially set with a timeout period of from one to five minutes. After the end of the period, the unit would shut down and the cashier would be required to re-enter his or her user name and password. This was such a nuisance that all the participants complained; in response, CCMI set all timeout periods to 255 minutes, the maximum allowed by the DCSU software. Another software problem was related to the sign-on process. After entering one’s name and password, the system would blank and return to the beginning screen, again asking for the password. The unit, however, would not accept the password the second time, causing one store not to use the system for the entire first week of September. When learning that stores were having problems with this, CCMI instructed them to press the CLEAR button on the DCSU keyboard to proceed. The problem was corrected by downloading new software during the September 18 visit.

Another inconvenience of the DCSU was that, during initial setup, it needed to be programmed to read UPCs correctly. This “programming” was done with a series of flash cards imprinted with bar codes. After sequentially scanning the bar codes, the unit was programmed. CCMI learned early in September, however, that the scanner lost this programming any time power to the unit was shut off. CCMI then provided each store with a set of flash cards and trained the owners and employees in how to program the unit upon start-up. The procedure was simple and not time-consuming, but it was still a nuisance.

By and large, however, the main problem was the need for cashiers to process all transactions twice—first on the DCSU and then on their regular cash register. After the data were examined it became evident that the DCSUs were simply not being used in many instances. Furthermore, when they were being used, cashiers appeared to be bypassing the scanning step oftentimes and simply entering the department code and price of the item being purchased. This inability or unwillingness to follow proper procedures was not universal, but it occurred often enough to conclude that this approach to collecting UPC data in non-scanning stores is not very reliable. This topic is addressed again in Chapter 6.

At the end of September, however, the project team did not know that the non-scanning stores were not using the DCSUs properly or consistently. All that was known was that, due to software problems, critical UPC data had not been collected. CCMI therefore proposed extending the data collection period for another month. This was initially deemed as potentially too burdensome on the participating stores; we did not want to deter their willingness to participate in a possible future data collection effort involving more stores in Georgetown County. Nevertheless, after discussions with FNS, CCMI was instructed to determine the stores' potential willingness to continue collecting data through the end of October without an increase in reimbursement. All four DCSU stores and the two supermarket chains agreed to this extension, largely because there was little burden on the supermarkets and the DCSU stores realized that they had not been consistent in scanning purchased items. The stores also recognized that, without an extension, the corrupted data from September would diminish the value of the study. FNS then approved an extension of the data collection period through the end of October.

After October 9, CCMI staff made three additional trips to Georgetown County to retrieve the weekly data. Upon extending the data collection through October 31, there were no additional hardware or software problems except at one DCSU store. When moving the DCSU to make room for a meat locker, the store damaged the unit and was unable to scan data during the last four days of the month.⁸ During the last site visit on November 3, CCMI staff collected the weekly data file from each store and removed the DCSUs.

3.5 Data Collection at the Supermarket Chains

Data collection at the two supermarket chains was more straightforward, but it too encountered some problems. The plan was to have headquarters staff at each chain copy the contents of their transaction logs (T-logs) for the participating stores to an extract file each week and to send that extract file to CCMI for processing. Prior to the start of data collection, the two chains would supply a sample data tape containing one week's worth of transaction data for initializing and test purposes. In addition, the stores were to supply an extract of their UPC master table.

Both store chains had IBM 4690 POS systems. Each chain's corporate headquarters retrieved daily transaction logs from each of the participating stores. The transaction files would be sent to CCMI on a weekly basis for processing. CCMI provided 4-mm DAT tapes and FedEx envelopes for each week and one spare for the expected weekly data.

Complications started showing up before data collection started. One of the chains nearly dropped out of the study due to their concern of providing proprietary data. The chain finally agreed to pre-processing their IBM 4690 T-logs to remove all non-food stamp items as well as all transactions that did not involve a food stamp payment.

As shown in Exhibit 3-4, problems also occurred with receipt of data. Chain A did not send data for October 13, 1997 for either of its two stores. After CCMI requested data for this date, Chain A responded that the data were lost. In addition, the process at Chain A was apparently a low priority one

⁸ The DCSU was permanently damaged when store staff accidentally tried to plug the power cord back into the unit's keyboard slot, bending a critical pin and short-circuiting the port.

due to resource constraints (hardware, software, and personnel). CCMI went over a month without receiving any data from Chain A, and then received three weeks of data at one time. These data were mislabeled, duplicative and, as mentioned, missing for one day in each store.

Upon reviewing the data provided by Chain B, CCMI found that most data were missing for several days, as shown in the exhibit. Chain B explained that the missing data for all stores on October 18 was due to disk drive problems in the chain’s communications server. No data from any of the chain’s stores were captured that day. The data from store B-1 on October 19 were lost due to a file being deleted mistakenly. The remaining data were missing due to unspecified communications problems between the affected stores and chain headquarters.

Despite these problems, nearly all data from the supermarkets chains were forwarded to CCMI by a week or so after the end of the data collection period.

Exhibit 3-4

Missing Data at Supermarkets

Store	Dates with Missing Data
A-1	October 13
A-2	October 13
B-1	October 18, 19, 23, 24, 25
B-2	October 18
B-3	October 18
B-4	October 10, 11, 18
B-5	October 18

Chapter 4

Processing Scanner Data

Scanner data hold the potential for exciting Food Stamp Program research, primarily because scanner data identify each purchased food item using Universal Product Codes (UPCs) that are uniform across retailers or price look-up codes (PLUs) that identify items sold in variable-weight packaging. A goal of this study was to obtain scanner data consisting of all data items appearing on a typical cash register receipt, from a sample of FSP retailers, and to analyze the food purchasing patterns of FSP households. The information items appearing on a typical cash register receipt are:

- a store identifier
- date and time of purchase
- subtotaled amount of purchase
- tax to be paid
- total due
- tender amounts, by tender type (e.g., cash, check, coupon, EBT, debit or credit)
- such food item specific information as:
 - item description¹
 - item price
 - item quantity
 - total expenditure on item

At the start of this project we identified three main challenges in processing scanner data:

1. Scanner log files generated by different manufacturers' POS systems are not necessarily formatted in a uniform manner. Before any analysis of scanner data can be done, scanner logs from individual chains must be processed into a consistent format and then added together.
2. Conclusive identification of FSP transactions and FSP tender amounts may not be possible using scanner data alone.
3. The number of UPCs in a typical supermarket is approximately 35,000. Meaningful analyses of scanner data must classify UPC information to a manageable taxonomy of food products.

The following sections describe the collected scanner data and the steps taken in processing the data to meet each of the challenges listed above.

¹ UPCs and PLUs are not typically printed on a cash register receipt. When a UPC or PLU is scanned, a computer lookup matches the code to an item description so that the item description can be printed on the receipt. Retailer's transaction log files contain UPCs and PLUs, but not item descriptions.

4.1 Description of Scanned Food Purchase Data

Scanner data collected from the eleven participating stores differed in terms of the scope of data and file format, as shown in Exhibit 4-1.

Exhibit 4-1

Characteristics of Collected Data

Data Source	# Stores	Data Format	Scope
Supermarket Chain A	2	IBM transaction logs; no processing by retailer prior to delivery.	All transactions
Supermarket Chain B	5	ASCII data files: fields extracted from IBM T-logs by retailer.	FSP transactions only
DCSU stores	4	Database files; file structure designed for study.	All transactions

Scope of Data

Chain A provided information about all transactions occurring at its two Georgetown County area supermarkets during the study period. These files contain food stamp and non-food stamp transactions. The files are large, consisting of approximately 90,000 records of transactions (food baskets) and 750,000 records of scanned item sales for each month of the study.

Chain B provided scanner data only for transactions involving food stamp tender. As noted earlier, Chain B nearly dropped out of the study after an initial agreement to participate. This was due to their concern that staff at CCMI, Abt Associates, and the USDA would have access to proprietary information, including customer names, credit card numbers, transaction data, and the like. A compromise was reached that involved Chain B's pre-processing of their transactions to remove all non-FSP eligible items as well as all purchases that did not involve a food stamp payment. Files provided by Chain B were quite manageable in size, consisting of approximately 4,300 transactions and 65,000 scanned item sales for each month of the study.

The four stores with DCSUs agreed to scan all transactions. As we show below, however, we found that retailers did not use the DCSUs consistently and the scanner data contain an incomplete record of transactions at those stores for each month of the data collection.

File Formats and Media

Scanner data at both Chain A and Chain B are stored in an IBM 4690 T-log (transaction log) format. Chain A provided its scanner data to CCMI in this IBM format on IBM tapes. These files were difficult to process because the IBM T-logs were binary and in a very complicated format (no CRLF, comma and colon delimited within the commas). In addition, Chain A delivered these files on IBM tapes in a proprietary format requiring access to an IBM tape reader. Because CCMI did not have an IBM tape

reader, Chain A agreed to send diskettes to CCMI. The diskettes, however, were in the IBM 4690 compressed form that could only be uncompressed by an IBM 4690 system. CCMI therefore enlisted the services of an outside vendor to uncompress and read the tapes and diskette T-log files from Chain A.

In contrast to the data from Chain A, scanner data from Chain B and the non-scanning stores were easier to read and process. Chain B processed its T-log files and sent to CCMI only those scanner records pertaining to FSP-eligible items from food stamp purchases. These extracted data were provided on floppy diskettes in ASCII files with a comma delimited format. Similarly, the scanner data collected from the data collection units in the four non-scanning stores were in an ASCII comma delimited format in files that could be transferred on floppy diskettes.

4.2 Initial Processing of the Scanned Data

In processing the scanner data, the plan was to create the following three main files:

- transaction summary file;
- item-detail file; and
- master item file.

The transaction summary file and item-detail file were constructed by reformatting into a uniform format the records of scanned transactions from the various stores. The transaction summary file contains information about each purchased food basket (e.g., total amount purchased and number of items purchased); the item-detail file contains information about each item purchased (e.g., item price, quantity, weight, and total expenditure on item). The master item file was constructed to be a combined list of all UPC and PLU codes scanned during the data collection period by the eleven retailers.

Reformatting the Scanner Data

After collecting scanner data from retailers, CCMI first converted the scanner data from each source into a consistent format. There were three sources of data: Chain A, Chain B, and DCSUs. The processing involved two steps: (1) parsing the POS data and writing “item-level” data and “basket-level” data to separate files (thereby writing data from all sources to a uniform format), and (2) reformatting files for analysis purposes.

For the first stage of processing, CCMI split the POS data into three intermediate files: (1) a transaction summary file, (2) a payment tender file, and (3) an item-detail file. The first two files are “basket-level” summaries; the third file contains records for each purchased item in the basket. The second step of processing simply combined the basket-level summaries into a single file so that the final analysis files, for each store, consist of: (1) a transaction summary file, and (2) an item-detail.

After this initial processing, CCMI delivered six data files to Abt Associates: a transaction summary file and an item-detail file for each of Chain A, Chain B, and DCSU stores (3 sets of data). The file structure

was consistent across each data source. The reformatted files, however, contained missing data reflecting differences in the number of data fields contained in the original source files.²

Composition of Reformatted Files

The data fields contained in the transaction summary files and item-detail files are shown in Exhibit 4-2, with missing data from each source noted. Because Chain A provided T-logs in their original format, these files contained all data items. Chain B provided an extract of their T-logs with the following data fields provided for each item purchased: transaction identifier, date, time, total amount of purchase, UPC, item description, price paid for item(s), and department code.³ Chain B did not provide all data items listed in Exhibit 4-2, and most items that were not provided could not be constructed. The impact of these missing data items for subsequent processing and analyses is discussed below. The DCSUs were custom designed for this study to collect nearly all data items listed in Exhibit 4-2.

Comparing a typical cash-register receipt with the list of items in Exhibit 4-2 would reveal one main difference: a cash-register receipt does not contain the UPC or PLU code, but it does contain a description of the purchased item. Item descriptions, however, are not typically saved to a retailer's T-logs because item descriptions, and other information about a product, reside in a master file. An item description is computer matched to a UPC/PLU at the time the item is scanned so that it can be printed on the cash register receipt. Likewise, for this study, product descriptions and other item-specific information were maintained in a separate master item file. Construction of the master item file is discussed in Section 4.3.

Impact of Missing Data in Scanner Files

Exhibits 4-1 and 4-2 showed that the main differences in scanner data collected from Chain A, Chain B, and DCSUs are as follows:

- Chain A files contained all required data fields and both FSP and non-FSP transactions.
- Chain B files lacked some important data fields and contained only FSP transactions.
- DCSU data files contained all data fields, but population of those data fields depended on retailer use of the DCSUs.

The most severe cases of missing data in the item and summary files are:

- Chain B did not provide data for:
 - tax amount
 - specific tender type amounts
 - item quantity
 - unit price

2 The DCSU data also differed from the supermarket data in that they were found to be less accurate when compared to EBT transaction logs. This is documented later in this chapter.

3 Chain B added product descriptions to this file in lieu of providing their host price file.

Exhibit 4-2a

Transaction Record Layout: Transaction Summary File

Field Name	Description	Nonmissing Data Items		
		Chain A	Chain B	DCSU
1: Rectype	Record type (1)	X	X	X
2: Txnum	Transaction number	X	X	X
3: Acctnum	Customer account number			
4: Txdate	Transaction date (yyyymmdd hh:mm)	X	X	X
5: Storenum	4-digit store number	X	X	X
6: Txtotal	Transaction total. Excluding taxes and coupons, this field should equal the sum of ext_price in the item file.	X	X	X
7: Taxamt	Tax amount	X		X
8: Cpn_count	Coupon count	X		
9: Tot_cpn	Total coupon amount	X		
10: Misc_tndr	Misc. tender	X		X
11: Txcount	Total count of items [derived from ITEM file]	X	X	
12: Amtten1	Amount cash tender	X		X
13: Amtten2	Amount check tender	X		
14: Amtten3	Amount food stamp tender	X	a	X
15: Amtten4	Amount miscellaneous tender	X		

a The FSP tender amount was constructed by CCMI and is always equal to the total tender amount on Chain B records.
 X denotes data items with non-missing values; no X denotes that all records in file have missing data.

Exhibit 4-2b

Item Record Layout: Item-Detail File

Field Name	Description	Nonmissing Data Items		
		Chain A	Chain B	DCSU
1: Rectype	Record type (2)	X	X	X
2: Txnum	Transaction number	X	X	X
3: Acctnum	Customer account number			
4: Txdate	Transaction date (yyyymmdd hh:mm)	X	X	X
5: Storenum	4-digit store number	X	X	X
6: UPC	UPC or coupon code	X	X	83% missing
7: Deptnum	Department number	X	X	X
8: Price_mult	Price multiplier	X	b	
9: Txprice	Item price. If weight_flag = 1, then txprice = ext_price/item_qty	X	b	X
10: Price_sign	0 = positive price, 1 = negative price	X		X
11: Ext_price	Total paid for item(s)	X	X	X
12: Item_qty	Quantity or weight	X	b	X
13: Weight_flag	0 = non-weighed item, 1 = weighed item and item_qty is lbs.	0-94.6% 1-5.4%	0-100%	0-100%
14: Qty_sign_flag	Quantity sign flag; 0 = add, 1 = void			X
15: Item_type	0 = product purchase, 1 = store coupon, 2 = electronic discount, 3 = manufacturer coupon	0-98% 1-0.1% 3-1.9%	0-100%	0-100%

b These data items were non-missing in the files delivered by CCMI to Abt Associates, but the data items either showed no variation (weigh_fl and price_mu = 0; item_qty = 1) or provided no additional information (item_pric was always equal to ext_pric).
 X denotes data items with non-missing values; no X denotes that all records in file have missing data.

- records of refunds
- records of manufacturers’ coupon redemption
- DCSU stores failed to scan items with the scanner in most cases (83.3 percent of item records are missing UPCs), although these stores entered price information using the keyboards of the data collection scanning units.

Missing data in the files provided by Chain B have the following consequences for analyses based on that chain’s purchases:

- Because information about manufacturers’ coupons was not provided on the item file, the coupon amounts could not be subtracted when aggregating item prices to create total “basket” purchase amount. Therefore, the tender total overstates the actual EBT debit on transactions that include coupon redemption.
- Tender totals overstate the actual EBT debit on transactions that include refunds or voids.
- It would not be feasible to analyze purchases standardized on a quantity or “per serving” basis because item quantities and unit prices were not provided.⁴
- The FSP tender amount could not be identified from the scanner data alone because only the tender total was provided and not a breakdown by tender type. (That is, FSP-only transactions could not be distinguished from transactions that included a combination of FSP tender and other tender such as cash or check). Direct examination of scanner data provided in this format would result in a biased account of the quantity and composition of purchases made with FSP benefits.

Missing data in the files collected from DCSU stores have the following consequences for analysis:

- Examination of the composition of FSP purchases at DCSU stores is not possible because the data do not contain UPCs to identify individual food items. (Also, without UPCs from DCSU stores, we cannot examine the composition of FSP purchases across different store types.)
- The sample of FSP households for which we collected complete scanner data for all monthly FSP redemptions is reduced.

Some of the biases caused by missing data from Chain B are trivial. Evidence from Chain A shows that manufacturers’ coupons were used for only 0.62 percent of items purchased in FSP transactions, and only 0.22 percent of the total purchase amount of FSP transactions was paid for with coupons. Likewise, only 0.42 percent of items in FSP transactions were returned, with 0.74 percent of purchase amounts refunded.

4 Quantity analyses also requires information about package size for each UPC. Package size data was provided in the “UPC lookup file” from Chain A, but the data were spotty and not standardized. We limited “quantity analyses” to an examination of the distribution of the nutrient characteristics of purchased “servings” of breakfast cereals. It did not seem feasible to do this on a larger scale.

Missing data about item quantities and/or unit prices are relevant in cases when multiple items are “rung up” at the same time. For example, if two boxes of Cheerios are purchased at the same time, then the data file from Chain A shows the unit price, the quantity, and the total price for the two boxes; data from Chain B show only the total price for the two boxes without indicating the quantity of items purchased. Because most of our analyses of food purchasing patterns focus on the distribution of food **expenditures** across different food groups, these missing data are not a serious drawback. A more ambitious analysis plan, focusing on FNS’ interest in the nutrient composition of food purchases, would require information about purchased quantities.⁵

4.3 Construction of the Master Item File

CCMI requested that each of the two participating supermarket chains provide its master UPC file. The plan was to merge these two files together and then add the item files that CCMI had constructed for the DCSU stores. The resulting master item file would serve as a standardized list of food items stocked by the participating retailers. Building and maintaining a master list would eliminate the need to link duplicate information (e.g., product category codes) on every item record in the scanner transaction file, and would ensure that item-level information was standardized across retailers.

As constructed, the master item file was to consist of the following data items:⁶

Item ID	Product identification number (UPC or PLU code)
Description	Product description (e.g., Kellogg’s Rice Crispies, 28 oz)
Department	Department identification number (e.g., grocery, meat, seafood)
Category	Product category (e.g., cereal, beef, scallops)
Retailer IDs (vector)	Retail store identification numbers for all retailers stocking the item
Food stamp flag	Flag indicating FSP eligibility (Y = yes; N = no)

All products have a unique product identification number. UPC codes are standardized across retailers. PLU (price look up) codes may vary across retailer locations, but they are unique within a given store. CCMI created a standardized subset of PLU codes across the two supermarket chains.

In order to construct the master item file, CCMI requested a “host price file” (also called the “UPC lookup file”) from Chain A and Chain B. Chain A provided their host price file at the beginning of the data collection period. This file contains the following information for each item stocked at Chain A:

-
- 5 Additional data requirements for nutrient composition analyses include package size information for each UPC, nutrient information corresponding to each UPC, and information about the composition of FSP households in terms of adult male equivalents. These requirements are difficult to meet with existing databases. We received package size information in the “UPC lookup” file from Chain A, but the information was spotty and inconsistently formatted. We know of no existing database that matches nutrient information to UPCs.
 - 6 Under our original design, the master item file was to contain a vector of prices for each product stocked at each retailer. The price vector was to be constructed based on observed item purchases in food stamp and non-food stamp transactions. It was expected that a complete price vector—that is, prices in effect each week of the sample period—would be observed on most commonly-purchased staple food items. Non-food stamp transactions, however, were not provided by Chain B, and there were insufficient food stamp transactions to construct a complete price vector for items stocked at Chain B. As a result, the consumer behavior analysis, which proposed to examine food stamp households’ price-responsiveness, was not possible at this stage of the study.

UPC or PLU, item description, commodity type, department code, food stamp flag, and product category code. Over 62,000 records appeared in the Chain A host price file, of which nearly 29,000 appeared in the scanner logs during the data collection period. Commodity type, department code, and product category code provide three overlapping categorization schemes for items stocked at Chain A stores. These categorization schemes are shown in Exhibit 4-3.

At the end of the data collection period, CCMI requested a second host price file from Chain A—a file that was current as of the end of October 1997. This second request was motivated by the fact that, by the end of September, Abt Associates had found that many items in the scanner logs did not match to an entry in the host price file. This second request would ensure that we had information on items added to inventory during the data collection period. In fact, we found that Chain A had recoded many PLU items during the data collection period, thus explaining our inability to find the items in the original host price file. Our second request for a host price file included a request that information about “quantity size” be added to the file, and Chain A met our request. The union of the old and new “host price files” contained nearly 98,000 items, of which 56 percent were general merchandise and non-food items.

Chain B did not provide their “host price file” due to confidentiality concerns. In lieu of providing this file, Chain B provided item descriptions within their scanner log files. CCMI delivered the original scanner logs from Chain B to Abt Associates so that the item descriptions could be extracted. For Chain B and the non-scanning stores, we therefore had a “host price file” that consisted of only the UPC/PLU and item description.⁷

The information in the Chain A host price files provided the basis for constructing our master item file. Construction of the master item file involved four main steps:

Step 1: As a first step, we assembled a list of all UPC/PLU codes appearing in the scanner logs of the eleven stores participating in the study. The number of unique UPC/PLU codes scanned by the DCSU stores was 595, the number scanned at Chain B was 9,213, and the number of unique codes scanned at Chain A was nearly 29,000. The union of these sets was a list of 33,810 items.

Of the 33,810 items scanned in the eleven stores, 42 percent were non-food and general merchandise items purchased in non-food stamp transactions at Chain A.⁸ We dropped non-food and general merchandise items from our master item file and were left with 19,700 items in the master list that required assignment of narrowly-defined product category codes.

7 Department codes were also provided by Chain B, but the coding scheme was not consistent with that used by Chain A. All items in our master file were standardized on the Chain A coding schemes.

8 Some non-food and general merchandise items were observed in food stamp transactions at Chain A when a combination of food stamp and non-food stamp tender was used in the transaction. We discuss these “combination tender” transactions in Section 4.4. Only 12 items categorized as non-food or general merchandise were observed in the Chain B scanner logs.

Exhibit 4-3

Categorization Schemes Provided in the Chain A Host Price File

Item Type	Product Category		
BOTTLE DEP	BEEF	CHICKEN	COCOA & MILK MODIFI
BOTTLE REF	GROUND BEEF	TURKEY	COFFEE
DAIRY	FRESH PORK	LUNCH MEATS	CONDIMENTS & SAUCES
DELI	FRYING CHICKEN	BARBEQUE MEATS	COOKIES CRACKERS BR
FEES	TURKEYS	SEAFOOD/FISH	DESSERTS
FROZEN	HENS	MISCELLANEOUS DELI	DIET FOODS
GENERAL	SMOKED MEATS	STEAM TABLE ENTREES	FISH CANNED
MERCH	BACON	CHIPS	FLOUR-MEAL-GRITS
GROCERY	CANNED HAMS	DELI INGREDIENTS IT	FRUIT CANNED
MEAT	FRESH SEAFOOD	CHEESE	FRUIT DRIED
PHARMACY	FROZEN SEAFOOD	SALAD BAR ITEMS	HOUSEHOLD CLEANERS
PHOTO	FRESH VEAL	PREPARED SALADS	HOUSEHOLD SUPPLIES
PRODUCE	FRESH LAMB	BAKERY	JAMS-JELLIES-SPREAD
	LUNCHEON MEATS	PIZZA	JUICES CANNED & BOT
	BEEF	HEADLESS SHRIMP	JUICE DRINKS
Department	BEEF OFFALS	MISCELLANEOUS SHRIMP	LAUNDRY SUPPLIES
BAKERY	GROUND BEEF	CATFISH	MACARONI & SPAGHETT
BOTTLE DEP.	FROZEN BEEF PATTIES	SALMON	MEAT CANNED
BOTTLE REF.	BONE-IN PORK	WHOLE FISH	MILK CANNED, POWDER
BULK FOODS	BONELESS PORK	FILLETS	PAPER & PLASTIC PRO
CHEESE	SMOKED PORK LOINS	LONGLINE SPECIES	PET FOOD
SHOP	MARINATED PORK	FARM RAISED TROUT	PICKLES AND RELISHE
COSMETICS	PORK OFFALS	CRAB	PREPARED FOODS
DAIRY	LAMB	VALUE ADDED	SALAD DRESSINGS-MAY
DELI	VEAL	SHELLFISH	SALT-SEASONING-SPIC
DELI TAX	WHOLE TURKEYS	SCALLOPS	SHORTENINGS OIL
FEES	TURKEY PARTS/GRINDS	LOBSTERS	SNACKS
FLORAL	CHICKEN	CRAWFISH	SOAPS - HAND & BATH
FROZEN	CHICKEN BREASTS	SURIMI	SOAPS & DETERGENTS
FOOD	IQF CHICKEN	SEAFOOD SAUCES	SOFT DRINKS & BEVER
GENERAL	MARINATED CHICKEN	SEAFOOD DIPS	SOUP
MCH	HENS	MILK	SUGAR
GROCERY	COOKED CHICKEN	EGGS	SYRUPS AND MOLASSES
MEAT	SMOKED HAMS	REFRIGERATED AND FROZ	TEA
NON-FOOD	SMOKED PICNICS	FROZEN BAKED GOODS	TOBACCO
NUTRITION	SALT BELLIES	MEAT AND FISH	VEGETABLES - CANNED
PHARMACY	COUNTRY HAM	FROZ FRUITS	VEGETABLES - DRIED
PRODUCE	BACON	FROZ JUICES AND D	BAKERY PRODUCTS
SALAD BAR	W/D FRANKS	FROZ PET FOOD	DAIRY PRODUCTS
TAX	FRANKS	FROZ POTATOES	FROZEN FOOD
SEAFOOD	SMOKED SAUSAGE	FROZ POT PIES	BULK MERCHANDISE
SERVICE	FRESH SAUSAGE W/D	FROZ PREPARED FOODS	BEER
MEAT	FRESH SAUSAGE, NAT	FROZ VEGETABLES -	WHISKEY
	SALADS	FROZ VEGETABLES -	PRODUCE MERCHANDISE
	POULTRY LUNCHMEATS	FROZ BREAKFAST FOODS	MARKET MERCHANDISE
	LUNCHEON MEATS	OTHER FROZEN FOODS	DELI MERCHANDISE
	THIN-SLICED LUNCHME	CHEESE	WINES
	SEAFOOD	YOGURT	BOOKS & MAGAZINES
	LUNCHABLES	MARGARINE AND BUTTE	PHARMACY
	MEAT BISCUIT/MUFFIN	REFRIGERATED COOKIE	GENERAL MERCHANDISE
	FULLY COOKED MEAT	MISC. REFRIGERATED	MISCELLANEOUS
	MEAT SEASONING/SAUC	BAKING MIXES	FRESH PRODUCE
	BREADED AND MISCELL	BAKING NEEDS	SPECIALTY PRODUCE
	MISCELLANEOUS MEATS	CANDY & GUM	PLANTS AND FLOWERS
	MEAT SUPPLY ITEMS	CEREAL	FLOWERS
			BALLOONS
			PLANTS

- Step 2:** Merge the old and new “host price files” from Chain A to our master list of scanned items. Of the 19,700 items in our master list, only 70 percent appeared in the host price files and hence had an assigned department and category code. The remaining 30 percent (nearly 6,000 items) required assignment of product category and department codes (2,400 items were scanned at Chain A but did not appear in the Chain A host price files; 3,200 items were unique to Chain B; and 200 items were unique to the DCSU stores).
- Step 3:** The third step in constructing the master item file involved examination of the product categorization schemes and “cleaning” the department and product category codes supplied by Chain A.
- Step 4:** Finally, the last step involved assignment of department and product category codes to the 2,400 Chain A items with missing codes and the 3,400 items added to the file from Chain B and DCSU logs.

Assigning Category Codes

The third and fourth steps listed above involved cleaning and assigning product category codes and department codes to items, reviewing the categorization schemes from the Chain A host price file (see Exhibit 4-3), and reassigning codes as necessary. As a final step, we developed procedures for assigning category codes to items with missing codes.

The item codes were initially “cleaned” by imposing a non-overlapping hierarchy of codes so that the narrowly-defined product category code would correspond to one and only one department.⁹ We reassigned the department code for 4.2 percent of items purchased at Chain A during the data collection period.¹⁰

The next step in cleaning involved review of products by product category code and reassignment of products to **new** categories. The new categories were created by splitting old categories in order to identify and tabulate purchases by narrow categories of interest.¹¹ We added 36 categories to the Chain A product category scheme; the old and new categories are shown in Exhibit 4-4. In addition, we reclassified beer and wine from grocery to non-food.

9 For example, prior to reassignment, cheese appeared in three departments: grocery, cheese shop, and deli; bread appeared in two departments: grocery and bakery; a few other product categories overlapped departments as well. A non-overlapping hierarchy was necessary because, in classifying Chain B items, we had no way of knowing which breads were in the bakery department versus the grocery department. Furthermore, this department distinction was not crucial for our analyses.

10 Of the 62,000 items in the original host price file, 29,000 were purchased at least once (in food stamp and non-food stamp transactions) at Chain A during the data collection period. We reassigned department code on 1,208 items.

11 This reassignment created a coding scheme that allowed us to group product categories to correspond to the categorization scheme used in the U.S. Department of Commerce, Consumer Expenditure Survey. It would not be difficult to recategorize the products to match other coding schemes, such as that used by the Continuing Survey of Food Intakes for Individuals (CSFII), which is conducted by the Agricultural Research Service of USDA.

Exhibit 4-4**Re-Classification of Items from Broad to Narrow Product Categories**

Old Category	New Category
MILK	Milk - Whole Milk - Lowfat Milk - Skim Milk - Flavored Milk - Buttermilk Milk - Lactose free Milk - Other
CHEESE	Natural cheese Processed cheese Cream cheese Sour cream Cottage cheese
MARGARINE AND BUTTER	Butter Margarine and spreads
COOKIES, CRACKERS, BREAD	Cookies Crackers Stuffing/Croutons
SNACKS	Chips Popcorn Nuts Pretzels Pastry Granola bars Rice cakes Fruit rolls
SOFT DRINKS AND BEVERAGES	Carbonated beverages Noncarbonated beverages Water Drink mixers
BAKERY PRODUCTS	Bread Cakes, cupcakes Bread and cracker products Pies, tarts, turnovers Biscuits, rolls, muffins Donuts
FRESH PRODUCE	Apples Bananas Oranges Other citrus fruit Other fruit Potatoes Lettuce Tomatoes Other vegetables

As noted above, 3,400 items were unique to Chain B or the DCSU stores and required assignment of department and category codes. An additional 2,400 items from Chain A were missing category and department codes. The process of assigning category codes involved a mix of automated and manual procedures, as follows:

- Step 1:** Assign category code based on manufacturer. For all manufacturers observed in the Chain A data with products in only a single category code, assign that code to the manufacturer's items with missing codes. (This step was fully programmed using the entire host price files, not the master item file, and categorized 829 items.)
- Step 2:** Assign category code within manufacturer by manually reviewing all item descriptions and categories for a manufacturer when any item for the manufacturer is missing a category code. Then assign a category code, constrained to be a category code already in use for the manufacturer. (This step categorized 3,344 items.)
- Step 3:** Drop items assigned codes in Steps 1 and 2 if those codes correspond to non-food and general merchandise categories. (This step dropped 1,052 items.)

The master item file contains approximately 19,700 items. The number of items for which we assigned category codes was 3,121; the number of items with missing category codes in the final file is 1,727 (these items could not be assigned a category code because the item description was missing or provided insufficient information). As a final step, the non-food and general merchandise items were added back to the master item file, bringing the total file size to 33,810.

Identifying Food Stamp-Eligible Items

One important data item that we expected to include in the master item file was an indicator of food stamp eligibility. Unfortunately, we were unable to determine the process by which retailers established and maintained food stamp eligible information on an item-by-item basis. Chain B extracted only food stamp-eligible items from their scanner logs in preparing scanner data for delivery to us, and they did not disclose their methodology for performing the extraction. Chain A provided all scanned data and a host price file with a food stamp eligible indicator, but they did not provide the documentation necessary to understand fully how the indicator should be used.

The food stamp indicator provided in the Chain A host price file did not appear to be a reliable indicator of food stamp-eligible items. Of the 62,000 items in the original host price file, only 4 percent (2,600 items) had a food stamp flag equal to 1. The food stamp flag appeared primarily on candy and soft drink items (70 percent of all occurrences of the food stamp flag were on candy and soft drinks; 75 percent of candy items and 81 percent of soft drinks were coded as food stamp-eligible). It initially seemed apparent to us, from the pattern of the data, that the food stamp flag must be used in conjunction with product category codes and a "taxable item" flag in order to be meaningful. For example, it may be that all taxable items are deemed food stamp-ineligible except those with a food stamp-eligible flag. The food stamp flag, however, was not consistently applied; for example, some sizes of Diet Pepsi were coded as food stamp-eligible whereas other sizes were not. Furthermore, a small number of non-eligible items appeared to be flagged as eligible. Without a taxable item code and evidence of consistent application of the food stamp flag, we chose not to use this data item for our analyses.¹²

¹² This experience may suggest an area of future FSP research. We know of no previous efforts to ascertain the accuracy with which scanning systems identify FSP-eligible food items.

4.4 Data Processing: Matching Scanner Data to ALERT EBT Log

The scanned food purchase data collected from retailers could not be immediately used for analysis, even after it was reformatted to be consistent across retailers. There were three issues to be resolved—it was necessary to:

1. Assess the completeness of the scanner data. That is, were all EBT food stamp transactions captured in the retailer scanner logs that we received?
2. Assess whether or not the tender type indicators in the scanner log were a reliable means of identifying food stamp transactions. This was especially an issue with the DCSU stores because identification of food stamp transactions depended on the diligence of cashiers in using the data collection scanning units.
3. Add a household identifier to food stamp transactions in the scanner log to enable household-level analyses.

To accomplish these tasks, we matched the scanner data to the ALERT log of EBT transactions.¹³

Matching Scanner Data to ALERT

The ALERT EBT log provided a record of all EBT transactions at the eleven stores during the sample period. ALERT data contain the following information:

- store identifier
- POS terminal identifier
- household FSP case ID
- card account number
- transaction date and time
- transaction amount
- transaction sign (\pm)
- transaction type (purchase, refund, void last transaction, balance inquiry)
- response code (approved, insufficient funds, invalid PIN)
- available balance prior to transaction

We initially expected to match ALERT to the scanner data through a merge by store, date, and time. In fact, we found that only 21 percent of ALERT transactions at the eleven stores matched a scanner record by this merge procedure.¹⁴ It was apparent that the clock time on retailers' POS terminals was not always in alignment with the clock time at the EBT vendor. Alternatively, we used a sequential matching

13 The USDA Food and Nutrition Service's Anti-Fraud Locator of EBT Retailer Transactions (ALERT) Subsystem contains data provided by EBT states' contracted EBT processors. FNS provided copies of South Carolina ALERT files for the months of September and October 1997.

14 The percent of ALERT records matching scanner data by store, date, and time was 36.9 percent for Chain A, 14.1 percent for Chain B, and 9.9 percent for DCSU stores.

procedure: first, the files were matched by store, date, and amount—allowing a match only if the time on the scanner log was within ten minutes of the time in ALERT.¹⁵ This matched 99.9 percent of ALERT records to scanner data for Chain A, 58.4 percent for Chain B, and 55.4 percent for DCSU stores.

Next, we worked with the unmatched ALERT records and unmatched scanner records; sorted both datasets by store, date, and time; and sequentially worked through the data determining matches according to the following rules: the time on the ALERT record must be within ten minutes of the time on the scanner record, and the scanner transaction dollar amount must exceed the ALERT dollar amount.^{16,17} As we worked sequentially through the ordered data, scanner transactions were discarded if the dollar amount was less than the ALERT dollar amount with which it was aligned, and the remaining data were realigned and again checked to see if they were within 10 minutes. The discarded scanner records were manually reviewed; it was found that most of these discards were due to a situation in which two scanner records matched to a single ALERT record. (In other words, the sum of the dollar amount on two adjacent scanner records matched the dollar amount on one ALERT record. This may have occurred if a shopper added items to their food basket after the cashier totaled the transaction, with these two transactions then combined for the purpose of debiting the EBT system.) This situation occurred 94 times and, in these cases, we aggregated the two scanner records.

Exhibit 4-5 summarizes the results of the matching algorithms and is discussed below. There were three main difficulties in matching scanner data to ALERT:

- Although Chain A scanner data contained separate data fields to identify FSP and other tender amounts (enabling a match to ALERT on the FSP amount), it became apparent that the FSP tender variable also identified WIC transactions. This problem was introduced during CCMI's processing of the T-logs from Chain A.¹⁸ This problem became apparent to us because the number of transactions with FSP tender in the scanner log exceeded the number of ALERT transactions for Chain A. The items purchased in these unmatched transactions were clearly WIC items. This accounts for the fact that only 77 percent of transactions in Chain A's scanner file matched to ALERT (see Exhibit 4-5).

15 The time constraint was necessary because there were often multiple transactions at a store for the same amount on a single day.

16 In matching scanner data to ALERT we found a reasonably close, but not exact, correspondence between the time posted by the POS to the scanner log and the time posted by the EBT terminal to the EBT vendor for all stores. For Chain A, the time difference between the scanner log and the ALERT file was within 5 minutes on only 55 percent of transactions, but an additional 37.6 percent of transactions showed a 6- to 7-minute difference. (It seemed likely that the clock setting on one EBT terminal was simply off by 6 to 7 minutes.) For Chain B, 90 percent of transactions posted times in the two files that were within 2 minutes, and 99.9 percent of transactions posted times within 5 minutes. For DCSU stores, 85 percent of transactions posted times in the two files within 5 minutes, and 90 percent of transactions posted times within 10 minutes.

17 The scanner dollar amount could exceed the ALERT amount if a combination of FSP and other tender was used for the transaction.

18 In conversations with CCMI after Abt Associates completed this stage of processing, it became apparent that the T-logs from Chain A included an FSP tender field and sub-fields. It seemed likely that the field that CCMI interpreted as FSP tender, was in fact a "public assistance" or "government tender" field, with subfields for FSP and WIC. A better understanding of the T-logs prior to initial processing probably could have avoided the coding of WIC transactions as FSP transactions.

Exhibit 4-5**Match of ALERT and Scanner Data**

Retailers	Number of Matched Trx	ALERT Data		Scanner Log	
		Number Trx	Percent Matched	Number Trx	Percent Matched
Chain A	3,835	3,905	98.2%	4,966	77.2%
Chain B	8,664	9,006	96.2%	8,822	98.2%
DCSU stores	617	1,428	43.2%	819	75.3%

All of Chain B scanner transactions were FSP transactions. For Chain A and DCSU stores, the number of scanned transactions shown in the table is the number of transactions with FSP tender identified in the scanner data. The total number of scanned transactions received from Chain A was 180,346, and the total number of scanned transactions received from the DCSU stores was 3,270.

- The Chain B scanner data did not separately identify FSP tender and other tender, so that a simple match to the ALERT data by store, date, and amount yielded a low match rate. For Chain B, most matches were determined by the order of the data within each store—requiring that matches be within ten minutes and the scanner amount exceed the ALERT amount. By this method we achieved a high match rate (98 percent of the scanner records were matched to ALERT records). We then assigned an FSP tender amount in the scanner data to be equal to the ALERT amount, and the residual scanner amount was put in the cash tender variable.
- The DCSU data were problematic for two reasons: First, the DCSUs captured only 43 percent of all FSP transactions recorded in the ALERT logs. (It was immediately apparent that the rate of capture was low because only 819 FSP transactions were scanned but 1,428 transactions were recorded in the ALERT log. In fact, only 617 scanned transactions matched ALERT records with a reasonable level of confidence.)

Second, although the data collection scanning unit keyboards allowed retailers to enter FSP tender and cash tender, retailers did not enter separate tender amounts in most cases. Of the 819 scanned transactions with FSP tender, the FSP tender amount was always equal to the total tender amount. Yet, the scanned transaction dollar amount matched the ALERT dollar amount exactly in only 55 percent of matched transactions.

4.5 Completeness of Captured Scanner Data

Exhibit 4-6 presents the results of matching the scanner data to the ALERT logs. By matching the scanner data to ALERT, we could assess whether or not all EBT food stamp redemptions at the participating stores were captured in the data files provided by the stores. The matches for the supermarket chains were quite good. Scanner data from Chain A represent 98.2 percent of all EBT food

stamp redemptions at Chain A stores during the study period; scanner data from Chain B represent 96.2 percent of EBT food stamp redemptions during the study period. In contrast, scanner data from the DCSU stores represent only 43.2 percent of EBT food stamp redemptions at these stores during September and October. Chain A and Chain B both experienced computer telecommunications problems resulting in a loss of archived data during the study period, thus accounting for nearly all “uncaptured” data. The low percent of captured data at the DCSU stores is explained by incomplete use of the temporary scanning units at those stores.

Exhibit 4-6 summarizes the number of EBT transactions and EBT redemptions captured in the scanner data, by store and month.

4.6 Adjustments to Combination-Tender Transaction Records

As noted above, a large percent of all FSP transactions occurred during shopping trips in which both FSP tender and other tender (e.g., cash, check, or manufacturers’ coupons) were used to purchase grocery items. FSP clients may use a combination of tender when purchasing both FSP-eligible and FSP-ineligible items in a single shopping trip; alternatively, FSP clients may use a combination of tender when exhausting their FSP benefits, regardless of the content of the basket of items.

Of the scanned transactions that were matched to the ALERT file, 36.9 percent from Chain A, 41.2 percent from Chain B, and 36.5 percent from DCSU stores were combination-tender transactions. (These percentages were measured after the match to ALERT provided confirmation of the FSP tender amount.)

We observed two types of combination-tender transactions: food baskets containing only FSP-eligible items paid for with a combination of FSP tender and other tender (usually cash), and food baskets containing a combination of FSP-eligible and FSP-ineligible items paid for with a combination of tender. Combination-tender transactions from Chain B, by construction, contained only FSP-eligible items, because Chain B stripped off all FSP-ineligible items.¹⁹

Combination-tender transactions pose a complication for the analysis of food stamp purchases. In combination-tender transactions, the total expenditure on grocery items exceeds the dollar amount debited to the FSP client’s EBT account. Clearly we do not want to count the entire food basket when reporting on food stamp purchases. Because our main goal in analyzing food stamp purchases is to examine the distribution of food stamp redemptions (i.e., dollars redeemed) across product categories, we adjusted the combination-tender transactions so that their weight in the data reflected only the FSP redemption amount.

19 Because of the construction of the Chain B scanner files, it is somewhat surprising that Chain B shows the highest percentage of combination-tender transactions. This result is due to the differing treatment of manufacturers’ coupons in the scanner logs of Chain A and Chain B.

Coupons appear as items with a negative price in the Chain A scanner logs and are not treated as tender. In contrast, coupons are not observed as items in the Chain B scanner logs (because Chain B did not provide this information), so that coupon redemptions contribute to the total tender. Hence, when coupons were redeemed at Chain B, the total tender in the scanner log exceeded the FSP dollar debit amount seen in ALERT.

Exhibit 4-6**EBT Transactions and Redemption Captured in Scanner Data**

Store	Number EBT Transactions			Amount of EBT Redemptions		
	Total Number	Captured Number	Percent	Total Amount	Captured Number	Percent
September 1997						
Chain A #1	391	389	99.5%	\$13,647	\$13,576	99.5%
Chain A #2	1,500	1,499	99.9	44,398	44,348	99.9
Chain B #1	358	352	98.3	11,199	10,888	97.2
Chain B #2	1,541	1,526	99.0	44,482	44,240	99.5
Chain B #3	904	900	99.6	22,777	22,432	98.5
Chain B #4	528	521	98.7	12,622	12,511	99.1
Chain B #5	965	960	99.5	22,389	22,285	99.5
Store C	196	32	16.3	910	94	10.3
Store D	265	181	68.3	1,791	1,282	71.6
Store E	114	4	3.5	1,286	24	1.9
Store F	95	2	2.1	1,531	7	0.5
Total	6,857	6,366	92.8%	\$177,032	\$171,687	97.0%
October 1997						
Chain A #1	429	415	96.7%	\$17,114	\$16,451	96.1%
Chain A #2	1,585	1,532	96.7	45,673	43,841	96.0
Chain B #1	367	334	91.0	12,031	11,171	92.9
Chain B #2	1,684	1,624	96.4	48,360	46,978	97.1
Chain B #3	936	895	95.6	24,005	23,198	96.6
Chain B #4	620	495	79.8	14,367	11,291	78.6
Chain B #5	1,103	1,057	95.8	23,263	22,358	96.1
Store C	222	16	7.2	922	56	6.1
Store D	286	243	85.0	1,826	1,598	87.5
Store E	151	121	80.1	1,496	1,286	86.0
Store F	99	18	18.2	2,140	126	5.9
Total	7,482	6,750	90.2%	\$191,197	\$178,354	93.3%
Totals						
Chain A	3,905	3,835	98.2%	\$120,832	\$118,216	97.8%
Chain B	9,006	8,664	96.2%	\$235,495	\$227,352	96.5%
Stores C-F	1,428	617	43.2%	\$11,902	\$4,473	37.6%

Source: ALERT files, USDA. Total number of EBT transactions and redemptions by all South Carolina FSP households. Captured data are measured by the number and amount of EBT transactions successfully matched to ALERT files.

The following procedure was followed to adjust combination-tender transactions:

- Step 1:** All FSP ineligible items were removed from the item file, and the total cost of the food basket was recalculated as the total expenditure on FSP-eligible items. (This step was unnecessary for Chain B transactions.)
- Step 2:** The ratio of the EBT debit amount to the cost of FSP-eligible items was calculated.
- Step 3:** For each FSP-eligible item in the food basket, an **FSP expenditure** was calculated by applying the ratio in Step 2 to the total item expenditure.

This method does not require arbitrary assumptions about which FSP-eligible grocery items were purchased with FSP benefits and which items were purchased with other tender.

With the completion of this adjustment to combination-tender transactions, we were ready to begin analysis of the scanner data, which is the subject of the next chapter.

4.7 Demographic Data

The previous section discussed the match of scanner transactions to ALERT data. This match was a one-to-one match of records that describe a transaction at a food store. The scanner data and ALERT data, however, differ in critical ways. The scanner data contain information about items purchased, but there is no information to identify the purchaser or to link together multiple transactions by the same household. In contrast, the ALERT data contain the FSP case identification number (CaseID), which allows multiple purchases by the same recipient to be linked together so that purchasing behavior over time may be examined. The CaseID also may be used to add demographic information about FSP households to the transaction records.

After we merged the ALERT and scanner data, the CaseID from the ALERT file was added to the scanner file and used as a merge variable to add demographic information. Demographic information was obtained from the South Carolina Department of Social Services for all active FSP participants as of September 1997. The information contained in the administrative records from South Carolina is presented in Exhibit 4-7.

Because the original data collection period was extended to a second month, FSP households that were new cases in October are not reflected in the September extract of the state administrative system. In addition, we do not have demographic information for food stamp households that left the program prior to September but spent remaining FSP benefits during September (because we received information only for **active** cases). We did not expect these numbers to be large, but were surprised to find that 12.3 percent of all FSP households who shopped at the eleven stores during the data collection period were not found in the South Carolina administrative extract. This 12.3 percent of FSP cases accounted for 9.0 percent of all EBT transactions and 9.6 percent of EBT redemptions at the eleven stores during the data collection period. The distribution of transactions for these cases was evenly split between the September and October calendar months.

Exhibit 4-7**Information Obtained from South Carolina Administrative Records**

Data Item	Definition
CASEID	Case number
BNFTAMNT	Benefit amount as authorized by E.W.
COUNTY	County
DISABLED	Elderly/Disabled indicator
DOB	Client date of birth
EARNED	Total gross earned income for household
EDUCLVL	Highest grade completed
ETHNIC	Ethnic status
HHSIZE	Total persons in assistance unit
HHTYPE	Household type
HOMELESS	Homeless household indicator
MARIT	Marital status
MIGRANT	Migrant household indicator
PARTCODE	Participation code
PGMTYPE	Program type
REGAMNT	Regular benefit amount before proration
UNEARN	Total gross unearned income for household
ZIPCODE	ZIP Code

The FSP households with missing demographic information showed the following pattern of association with ALERT transactions in the two months of collected data.

Transactions in:	Percent with Missing Demographics		
	Cases	Transactions	Redemptions
September only	14%	6%	4%
October only	60%	41%	44%
Both months	26%	53%	52%
Total	100%	100%	100%

The pattern suggests that most cases (60 percent) for which we are missing administrative (demographic) information are cases that were new to the FSP in October 1997; an additional 26 percent of cases redeemed benefits in both September and October, but may have been approved for benefits after the start of the September disbursement month (as suggested by the disproportionately large percentage of

transactions associated with these cases). The few cases (14 percent) observed only in September are likely to be cases that did not receive a disbursement in September but spent down benefits received in prior months. This hypothesis is supported by the relatively few transactions and low level of redemptions associated with these cases.

Chapter 5

Food Purchasing Patterns of Food Stamp Households

The two main research goals of this study are to explore the feasibility, difficulties and limitations of using scanned food purchase data to:

1. Describe the food purchasing patterns of FSP participants, and
2. Compare the purchasing patterns of FSP participants with non-FSP participants.

These goals were achieved by collecting scanner data at checkouts, identifying FSP purchases within the scanner data, and tabulating and comparing FSP purchases with non-FSP purchases. Results of the analyses are presented below, following discussion of data limitations.

Due to the small number of stores that participated in the study, the analysis results are not intended to be representative of FSP or non-FSP purchases in Georgetown County or elsewhere. Rather, they are suggestive of the types of analyses that can be done with scanner data.

This chapter begins with a discussion of the sample of collected food purchase data and the sample of households represented in the data. Section 5.2 reviews the analysis tasks proposed for the study, noting those that were achievable with the collected scanner data and those that were not. Sections 5.3 and 5.4 present analyses of scanner data from the seven supermarkets in our sample. Section 5.5 presents a limited analysis of scanner data collected from the four DCSU stores.

5.1 Sample of FSP Retailers and Households

The study design called for selection of a single county site with: (a) a moderate number of retailers, and (b) a high proportion of FSP redemptions by county residents occurring within the county.¹ The goal was eventually to collect scanner data that would represent most of the monthly FSP redemptions of county residents, although issues of sample representativeness were not the main focus of the current study.²

There were 53 active FSP-authorized retailers in Georgetown County and at least 2,427 active FSP households during September and October 1997.³ The 53 retailers redeemed approximately \$450,000 of

1 The goal was to select a county in which residents of the county shopped within the county. This was important because we wanted to recruit a group of retailers and collect data representing monthly household FSP redemption by a sample of households.

2 Recall that a second study phase, to include more retailers and more scanner data, was originally planned.

3 The number of Georgetown County FSP households may be understated because we could not assign a county code to about 12 percent of FSP households shopping at the eleven stores, as explained in Section 4.7.

FSP benefits per month; approximately 87 percent of these redemptions were by Georgetown County households.

The eleven retailers participating in the study accounted for 41 percent of all FSP redemptions in Georgetown County during the two-month study period. Seventy-five percent of Georgetown County FSP households (1,827) shopped at one of the eleven stores at least once. Additional FSP households from neighboring counties also shopped at these stores, so that the data include a total of 2,795 FSP households shopping at the eleven participating retailers, with 65 percent being Georgetown residents.⁴

The study's initial research design called for an analysis of FSP redemption behavior by Georgetown County residents. It quickly became apparent, however, that there was no reason to restrict the sample of FSP households to those residing in Georgetown County. One reason to drop the restriction was that, in comparing FSP and non-FSP food purchases, we would not be able to restrict non-FSP data to Georgetown County residents. Another reason was that we did not know the county of residence for 12.3 percent of all households shopping at the eleven stores. Finally, as discussed below, the collected scanner data would not support household-level analyses (i.e., an examination of the distribution of FSP households according to their purchasing behavior), so that it was no longer important to examine a sample of households representing a single county.

Exhibit 5-1 summarizes the samples of: (1) all FSP participants residing in Georgetown County ($n = 2,427$), and (2) South Carolina FSP participants who redeemed benefits at the eleven stores participating in our study ($n = 2,795$). We chose to define the sample according to (2) because it yielded a larger sample of households with a higher concentration of FSP benefit redemption at the sample stores.⁵ Exhibit 5-1 shows that Georgetown County FSP households redeemed 35.7 percent of their FSP benefits at the eleven sample stores, whereas households who shopped at the eleven stores at least once redeemed 40.4 percent of their FSP benefits at the eleven sample stores.

Captured Scanner Data, by FSP Household

At this first stage of the feasibility study, we hoped to capture a high percentage of the FSP redemptions of some sample of households. The first column of Exhibit 5-2 shows, however, that only 13.5 percent of FSP participants known to reside in Georgetown County redeemed **all** of their FSP benefits at the eleven sample stores. Among all South Carolina residents, 2,795 FSP households shopped at the eleven stores at least once during the sample period, and 13.4 percent redeemed all of their FSP benefits at the eleven stores. When looking only at FSP redemptions in supermarkets, between 16 and 17 percent of the FSP households in the sample did all their supermarket shopping in one or more of the study's seven participating supermarkets.

Exhibit 5-2 shows the distribution of FSP households by the percent of redemptions actually made at the eleven stores, based on the ALERT record of EBT transactions. Exhibit 5-3 shows the total FSP redemptions of households in our sample (by store type), the amount and percent of redemptions that

4 Recall from Chapter 3 that one or more of the eleven participating stores are located in adjacent counties.

5 As discussed in Section 4.7, administrative data were unavailable for 12.3 percent of FSP cases. We do not know the county of residence for these cases, and use of definition (2) allowed us to include the FSP households for which we did not know the county of residence.

Exhibit 5-1
EBT Transactions in Georgetown County South Carolina

I. EBT Redemptions by Georgetown County Residents (N=2,427)

	All Stores	Georgetown County Stores		Eleven Stores in Study	
		Number	Percent	Number	Percent
Number of FSP-authorized retailers where benefits redeemed	391	53	13.6%	11	2.8%
Number of EBT purchase transactions	32,214	28,438	88.3%	11,244	34.9%
EBT redemptions (dollars)	\$777,317	\$665,399	85.6%	\$277,705	35.7%

Source: ALERT files for September and October 1997.

II. EBT Redemptions by South Carolina FSP participants who EVER shopped at Eleven Stores in Study (N=2,795)

	All Stores	Georgetown County Stores		Eleven Stores in Study	
		Number	Percent	Number	Percent
Number of FSP-authorized retailers where benefits redeemed	574	53	9.2%	11	1.9%
Number of EBT purchase transactions	38,452	30,137	78.4%	14,462	37.6%
EBT redemptions (dollars)	\$912,657	\$714,225	78.3%	\$368,487	40.4%

Source: ALERT files for September and October 1997.

Exhibit 5-2
Distribution of FSP Households, By Percent of EBT Redemptions at Sampled Stores

	Georgetown County	Non-Georgetown Counties	Entire State
Number households	2,427	142,762	145,189
FSP households EVER shopping at sample stores			
Number	1,827	968	2,795
Percent	75.3%	0.7%	1.9%
Distribution of FSP households by percent of FSP redemptions at sample stores			
1-10%	11.0%	24.7%	15.7%
11-25%	15.4%	21.4%	17.5%
26-50%	23.6%	18.5%	21.9%
51-75%	20.5%	12.3%	17.6%
76-90%	10.2%	5.9%	8.7%
91-99%	5.8%	3.9%	5.2%
100%	13.5%	13.3%	13.4%
Distribution of FSP households by percent of FSP supermarket redemptions at sample stores			
Zero	2.6%	2.3%	2.5%
1-10%	8.7%	20.8%	12.9%
11-25%	14.1%	20.6%	16.4%
26-50%	22.2%	17.9%	20.7%
51-75%	19.5%	11.5%	16.7%
76-90%	10.0%	6.3%	8.7%
91-99%	6.3%	4.3%	5.6%
100%	16.7%	16.4%	16.6%

Note: EBT redemptions at sampled stores are measured from the ALERT file and exceed the amount of data captured at the sample stores, which is shown in Exhibit 5-4.

Source: ALERT, September and October 1997.

Exhibit 5-3

Captured and Uncaptured FSP Redemptions for Households EVER Shopping at Sample Stores, September and October 1997



Distribution of Uncaptured FSP Redemptions for Households EVER shopping at Sample Stores

Store Type	Total Redemptions Amount	Uncaptured Redemptions		Percent of uncaptured data by reason					
				Amount	Percent	Store not in study		Disk failure/ lost data*	No match to ALERT
						Georgetowr County	Other counties		
SEPTEMBER 1997									
Supermarket	\$364,581	\$194,302	53.30%	74.1%	25.3%	-	0.6%		
Grocery	\$9,085	\$7,803	85.90%	14.1%	79.4%	-	6.5%		
Specialty	\$11,810	\$11,810	100.00%	85.2%	14.8%	-	0.0%		
Convenience	\$17,590	\$17,464	99.30%	38.7%	40.7%	-	20.6%		
Other	\$2,918	\$2,918	100.00%	57.1%	42.9%	-	0.0%		
Total	\$405,984	\$234,297	57.71%	69.8%	28.0%	0.0%	2.2%		
OCTOBER 1997									
Supermarket	\$396,285	\$221,007	55.80%	71.0%	24.7%	2.2%	2.1%		
Grocery	\$10,349	\$8,751	84.60%	9.6%	87.7%	-	2.7%		
Specialty	\$12,980	\$12,980	100.00%	86.4%	13.6%	-	0.0%		
Convenience	\$18,388	\$16,921	92.00%	41.3%	40.4%	-	18.3%		
Other	\$2,213	\$2,213	100.00%	65.3%	34.7%	-	0.0%		
Total	\$440,216	\$261,872	59.49%	67.7%	27.4%	1.9%	3.0%		

*One supermarket chain experienced disk failures on two dates, resulting in an unrecoverable loss of transaction data.

were not captured, and the distribution of uncaptured redemptions by reason. The main reason that the study failed to capture all redemptions for these households is that households shopped at multiple stores both inside and outside of Georgetown County. A small number (2.2 percent) of supermarket redemptions in October were not captured because of disk failures and telecommunication problems at the stores in the study. Nearly 20 percent of the convenience store redemptions were not captured because the stores using the DCSUs did not use the units properly.

A large percentage of households who shopped at the eleven stores redeemed more than 50 percent of their FSP benefits at stores that were not in our sample for data collection.⁶ This finding led us to preclude household-level analyses at this stage of the feasibility study.⁷

Demographic Characteristics of FSP Households

The 2,795 FSP households who shopped at the eleven sample stores are described in Exhibit 5-4. The characteristics of the entire sample appear in the far right column labeled “Total.” Other columns show the characteristics of subsamples defined by the percent of FSP redemptions captured in the scanner data.

Overall, the households represented in the scanner data received an average monthly FSP allotment of \$188.59, made 7.9 EBT transactions per month on average, and redeemed 91 percent of their allotment at supermarkets. In comparison, all FSP households in South Carolina had an average monthly allotment of \$163.66, made 6.7 EBT transactions per month on average, and redeemed 86 percent of their benefits at supermarkets.

The difference in demographic characteristics of the sample, by percent of redemptions captured by the study, shows that the small number of households for which we have complete food purchasing data **is not representative of the entire sample of households that shopped at these eleven stores.** The households for which we captured complete data are smaller on average, with fewer children; they are more likely to have elderly persons in the household, they have a smaller FSP allotment, and they make fewer EBT transactions.

In the subsequent examination of food purchasing patterns in Sections 5.3 and 5.4, we compare food purchasing patterns of three types of FSP households: households with elderly persons and no children (9.8 percent of total sample), households with children (57.8 percent), and households with neither elderly nor children (20.1 percent). An additional 12.3 percent of households could not be characterized because they did not appear in the FSP administrative data received from the state of South Carolina.⁸ The characteristics of these four groups of households are shown in Exhibit 5-5.

6 Half of Georgetown County FSP households redeemed more than 50 percent of their benefits at stores outside our sample for data collection; 55 percent of households **ever** shopping at sample stores redeemed more than half of their FSP benefits at non-sample stores.

7 By “household-level analyses” we refer to analyses examining the distribution of **households** according to characteristics of their monthly benefit redemption, rather than an aggregate analysis examining the distribution of transactions or redemptions. Household-level analyses are not valid when we observe only a portion of the monthly benefit redemption of most households.

8 See Section 4.7.

Exhibit 5-4

Characteristics of FSP Households and FSP Redemptions, By Percent of Captured Data: Households EVER Shopping at Sample Stores

	Percent of Redemptions Captured						TOTAL
	1-25%	26-50%	51-75%	76-90%	91-99%	100%	
FSP HOUSEHOLDS:							
Number	929	611	493	243	144	375	2,795
Percent	33.2%	21.9%	17.6%	8.7%	5.2%	13.4%	100.0%
DEMOGRAPHICS¹							
Percent in Georgetown county	52.0%	70.7%	75.9%	76.5%	73.6%	65.6%	65.4%
Avg household size	3.2	3.0	2.7	2.7	2.9	1.6	2.8
Avg number children	1.8	1.6	1.4	1.5	1.5	0.4	1.5
<i>Percent of households with:</i>							
Elderly	8.2%	7.9%	11.6%	9.5%	9.7%	32.0%	12.1%
Children	67.0%	62.4%	61.1%	58.9%	60.4%	21.9%	57.8%
Spouse present	14.0%	10.5%	11.0%	7.8%	14.6%	9.3%	11.6%
AFDC/TANF	16.5%	12.0%	12.0%	9.1%	11.8%	2.4%	11.9%
Earned income	34.3%	34.0%	35.3%	34.2%	38.2%	18.9%	32.6%
Demographics not known ²	12.6%	12.3%	8.9%	10.7%	13.2%	16.8%	12.3%
EBT REDEMPTIONS³							
Avg number of monthly EBT trx	10.2	8.7	7.5	7.5	6.6	2.1	7.9
Avg number of monthly EBT trx at sample stores	1.7	3.2	3.9	5.2	5.4	2.1	3.0
Avg monthly FSP disbursement	\$232.09	\$205.89	\$186.96	\$186.77	\$175.22	\$61.11	\$188.59
Percent of EBT redemptions by store type:							
Supermarkets	86.4%	89.8%	92.3%	94.3%	96.3%	97.9%	90.9%
Grocery stores	3.8%	2.5%	1.4%	0.6%	0.2%	0.9%	2.2%
Specialty stores	3.3%	3.6%	2.4%	2.2%	0.9%	0.0%	2.5%
Convenience stores	5.7%	3.8%	3.4%	2.7%	2.5%	1.2%	3.9%

Notes

¹ Source is South Carolina State Food Stamp Program caseload extract for October 1997.

² Demographics are not known for FSP households who redeemed benefits in September or October but did not receive a disbursement in October, when the caseload extract was drawn.

³ Source is Food Stamp Program ALERT data (EBT transaction log for September and October 1997).

Exhibit 5-5

Characteristics of FSP Redemptions, By Household Type: Households EVER Shopping at Sample Stores

	Household Type				TOTAL
	Households with elderly and no children	Households with no children	Households with children	Unknown household type	
FSP HOUSEHOLDS:					
Number	273	562	1,616	344	2,795
Percent	9.8%	20.1%	57.8%	12.3%	100.0%
DEMOGRAPHICS¹					
Percent in Georgetown county	74.7%	76.2%	73.9%	.	65.4%
Avg household size	1.2	1.2	3.6	.	2.8
Avg number children	0	0	2.2	.	1.5
<i>Percent of households with:</i>					
Elderly	100.0%	0.0%	4.0%	.	13.8%
Children	0.0%	0.0%	100.0%	.	65.9%
Spouse present	11.0%	6.4%	15.9%	.	13.2%
AFDC/TANF	0.4%	2.7%	19.6%	.	13.6%
Earned income	2.2%	14.1%	51.1%	.	37.1%
EBT REDEMPTIONS²					
Avg number of monthly EBT trx	3.2	5.5	9.6	7.6	7.9
Avg number of monthly EBT trx at sample stores	1.8	2.4	3.4	2.8	3.0
Avg monthly FSP disbursement	\$ 54.44	\$ 101.06	\$ 240.59	\$ 193.83	\$ 188.59
Percent of EBT redemptions by store type:					
Supermarkets	96.0%	90.4%	90.3%	90.6%	90.9%
Grocery stores	1.3%	2.0%	2.6%	1.8%	2.2%
Specialty stores	1.3%	2.5%	2.7%	2.8%	2.5%
Convenience stores	1.3%	4.8%	3.8%	4.5%	3.9%

Notes

¹ Source is South Carolina State Food Stamp Program caseload extract for October 1997.

² Source is Food Stamp Program ALERT data (EBT transaction log for September and October 1997).

The Sample Period

In South Carolina, FSP benefits are disbursed over a ten-day window during the first ten days of each calendar month. The disbursement schedule implies that “monthly” redemption behavior occurs over a calendar period that varies by household according to the date of benefit disbursement. Previous research of FSP redemption patterns has been based on the disbursement month rather than the calendar month. To examine monthly behavior, households are aligned so that, for each household, “Day One” is the day of disbursement.⁹

For this study, scanner data were collected for two calendar months: September and October 1997. After matching the scanned food purchase data to the ALERT data, we assigned a date of disbursement to each FSP household in the sample.¹⁰ This allowed us to contrast the monthly pattern of FSP redemption with the monthly pattern of food purchase transactions displayed in non-FSP transactions. Because we did not capture all FSP redemptions for each household in our sample, however, we did not examine the composition of food purchases by time of month.

Summary of Sample Definition

- All analyses included in this report are based on the sample of FSP households who **ever** shopped at sample stores during September and October 1997. Sixty-five percent of these households reside in Georgetown County (see Exhibit 5-2).
- September and October data are pooled for nearly all analyses.
- Data from the DCSU stores are not included in the main analyses in this chapter because we were unable to characterize most of the purchased items in these transactions.
- Comparisons of FSP and non-FSP transactions are based on data from the single supermarket chain (Chain A) that provided both FSP and non-FSP transactions.
- Data from both Chain A and Chain B (seven supermarkets) are used to examine the distribution of items purchased in FSP transactions.
- Although the data do not support household-level analyses of FSP redemption, FSP transactions are aggregated into three groups to compare the composition of grocery purchases across types of FSP households (elderly, households with children, households without children).

9 Cole, Nancy, “Evaluation of the Expanded EBT Demonstration in Maryland: Patterns of Food Stamp and Cash Welfare Benefit Redemption,” USDA Food and Consumer Service, February 1997.

10 ALERT does not include information about disbursement dates, but with multiple months of ALERT data the disbursement date is implied by the trend in EBT account balances. The disbursement date is observed with certainty when transactions occur on consecutive days before and on the date of disbursement. Only a fraction of households statewide made consecutive transactions revealing their disbursement date, but these households revealed a clear pattern of disbursement whereby the calendar date (day) of disbursement corresponded to the last digit of the case ID number (i.e., IDs ending in “1” received disbursement on the first of the month).

5.2 Application of Scanner Data to Proposed Analyses

The potential analyses of scanner data included two broad goals: first, a descriptive analysis of the food purchasing patterns of FSP households, and second, an examination of consumer behavior with respect to food purchases. At this initial stage of the feasibility study, we did not expect to have the breadth of data necessary to pursue the second level of analysis.

The specific analysis goals are as follows:¹¹

- **Examine the completeness of captured food purchase data.** Examine the percentage of total EBT food purchases captured in scanner files, the reason for lack of capture, the potential for imputing missing data, and a comparison of the study site with other EBT sites for which transactions data have been analyzed.
- **Food purchasing patterns.** Examine households' allocations of FSP allotment among types of food products (by product category, branded vs. non-branded, nutritional characteristics).

Examine the variation in purchase allocations by household characteristics, by store type, and by urban/non-urban location.

Compare overall patterns of FSP purchases to a sample of non-FSP purchases in the scanner data. Compare patterns by demographic group to published data.

- **Consumer behavior.** Evaluate households' responsiveness to price changes (price elasticity) and the variation in price responsiveness by product category and by household characteristics.

Completeness of Scanned Data

The completeness of captured food data was examined in Sections 4.5 and 5.1 above. The study captured nearly all food stamp EBT transactions at the supermarkets participating in the study, but was not successful in capturing "food data at checkout" at stores equipped with DCSUs. Even if the study had been successful at DCSU stores, we would have captured complete food stamp redemption information for only a small number of food stamp households because households also shopped at stores outside our sample.

Due to the small number of retailers participating in this stage of the study, we captured complete monthly FSP redemptions for only 375 households. These households are not representative of the FSP caseload in South Carolina, and this sample was not large enough to support household-level analyses.

¹¹ Abt Associates Inc., "Feasibility Study of Capturing Food Data at Checkout, Technical Proposal: Final Submission," September 18, 1996.

Food Purchasing Patterns

The examination of food purchasing patterns, to be presented in Sections 5.3 and 5.4, is limited by the collected data in two respects. First, we are unable to examine the composition of food purchases by store type because there is no variation in the captured scanner data by store type—only the data from supermarkets were useable. Second, we were unable to examine the composition of food purchases by urban/non-urban location because the extract provided by the South Carolina Department of Social Services did not include the address of the recipients, which would have allowed us to code recipients as residing in urban or non-urban census tracts.¹²

Consumer Behavior

The consumer behavior analyses were planned only for the second stage of data collection, after enrolling a larger number of retailers in the study. Nonetheless, it is useful to review the reasons why the scanned food purchase data collected at this stage would not support this analysis.

The purpose in estimating price elasticities of demand is to ascertain the willingness of consumers to “shop around”—between stores and over time—in response to price changes. At this stage of the study we obtained data from only two supermarket chains. Of the FSP households that shopped at Chains A and B, 69 percent shopped **either** at Chain A or Chain B, and 31 percent shopped at both chains. Therefore, there is probably enough “store-switching” in these data to yield estimates of price elasticities of demand that reflect both temporal changes in price and the willingness of consumers to “shop around” seeking bargains.

There are two main reasons why the present scanner data would not support valid estimates of price elasticities. First, we did not obtain scanner data on all FSP purchases for a sample of FSP households; incomplete information would produce biased estimates of household behavior. Second, we did not obtain price information in the scanner data provided by Chain B. Chain B provided the following information about purchased items: UPC or PLU, item description, and item expenditure. A crucial piece of information missing from the Chain B item file was the quantity (number of items) purchased. Because we did not have the quantity information, we could not divide expenditure by quantity to calculate the item price.

5.3 Characteristics of Food Baskets Captured at Checkout

This section provides a description of FSP and non-FSP food basket transactions observed in the scanner data, including the number and size of food baskets, the distribution of food baskets throughout the month, and the content of food baskets, particularly the distribution of purchased items across product category. **Readers are reminded that the following analysis results are not representative of purchases made by FSP households in Georgetown County or elsewhere.** The section demonstrates the types of analyses that can be conducted with scanned food purchase data.

¹² The failure to request address information was an oversight on the part of the study team.

Number and Size of Food Baskets

Scanner data were collected from seven supermarkets in two supermarket chains, and from four non-scanning stores (via DCSUs). Exhibit 5-6 shows the number of transactions (baskets) scanned at each store, the average basket size (measured by number of items), and the average dollar cost of the basket.

As discussed in Chapter 4, we observe two types of food stamp transactions in the scanner data: (1) “FSP only” transactions, in which FSP benefits are the sole source of tender; and (2) combination-tender transactions, in which FSP benefits are combined with other tender (cash, check, miscellaneous) to pay for the food basket.¹³ Exhibit 5-6 shows the characteristics of FSP-only transactions, non-FSP transactions (at Chain A and the DCSU stores), and combination-tender transactions.

FSP transactions show considerable variation across stores in this sample. The average number of items per basket is lowest at the DCSU stores and highest at Chain A stores, with the average basket cost varying accordingly. Combination-tender transactions represent larger food baskets than FSP-only transactions; interestingly, even the FSP-tender portion of combination-tender transactions exceeds the size of FSP-only transactions.

Distribution of Captured FSP and Non-FSP Food Baskets, By Time of Month

In comparing the characteristics of FSP transactions and non-FSP transactions, it is important to recall that monthly food stamp benefit redemption is characterized by a rapid “spend-down” of the FSP allotment. Previous research shows that 70 percent of FSP benefits are redeemed within one week of disbursement.¹⁴ The rapid spend-down may account for the larger average size of FSP food baskets.

Exhibits 5-7 through 5-10 show the time trend in the characteristics of scanned food baskets at Chain A and Chain B. There are several important differences between FSP and non-FSP transactions. (FSP transactions include FSP-only and combination-tender transactions.) First, the number of FSP transactions decline steadily throughout the month, whereas non-FSP transactions display a consistent pattern throughout the month, with notable peaks on Saturdays.¹⁵ Exhibit 5-8 shows the pattern of transactions with daily activity expressed as a percentage of the monthly total: this graph shows that the distribution of FSP transactions at Chain A and Chain B is nearly identical, and that the Saturday peaks in non-FSP activity are not very large in percentage terms.

13 The following specific tender types were identified in the Chain A data: FSP, cash, check, manufacturers’ coupons, other. For Chain B and the DCSU stores, combination-tender transactions were identified in the data by the fact that the scanned transaction total exceeds the transaction total in the ALERT EBT log; the difference between the store transaction total and the EBT total was inferred to be non-FSP tender.

14 Cole, Nancy, “Evaluation of the Expanded EBT Demonstration in Maryland: Patterns of Food Stamp and Cash Welfare Benefit Redemption,” *op. cit.*

15 The FSP transaction data are graphed by “days since disbursement” rather than by calendar day because disbursement is spread over a ten-day window at the beginning of each month. Therefore, retailers do not experience the extreme peak of FSP activity that is displayed on the graph. A graph of FSP transactions by calendar day does **not** display the Saturday peaks that are seen in non-FSP data.

Exhibit 5-6
Characteristics of Transaction Baskets Scanned at Checkout, By Tender Type

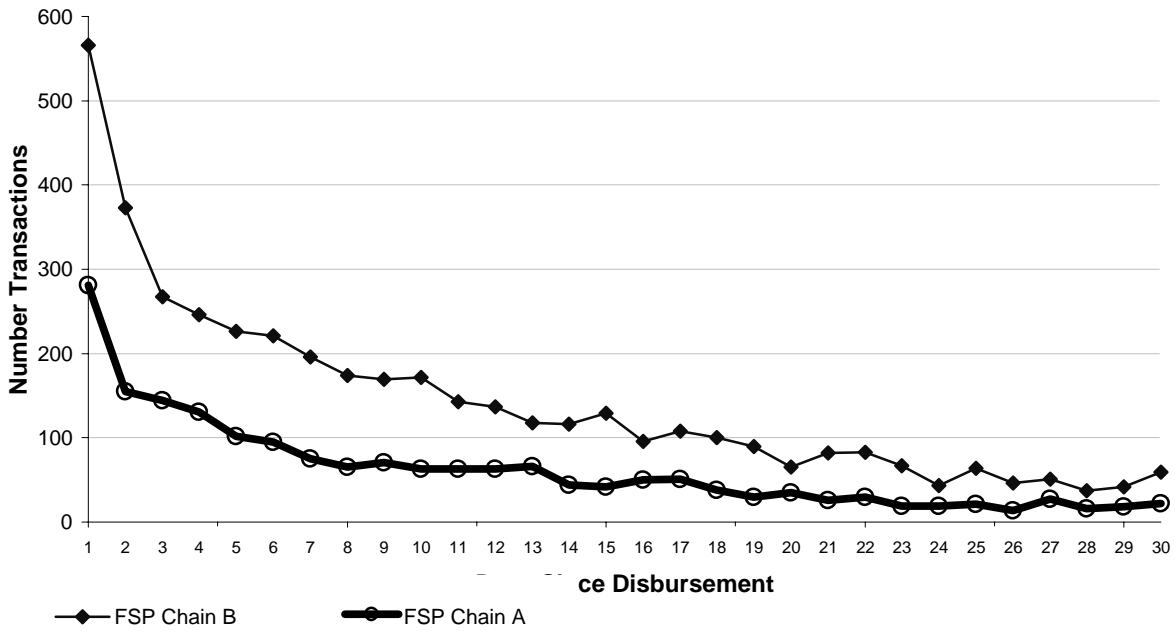
Store	FSP-Only Tender			Non-FSP Tender			Combination FSP and Non-FSP			
	Number baskets	Average size* (#items)	Average amount	Number baskets	Average size* (#items)	Average amount	Number baskets	Average size* (#items)	Average total amount	Average FSP amount
SEPTEMBER 1997										
Supermarkets										
Chain A #1	213	14.6	\$ 28.46	61,707	8.8	\$ 20.37	176	25.9	\$ 52.41	\$ 42.70
Chain A #2	967	14.1	\$ 26.89	27,565	7.3	\$ 17.50	532	22.0	\$ 40.73	\$ 34.56
Chain B #1	165	9.6	\$ 18.61				187	25.5	\$ 48.13	\$ 41.80
Chain B #2	936	9.0	\$ 17.16				590	27.9	\$ 51.92	\$ 47.86
Chain B #3	536	7.9	\$ 15.29				364	23.9	\$ 44.96	\$ 39.11
Chain B #4	333	9.8	\$ 19.03				188	21.9	\$ 39.90	\$ 32.83
Chain B #5	572	7.0	\$ 12.71				388	22.4	\$ 41.72	\$ 38.68
DCSU Stores										
Store C	28	3.4	\$ 2.67	1,078	2.9	\$ 3.58	4	4.3	\$ 8.36	\$ 4.70
Store D	106	5.1	\$ 6.38	499	2.8	\$ 3.25	75	6.9	\$ 8.51	\$ 7.92
Store E	4	3.5	\$ 6.10	312	3.6	\$ 5.95				
Store F	1	3.0	\$ 4.20	147	3.0	\$ 6.13	1	3.0	\$ 3.07	\$ 3.01
OCTOBER 1997										
Supermarkets										
Chain A #1	206	17.8	\$ 34.61	59,265	8.7	\$ 20.19	208	28.5	\$ 54.19	\$ 44.75
Chain A #2	1,035	13.1	\$ 24.17	27,975	7.4	\$ 17.97	497	23.8	\$ 43.36	\$ 38.12
Chain B #1	158	10.8	\$ 20.92				176	25.9	\$ 49.04	\$ 44.71
Chain B #2	901	8.5	\$ 16.52				723	25.7	\$ 48.67	\$ 44.40
Chain B #3	548	8.8	\$ 17.50				347	23.2	\$ 44.28	\$ 39.21
Chain B #4	294	8.5	\$ 16.69				201	20.6	\$ 39.76	\$ 31.76
Chain B #5	649	7.4	\$ 13.96				408	19.0	\$ 35.57	\$ 32.60
DCSU Stores										
Store C	14	3.6	\$ 3.50	341	4.5	\$ 5.52	2	4.5	\$ 3.68	\$ 3.39
Store D	116	5.0	\$ 6.40	128	2.5	\$ 3.31	127	6.6	\$ 7.12	\$ 6.61
Store E	115	6.6	\$ 10.61	10	3.2	\$ 24.03	6	7.2	\$ 15.76	\$ 10.94
Store F	8	2.8	\$ 4.93	138	3.0	\$ 5.45	10	5.3	\$ 10.31	\$ 8.57

* Data includes all captured baskets. FSP baskets are identified by successful match to the EBT (ALERT) log file of FSP transactions by South Carolina FSP households. The average basket size is measured as the number of entries on the cash register receipt. This measure approximates the number of items purchased.

Exhibit 5-7

**Number of Food Stamp and Non-Food Stamp Transactions at Sample Stores, By Day:
September 1997**

Number of Food Stamp Transactions, By "Days Since Disbursement"



Number of Non-Food Stamp Transactions, By Calendar Day

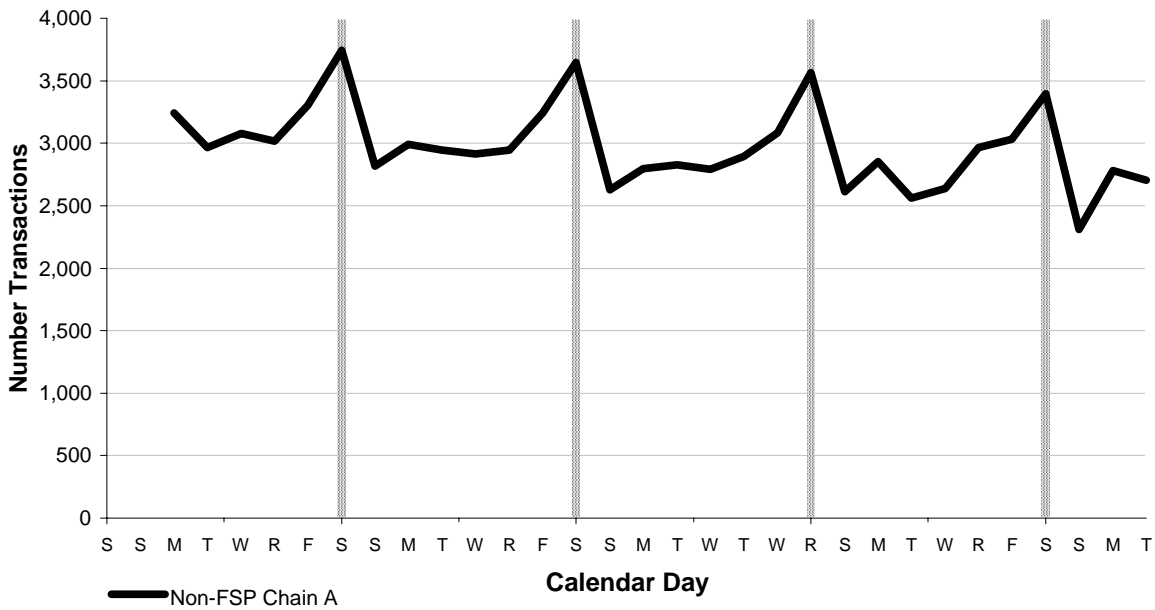
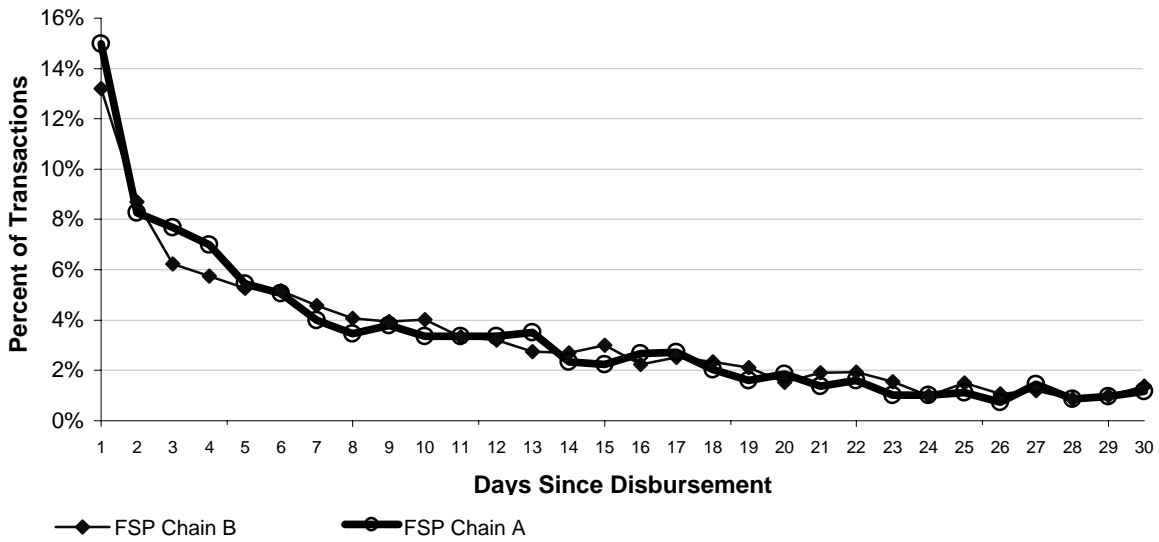


Exhibit 5-8

**Percent of Food Stamp and Non-Food Stamp Transactions at Sample Stores, By Day:
September 1997**

Distribution of Food Stamp Transactions, By "Days Since Disbursement"



Distribution of Non-Food Stamp Transactions, By Calendar Day

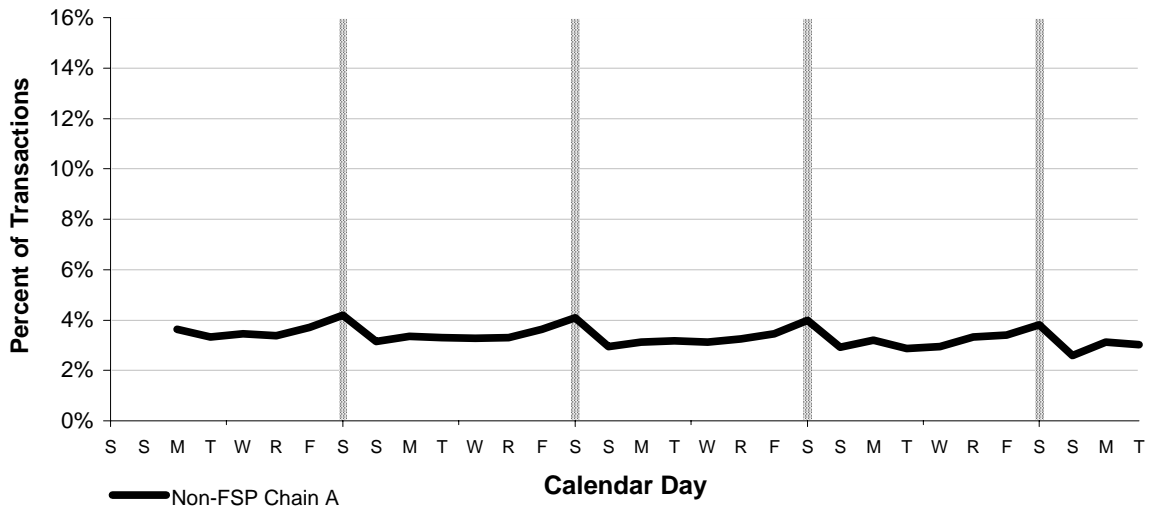


Exhibit 5-9

Average Number Items Per "Basket," by Day: Food Stamp and Non-Food Stamp Transactions at Sample Stores, September 1997

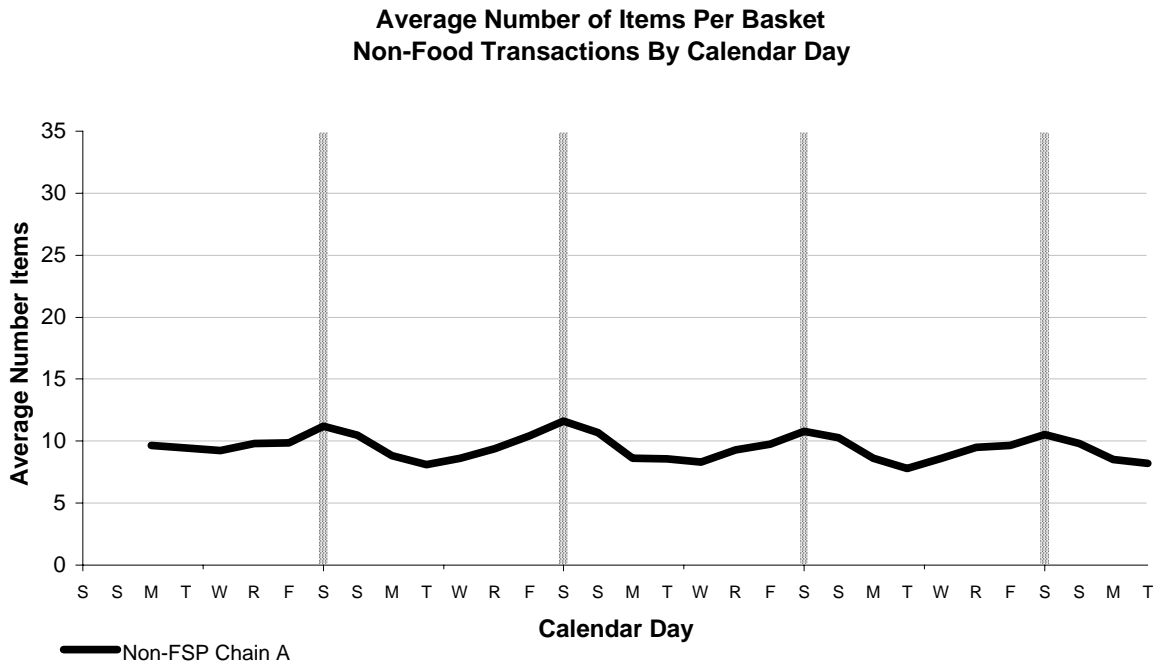
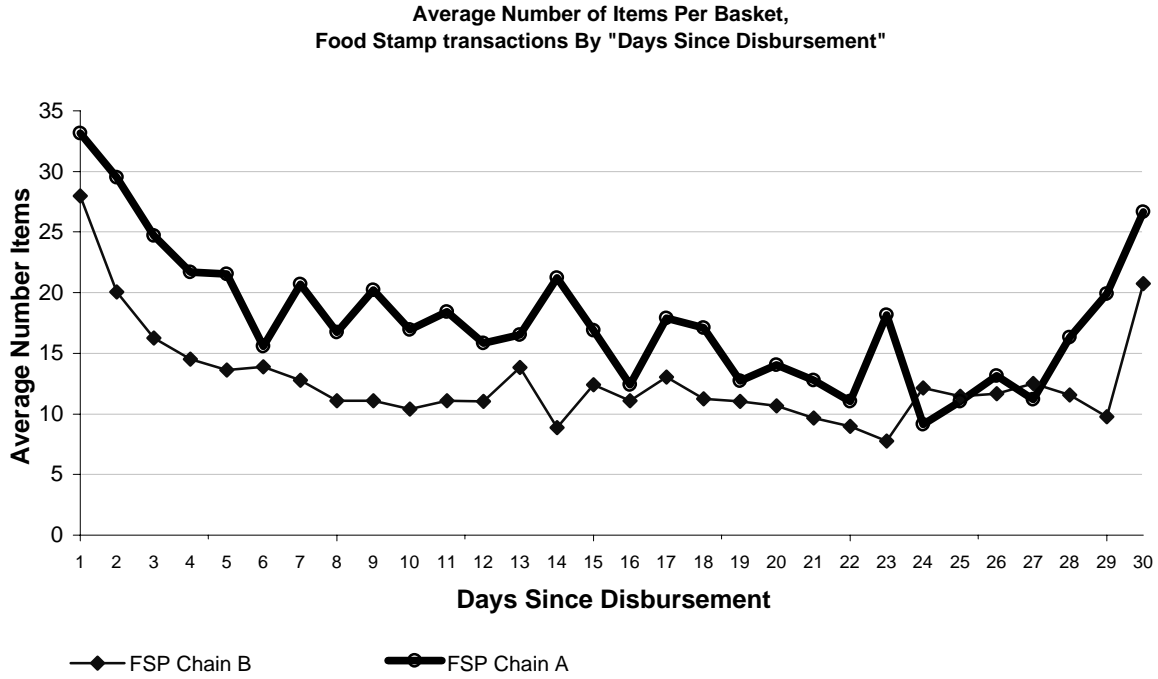
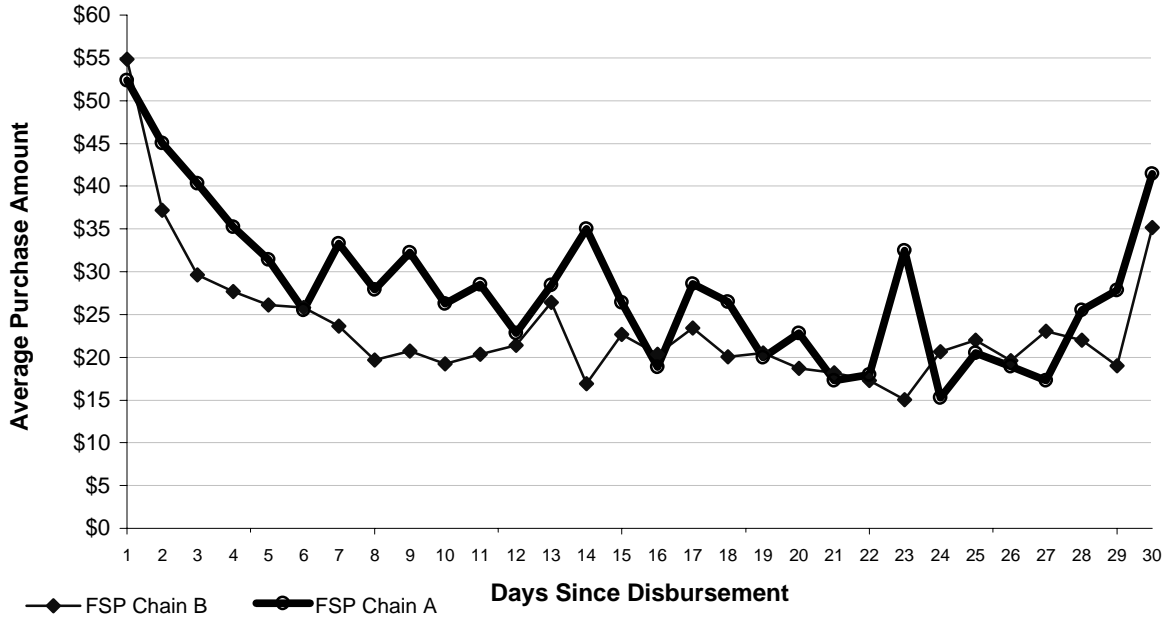


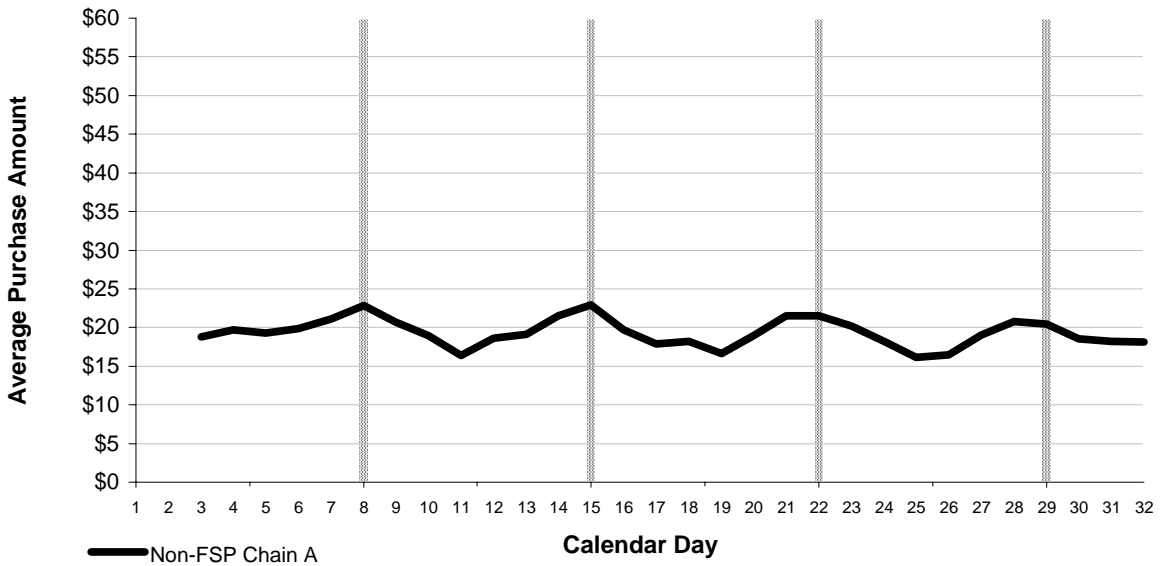
Exhibit 5-10

Average Basket Value (\$) by Day: Food Stamp and Non-Food Stamp Transactions at Sample Stores, September 1997

Average Transaction Purchase Amount, Food Stamp Transaction By "Days Since Disbursement"



Average Transaction Purchase Amount, Non-Food Stamp Transactions By Calendar Day



Exhibits 5-9 and 5-10 show that the average size of FSP transactions declines somewhat throughout the month, whereas the average size of non-FSP transactions displays a weekly “U-shaped” pattern that repeats throughout the month. The increase in basket size for FSP purchases at the end of the month may be an anomaly due to a limited number of observed transactions, or it could be due to an error in the identification of disbursement date for a small number of recipients.

5.4 Characterization of the Content of Food Baskets

We use two product categorization schemes to describe the composition of FSP and non-FSP food purchases in Exhibits 5-11 through 5-17: **department** and **product category**. Assignment of department and product category codes to individual UPC and PLU codes was described in Section 4.3.

Distribution of Spending on Items Purchased, by Store Department

The broad **department** category scheme characterizes food purchases primarily by the area of the supermarket where the item is stocked: meat, produce, dairy, deli, and frozen food define specific supermarket areas; the grocery department encompasses most “dry” food items.¹⁶ Because this scheme primarily reflects the layout of supermarkets, it is important to realize that items that are equivalent from a nutritional perspective may be found in multiple departments. For example, juices appear in the grocery, dairy, and frozen food departments; vegetables appear in produce, grocery, and frozen food departments.

Exhibit 5-11 compares the distribution of items purchased in FSP and non-FSP transactions at Chain A. We limit this analysis to Chain A transactions because we do not have a non-FSP sample from Chain B. Inclusion of FSP transactions from Chain B may confound the FSP/non-FSP differences with differences in product offerings across store chains.

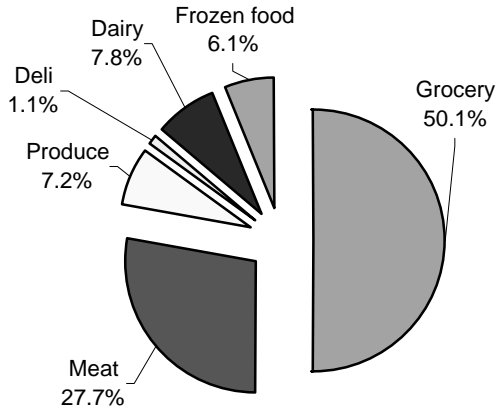
The top panel of Exhibit 5-11 compares FSP and non-FSP food baskets. FSP food baskets contain a somewhat higher percent of grocery and meat items, compared to non-FSP baskets, and FSP baskets contain lower percentages of frozen food, produce, dairy, and deli items. The bottom panel of Exhibit 5-11 shows this comparison of FSP/non-FSP food baskets for each of the two stores in Chain A.

Surprisingly, the distributions of FSP and non-FSP purchases across departments are very comparable **within** each store, whereas there are differences in the distribution of both FSP and non-FSP transactions **between** stores. For example, meat represents about 23 percent of FSP purchases and 20 percent of non-FSP purchases at Chain A Store #1, but it represents 29 percent and 26 percent of FSP and non-FSP purchases, respectively, at Chain A Store #2. Thus, the apparent difference in FSP/non-FSP baskets at the top of Exhibit 5-11 is a bit misleading. The differences in the distribution of spending, by department, across individual stores may reflect differences in the variety or quality of product offerings

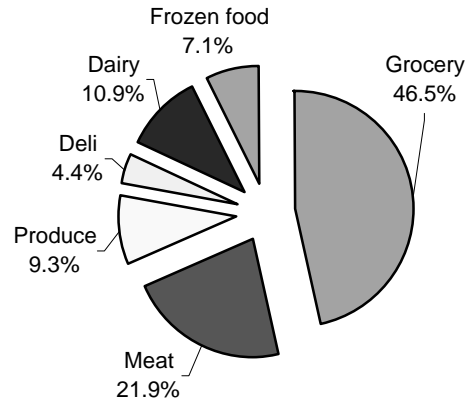
¹⁶ Non-food items are dropped from non-FSP transactions for the purpose of comparing the distribution of FSP and non-FSP purchases across product type. Non-food items are classed in the following departments: **non-food** (includes soaps, laundry supplies, household cleaners, household supplies, paper and plastic products, pet food); **general merchandise** (includes books, magazines, housewares, hardware); money orders and fees. We also classed tobacco, beer, and wine as non-food items.

**Exhibit 5-11
Distribution of FSP-eligible Items in Food Stamp and Non-Food Stamp Purchases, By
Department**

Food Stamp Transactions



Non-FSP Transactions



Note: Pie charts show distribution of purchases/redemptions across six department categories; this includes purchases/redemptions except those for non-food and general merchandise items.

**Distribution of FSP-eligible Items in Food Stamp and Non-Food Stamp Purchases, By
Department and Store**

DEPARTMENT	Chain A #1		Chain A #2	
	Food Stamp Transactions	Non-FSP Transactions	Food Stamp Transactions	Non-FSP Transactions
Grocery	47.8%	46.0%	50.8%	47.8%
Meat	22.8%	20.3%	29.1%	26.1%
Produce	8.4%	9.5%	6.8%	8.9%
Deli	2.4%	5.2%	0.7%	2.2%
Dairy	10.5%	11.4%	7.0%	9.4%
Frozen food	8.2%	7.7%	5.5%	5.5%
Total purchases/redemptions	\$ 19,946	\$ 1,653,725	\$ 69,480	\$ 606,626

Notes

Food Stamp transactions include all transactions paid for in whole or part with FSP tender. Non-food stamp transactions are transactions with no FSP tender. Transactions paid for with a combination of FSP tender and other tender (cash, check, coupons) may or may not include FSP-eligible items. To examine the distribution of items purchased with FSP benefits, we exclude FSP-eligible items from combination transactions and weight the item prices on all FSP-eligible items so that the purchase prices sum to the FSP redemption posted to the EBT log.

in departments of individual stores, which could bias the FSP/non-FSP comparison in the aggregate.¹⁷ In other words, if the distribution of FSP purchases across supermarket departments depends on the choice set of product offerings at individual stores, then it is important to observe FSP transactions at a representative sample of stores before drawing conclusions about the content of FSP food baskets.

Exhibit 5-12 shows the distribution of all captured FSP item purchases across department, by household type, including FSP transactions at Chain A and Chain B. FSP households are grouped as follows: households with elderly persons and no children, households with no elderly and no children, and households with children.¹⁸ Exhibit 5-12 reveals very little difference in the purchases of these types of households across the broad department categories, with one exception: households with elderly persons redeem a higher percentage of their FSP benefits on produce than other FSP households.

Percentage of Item Purchases on Store-Brand Items

Purchase of store-brand (i.e., generic) food items is one means by which FSP households can stretch their food stamp dollars. We identified store-brand items at Chain A and Chain B by coding items in the Master Item File according to the stores' manufacturer code, which appears as the first five digits of the UPC code.¹⁹ Among all items appearing in our Master Item File, of 33,810 items observed in transactions during the data collection period, 11 percent were store-brand items.

Exhibits 5-13 through 5-15 show the percent of FSP redemptions and non-FSP purchases spent on store-brand items. At Chain A, the percentage of FSP benefits spent on store-brand items was six percentage points greater than the percentage of non-FSP purchases spent on store-brand items. Exhibit 5-14 shows that most of the difference in store-brand purchases between FSP and non-FSP transactions is due to higher spending by FSP households on store-brand **grocery** items.²⁰ Exhibit 5-15 shows that there are some differences in spending on store-brand items within FSP households, by household type. FSP households with elderly persons purchase a higher percentage of store-brand items than other FSP households, especially in the grocery and frozen food departments.

Distribution of Item Purchases by Product Category

The second categorization used to describe the composition of food baskets is a taxonomy of food items based on food groups. This categorization scheme allows us to aggregate individual food purchases into categories that are much more narrowly defined than supermarket departments: there are ten departments for categorizing food items and 168 product categories. The assignment of product category codes to individual food items was described in Section 4-3.

17 Note that the FSP percentage of all transactions at Stores #1 and #2 is 1.2 and 10.3, respectively.

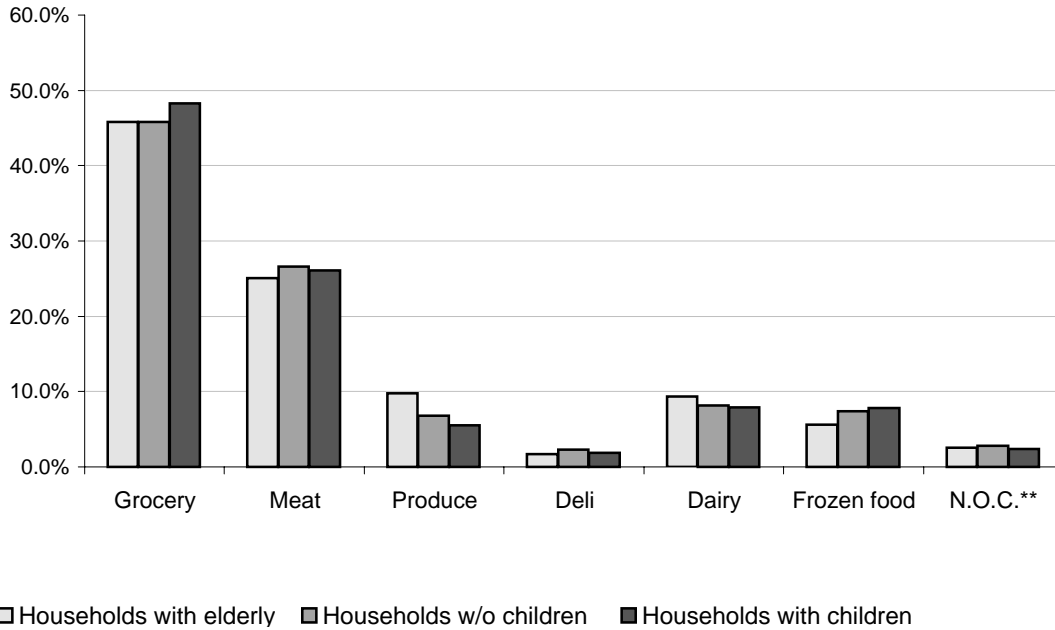
18 Four percent of "households with children" also include elderly persons.

19 We recognize that there are "generic" brands that are not store-brands. Identification of all "generic" items at Chains A and B was beyond the scope of this study.

20 Unfortunately, because we do not have a comparison non-FSP group at Chain B, we cannot draw conclusions about the percentage of spending on store brand items by FSP households at Chain B. Significant differences in the quality and breadth of store brand offerings across supermarket chains may lead to significant variation in store brand purchases by store chain.

Exhibit 5-12

Distribution of Food Stamp Redemptions, By Supermarket Department and Household Type: Redemptions at Seven Supermarkets in Study Site



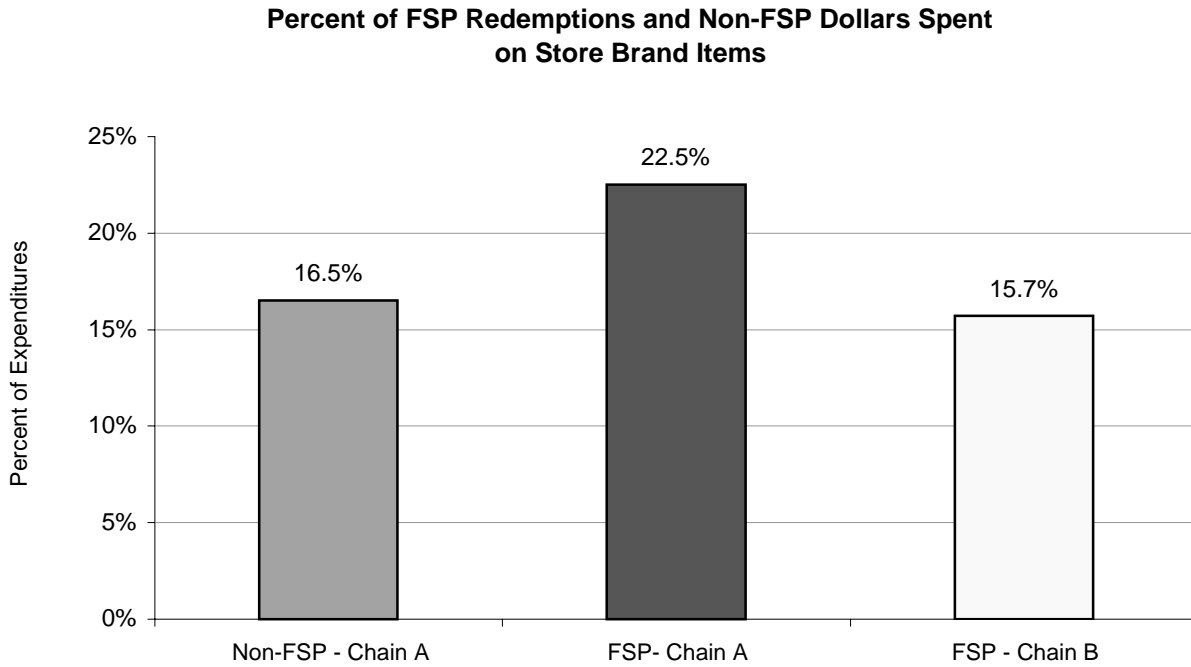
	Households with elderly	Households w/o children	Households with children	Unknown Household Type*	Total
DEPARTMENT					
Grocery	45.8%	45.8%	48.2%	47.3%	47.7%
Meat	25.0%	26.6%	26.1%	26.3%	26.1%
Produce	9.8%	6.8%	5.5%	6.4%	5.9%
Deli	1.7%	2.3%	1.9%	1.5%	1.9%
Dairy	9.4%	8.2%	7.9%	8.1%	8.0%
Frozen food	5.6%	7.4%	7.8%	7.8%	7.7%
N.O.C.**	2.5%	2.8%	2.4%	2.4%	2.4%
Total purchases/redemptions	13,981	42,724	244,291	32,465	333,461
Number FSP households	267	535	1,577	325	2,704

Notes: Sample includes FSP redemptions by South Carolina FSP households at seven supermarkets in Georgetown County during September and October 1997.

*Household type is from the South Carolina FSP administrative database of active FSP cases during October 1997. Household type is not known for FSP cases redeeming benefits in September or October 1997 but not receiving a disbursement in October.

** N.O.C. means not otherwise classified.

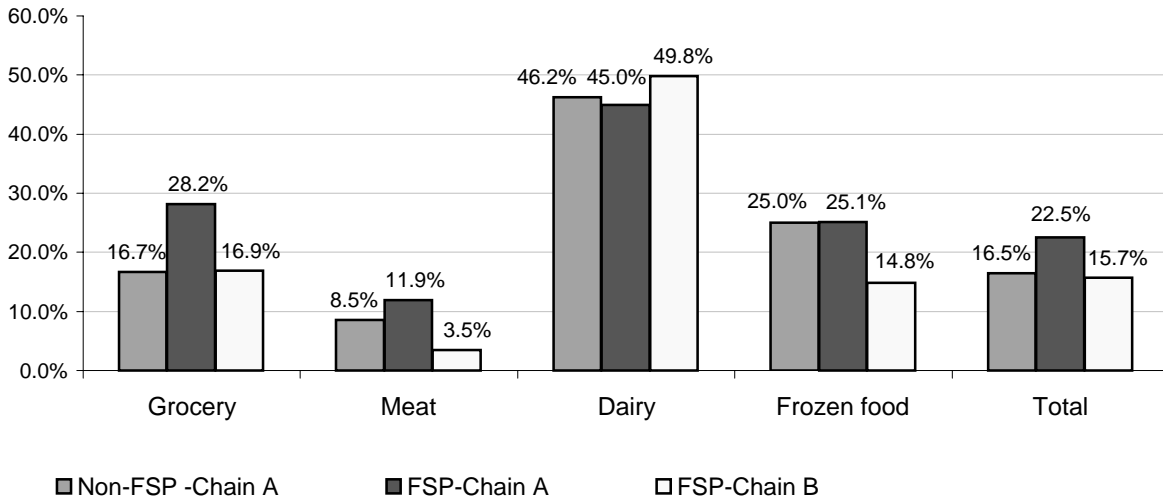
Exhibit 5-13
Percent of FSP and Non-Food Stamp Expenditures on Store Brand Items



Note: Percent of purchases on store-brand items is measured over FSP-eligible product categories.

Exhibit 5-14

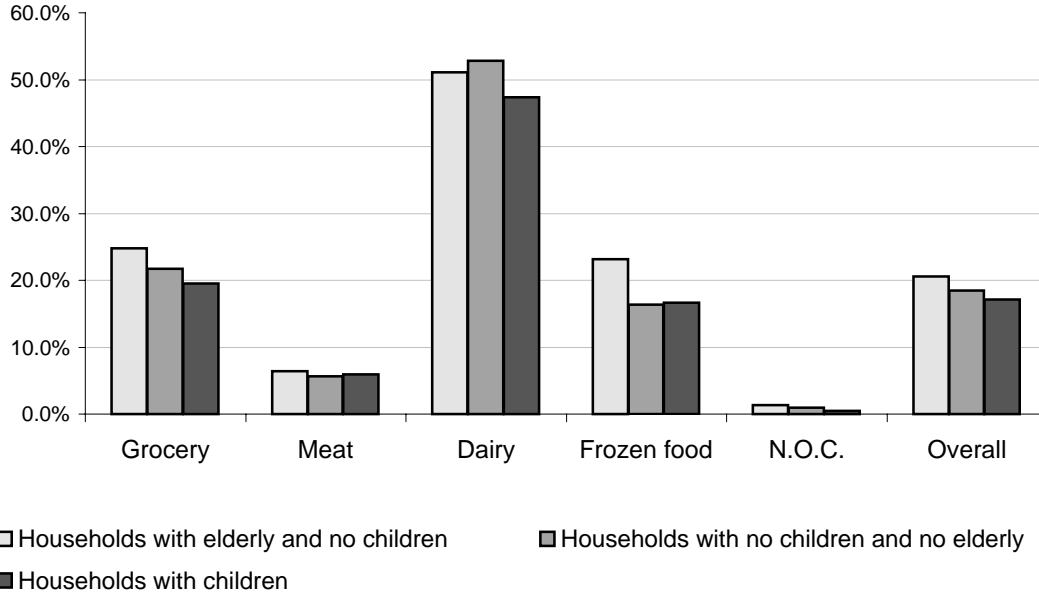
Store Brand Percent of Food Stamp Redemptions and Non-Food Stamp Purchases, By Supermarket Department



Notes

Food Stamp transactions include all transactions paid for in whole or part with FSP tender. Non-food stamp transactions are transactions with no FSP tender.

Exhibit 5-15
Store Brand Percent of Food Stamp Redemptions, By Household Type and Supermarket Department



Percent of Redemptions on Store Brand Items	Households with elderly and no children	Households with no children and no elderly	Households with children	Unknown Household Type*	Total
Overall	20.5%	18.4%	17.2%	18.0%	17.5%
By Department					
Grocery	24.7%	21.8%	19.5%	20.6%	20.1%
Meat	6.4%	5.7%	5.9%	6.2%	5.9%
Dairy	51.1%	52.8%	47.4%	50.0%	48.5%
Frozen food	23.1%	16.4%	16.6%	18.9%	17.1%
N.O.C.	1.3%	1.0%	0.5%	0.5%	0.6%
Total Redemptions	13,981	42,724	244,291	32,465	333,461
Number FSP households	267	535	1,577	325	2,704

Notes: Sample includes FSP redemptions by South Carolina FSP households at seven supermarkets in Georgetown County during September and October 1997.

Household type is from the South Carolina FSP administrative database of active FSP cases in during October 1997.

Household type is not known for FSP cases that redeemed benefits in September or October 1997 but did not receive a disbursement in October.

Exhibit 5-16 compares the content of FSP and non-FSP food baskets purchased at Chain A, by product category. (FSP-ineligible items are dropped from non-FSP transactions for the purpose of calculating this distribution.) Recall that FSP food baskets contain a higher percentage of grocery and meat items than non-FSP food baskets (see Exhibit 5-11). The distribution of items by product category shows that the difference in the meat department is due to higher expenditures on red meat and miscellaneous meat products (meat, NFS) by FSP households. The difference in the grocery department is due to higher expenditure by FSP households on grain-based prepared foods, flour, rice and pasta, fats and oils, sugar and candy, and beverages.

Exhibit 5-17 shows the distribution of FSP redemptions across product category by FSP household type. FSP households with elderly persons spend nearly twice as much of their FSP allotment on fruit than other households, and they spend two percentage points more on vegetables than households with children. All households spend the highest percentage of their allotment on red meat. The next most important product category, in terms of expenditures, for households with children is soft drinks; for households without children it is lunch meat; and for households with elderly it is vegetables.

Distribution of Purchases of Specific Items, by Quantity Size

As discussed above, purchase of store-brand or generic grocery items is one way in which FSP households may maximize the value of their FSP benefits, because generic or store-brand items typically have lower unit prices than branded items. Similarly, FSP households may maximize the purchasing power of their FSP allotment by choosing larger commodity sizes, which typically have lower unit prices.

Package size information was obtained from Chain A as part of their host price file (see Section 4.3), but these data were not comprehensive and the information was not standardized.²¹ We cleaned the package size information for a selection of staple food items: milk, sugar, flour, and breakfast cereals. Exhibits 5-18 through 5-21 show the distribution of food purchases by product size, comparing FSP and non-FSP purchases at Chain A, and FSP purchases at Chain A and Chain B across household type. Evidence from milk, sugar, flour, and breakfast cereal purchases at Chain A reveals that FSP households purchase larger item sizes than non-FSP households.

Milk

The difference between FSP and non-FSP purchases by commodity size is especially pronounced for milk purchases: 70 percent of FSP purchases of milk are for gallon sizes, whereas less than 50 percent of milk purchases of non-FSP households are for gallon sizes. (The distribution is measured over dollar expenditures.)

Sugar

Most sugar purchases, by FSP and non-FSP households, are for a 5-pound package; but nearly 14 percent of sugar purchases by FSP households are for the 10-pound package, compared to 2.4 percent of

21 The single largest problem in using the package size data was that the data field contained a mix of numeric and alphabetic data to denote the package size and the units of measurement, but the units of measurement information was not standardized. For example, we might see “10 ounces” or “10 oz” or “10 z.” A considerable effort would be necessary to clean these data for all items in the master item file.

Exhibit 5-16

Distribution of Food Stamp and Non-Food Stamp Item Purchases, By Product Category

	Food Stamp Redemptions	Non-FSP Transactions
MILK, CHEESE, CREAM	4.8%	7.8%
Milk	2.4%	3.9%
Cheese	2.1%	3.3%
Cream	0.1%	0.3%
Yogurt	0.2%	0.3%
FRUIT AND VEGETABLES	12.2%	14.9%
Fruit	3.6%	4.4%
Juice	1.3%	2.4%
Vegetables	7.4%	8.1%
PREPARED FOODS	9.6%	9.0%
Frozen/Refrigerated	2.4%	2.4%
Deli/Salad bar	0.5%	1.4%
Grain Based	4.2%	2.1%
Meat Based	0.4%	0.4%
Soup	1.0%	1.3%
Diet	0.4%	0.3%
Baby Food	0.8%	1.0%
MEAT AND MEAT ALTERNATIVES	32.2%	26.8%
Red meat	12.2%	10.5%
Poultry	3.2%	3.8%
Fish	2.2%	2.3%
Lunch meats	6.1%	5.4%
Eggs	0.8%	0.9%
Legumes	0.2%	0.1%
Nuts	0.7%	0.8%
Meat, NFS*	6.8%	3.1%
GRAIN PRODUCTS	10.9%	10.6%
Breakfast cereals	3.3%	2.5%
Flour, rice, pasta	3.5%	1.8%
Bread	1.7%	3.1%
Bakery products	2.3%	3.2%
OTHER FOODS	30.3%	31.0%
Fats, oils	3.0%	2.5%
Sugar, candy	4.7%	3.7%
Seasonings	1.3%	0.7%
Soft drinks, ades	7.9%	6.3%
Coffe, tea	1.1%	2.4%
Salty snacks	2.1%	2.2%
Cookies, cakes, pies	4.9%	5.4%
Ice cream, jello, pudding	1.5%	2.5%
Condiments	2.7%	2.5%
NFS	1.2%	2.8%
Total purchases/redemptions	89,426	2,260,366

Notes: Table excludes items in the following broad product categories; Non-food, General Merchandise, Fees.

See notes to Exhibit 10 for definition of Food Stamp and Non-Food Stamp purchases.

* NFS means not further specified.

Exhibit 5-17

Distribution of Food Stamp Redemptions, By Product Category and Household Type

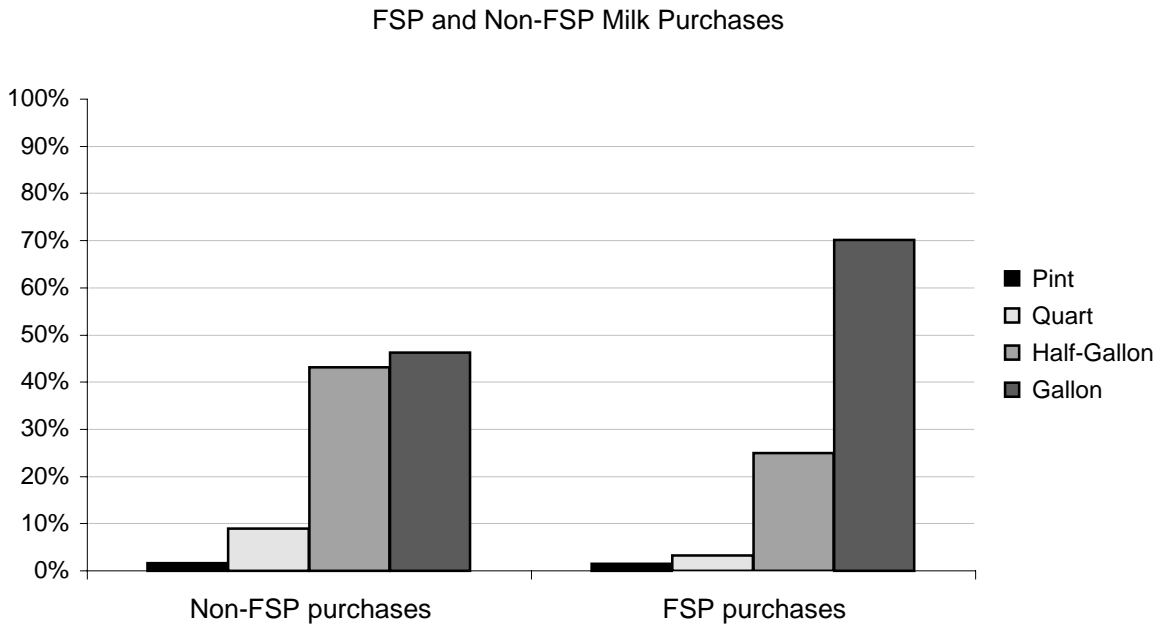
Product Category	Households with elderly	Households w/o children	Households with children	Unknown	Total
MILK, CHEESE, CREAM	5.6%	5.4%	4.9%	5.1%	5.0%
Milk	3.3%	2.7%	2.4%	2.4%	2.5%
Cheese	2.0%	2.5%	2.3%	2.5%	2.3%
Cream	0.2%	0.1%	0.1%	0.1%	0.1%
Yogurt	0.1%	0.1%	0.1%	0.2%	0.1%
FRUIT AND VEGETABLES	16.9%	12.7%	10.4%	11.4%	11.1%
Fruit	6.0%	3.6%	3.0%	3.6%	3.3%
Juice	2.9%	1.8%	1.4%	1.2%	1.5%
Vegetables	8.0%	7.3%	6.1%	6.7%	6.4%
PREPARED FOODS	7.8%	8.8%	10.4%	9.8%	10.0%
Frozen/Refrigerated	1.8%	2.6%	2.8%	2.8%	2.8%
Deli/Salad bar	0.5%	0.4%	0.4%	0.2%	0.4%
Grain Based	2.9%	3.5%	4.4%	4.2%	4.2%
Meat Based	0.3%	0.2%	0.5%	0.5%	0.5%
Soup	1.4%	1.1%	1.0%	1.1%	1.0%
Diet	0.7%	0.3%	0.3%	0.2%	0.3%
Baby Food	0.1%	0.7%	0.9%	0.9%	0.9%
MEAT AND MEAT ALTERNATIVES	31.1%	33.1%	30.9%	30.7%	31.2%
Red meat	11.6%	14.0%	13.6%	14.5%	13.6%
Poultry	4.4%	3.9%	3.9%	3.7%	3.9%
Fish	1.6%	1.6%	1.2%	1.6%	1.3%
Lunch meats	7.7%	7.7%	7.2%	6.7%	7.2%
Eggs	1.3%	0.8%	0.7%	0.6%	0.7%
Legumes	0.3%	0.3%	0.2%	0.2%	0.2%
Nuts	0.6%	0.7%	0.6%	0.7%	0.6%
Meat, NFS*	3.7%	4.2%	3.6%	2.8%	3.6%
GRAIN PRODUCTS	10.9%	9.0%	10.6%	10.5%	10.4%
Breakfast cereals	2.7%	2.2%	3.3%	3.6%	3.2%
Flour, rice, pasta	4.5%	2.8%	3.2%	3.0%	3.2%
Bread	2.0%	2.0%	1.9%	1.8%	1.9%
Bakery products	1.7%	2.0%	2.3%	2.1%	2.2%
OTHER FOODS	27.8%	30.8%	32.7%	32.4%	32.2%
Fats, oils	3.4%	2.8%	2.9%	3.2%	2.9%
Sugar, candy	4.2%	4.2%	4.5%	4.6%	4.4%
Seasonings	1.5%	1.0%	1.1%	1.4%	1.1%
Soft drinks, ades	6.3%	7.2%	8.5%	7.8%	8.2%
Coffe, tea	1.6%	1.8%	1.0%	1.3%	1.1%
Salty snacks	1.4%	2.2%	3.0%	2.5%	2.8%
Cookies, cakes, pies	3.4%	3.9%	4.2%	4.0%	4.1%
Ice cream, jello, pudding	1.6%	1.9%	1.9%	1.9%	1.9%
Condiments	1.5%	2.7%	3.0%	2.8%	2.9%
NFS	2.9%	3.1%	2.7%	2.8%	2.8%
Total purchases/redemptions	13,981	42,724	244,291	32,465	333,461

Notes: Table excludes items in the following broad product categories; Non-food, General Merchandise, Fees.

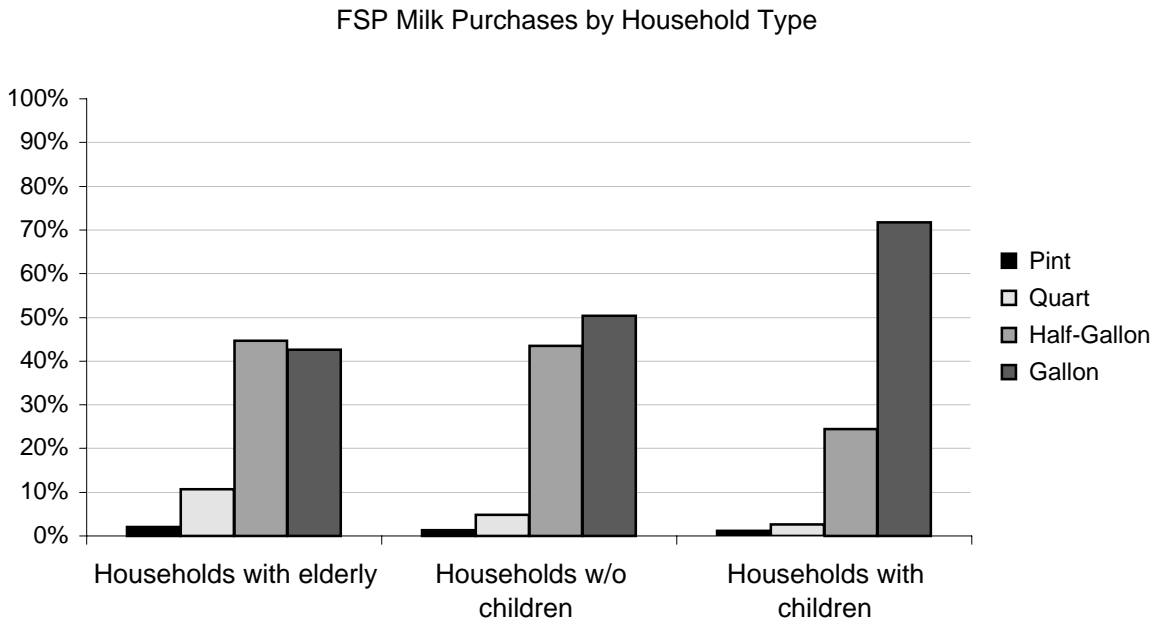
See notes to Exhibit 10 for definition of Food Stamp and Non-Food Stamp purchases.

* NFS means not further specified.

Exhibit 5-18
Distribution of Milk Purchases By Quantity Size: Food Stamp and Non-Food Stamp Purchases

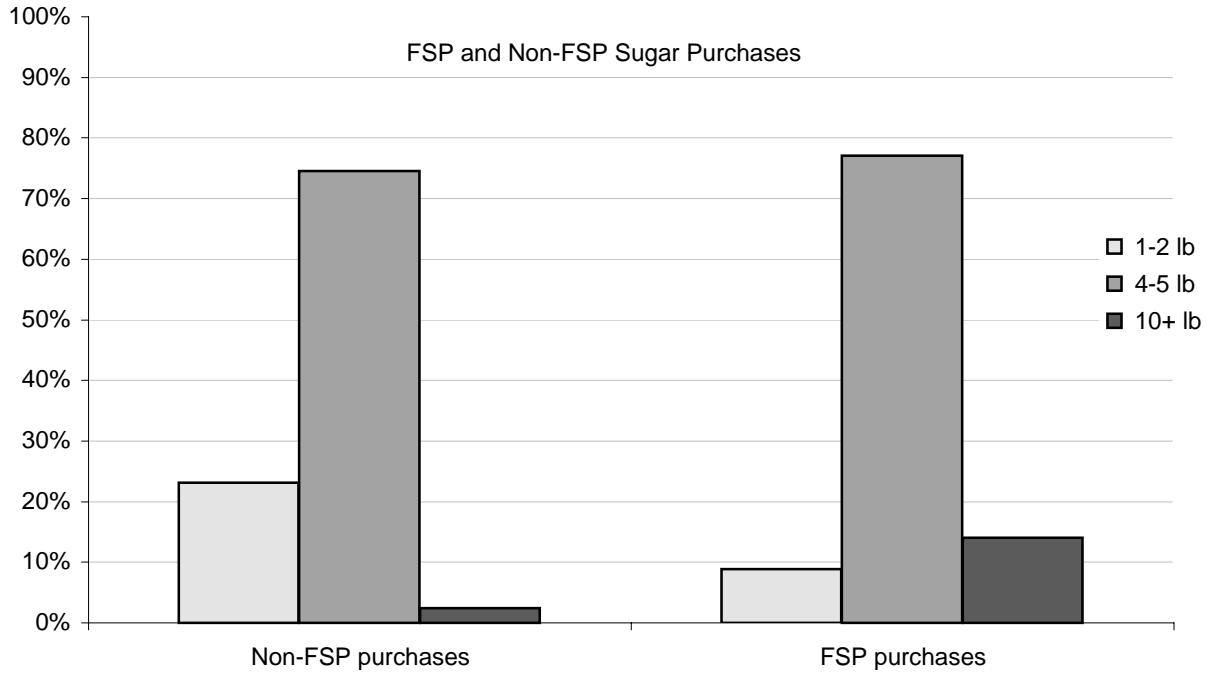


Note: Data for FSP and Non-FSP purchases are from two supermarkets in Georgetown County.

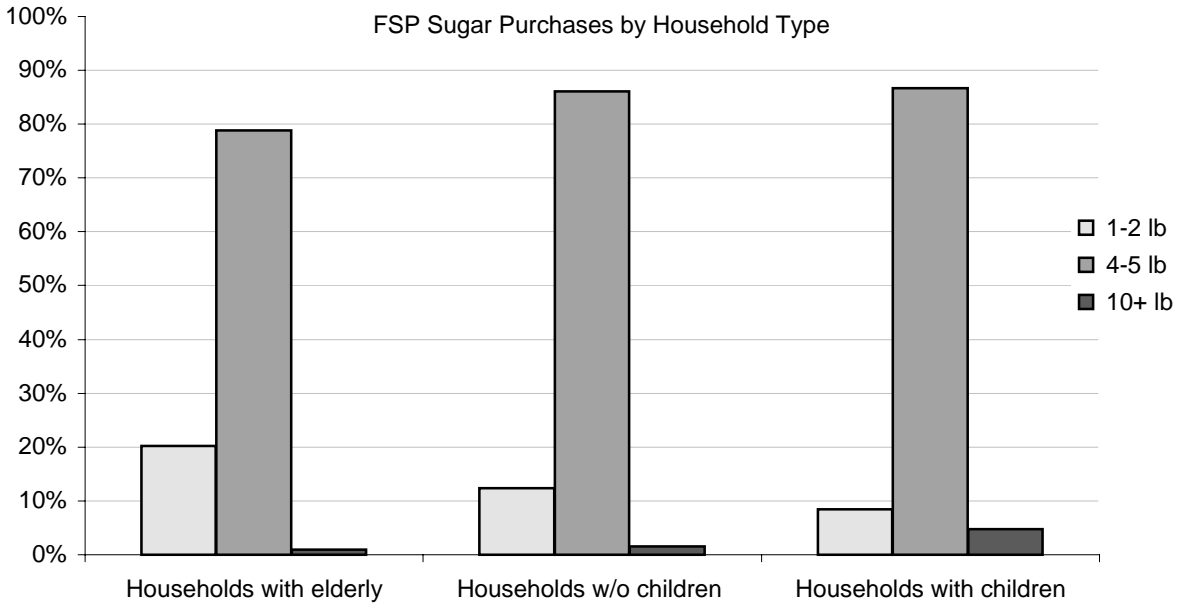


Note: Data for FSP purchases are from seven supermarkets in Georgetown County.

Exhibit 5-19
Distribution of Sugar Purchases By Quantity Size: Food Stamp and Non-Food Stamp Purchases

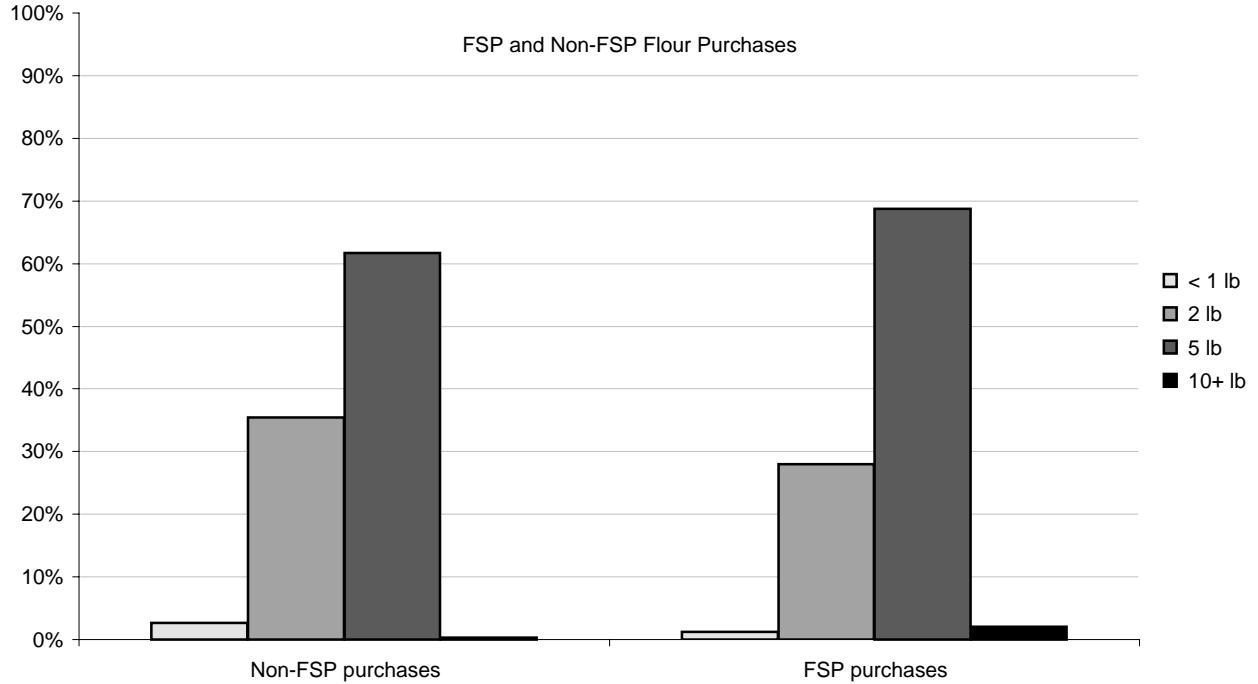


Note: Data for FSP and Non-FSP purchases are from two supermarkets in Georgetown County.

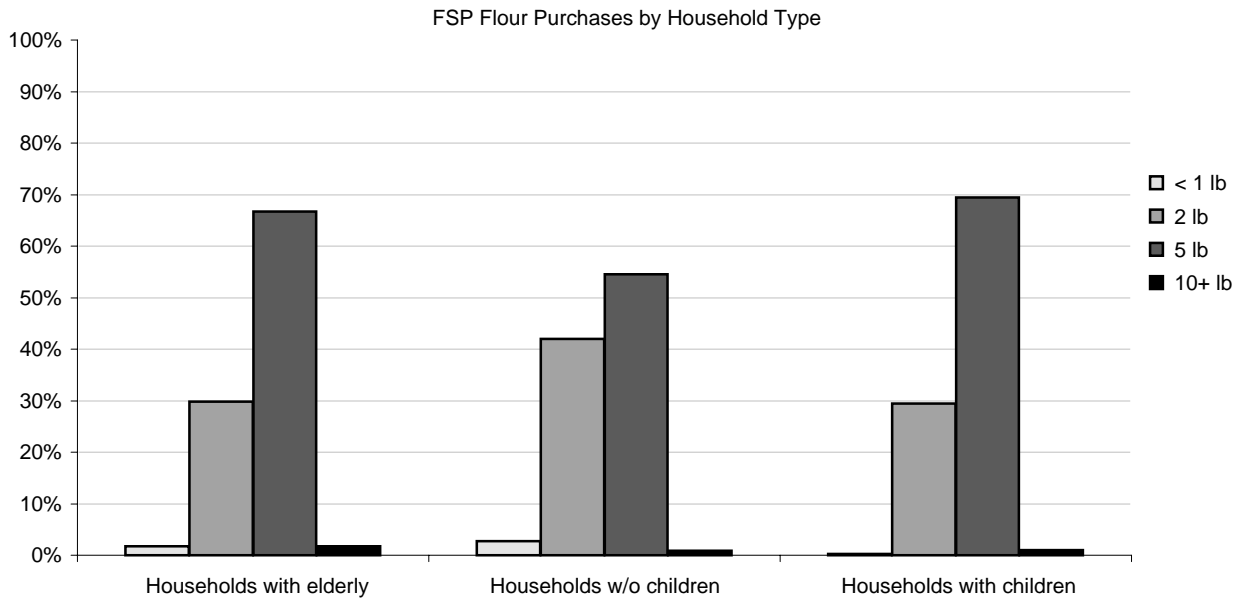


Note: Data for FSP purchases are from seven supermarkets in Georgetown County.

Exhibit 5-20
Distribution of Flour Purchases By Quantity Size: Food Stamp and Non-Food Stamp Purchases



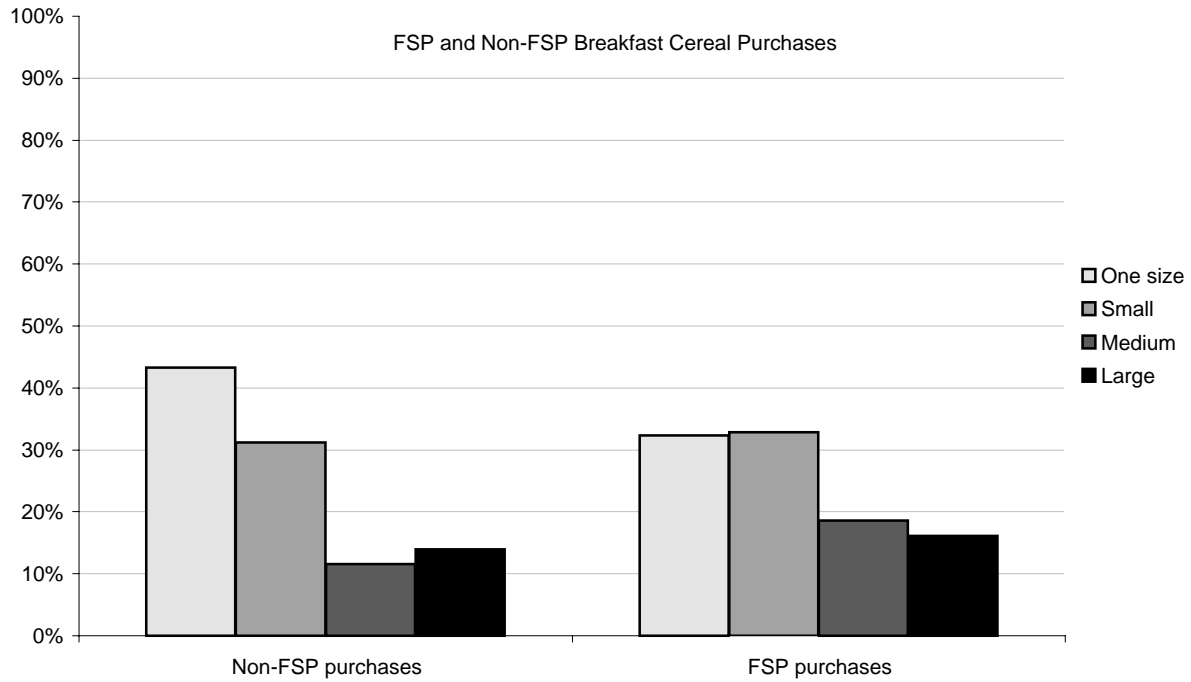
Note: Data for FSP and Non-FSP purchases are from two supermarkets in Georgetown County.



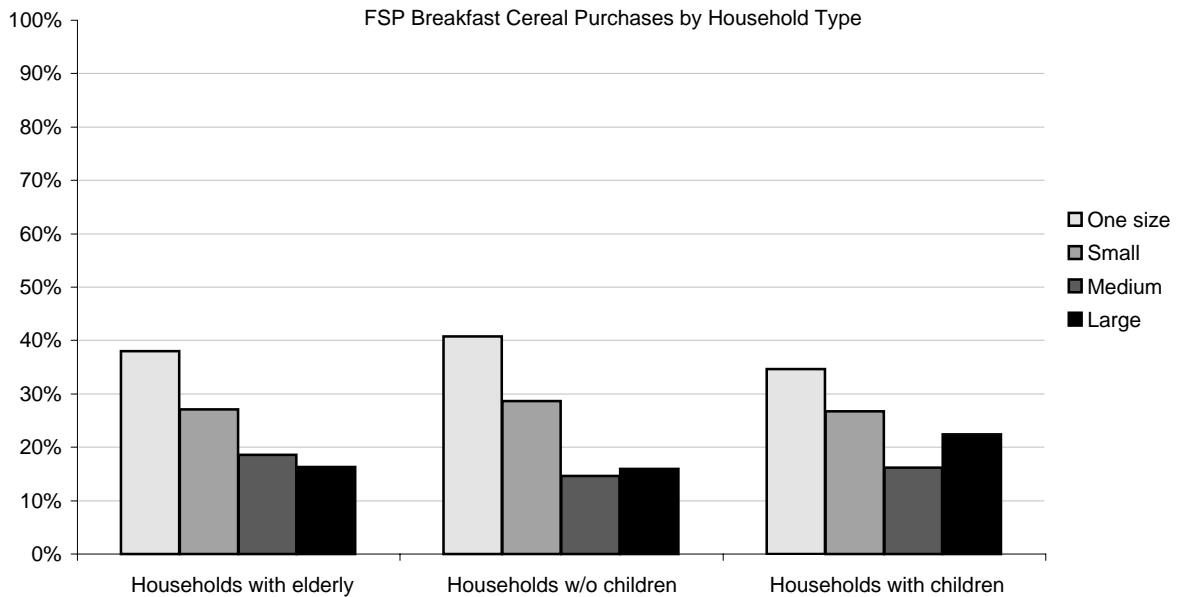
Note: Data for FSP purchases are from seven supermarkets in Georgetown County.

Exhibit 5-21

Distribution of Breakfast Cereal Purchases By Quantity Size: Food Stamp and Non-Food Stamp Purchases



Note: Data for FSP and Non-FSP purchases are from two supermarkets in Georgetown County.



Note: Data for FSP purchases are from seven supermarkets in Georgetown County. Breakfast cereals include "ready-to-eat" cereal only. "Single size" refers to cereal brands available in only one package size at the sample stores.

non-FSP purchases. Finally, less than 10 percent of FSP purchases of sugar are on the 2-pound package, whereas 23 percent of non-FSP purchases are for the 2-pound package.

Flour

Flour purchases show a pattern that is similar to sugar purchases: FSP households are somewhat more likely to purchase the 5-pound size, and less likely to purchase the 2-pound size. There are few purchases of flour in a 10-pound size by FSP or non-FSP households.

Breakfast Cereal

The tabulation of breakfast cereals by package size is complicated by the fact that different brands of cereal are packaged in different sizes. For example, flakes are a lightweight cereal, whereas raisin bran is a heavy cereal. A medium-size box of flakes weighs considerably less than a medium-sized box of raisin bran. We could not, therefore, characterize a “small” package size according to the weight of the box, because “small” means different things to different brands. To represent the choices that shoppers face in the breakfast cereal aisle, we sorted all breakfast cereals by cereal brand and package size and, if a brand had three sizes, we coded small, medium, and large in order of package size. If a brand had only two sizes, we coded the two packages as small and large; if a brand had only one size, we coded the cereal as “one size.” Exhibit 5-21 shows that FSP households are more likely to purchase medium and large size breakfast cereal packages and less likely to purchase one-size brands of cereal, compared to non-FSP households. Among FSP households, households with children spend a larger percentage of cereal purchases on large size packages than other households.

The differences in commodity size purchased by FSP and non-FSP households may be due to the fact that FSP households seek the economy of larger item sizes, or it may simply reflect differences in average household size between FSP and non-FSP households. Ideally, in determining whether or not FSP households are more likely to purchase larger commodity sizes, we would like to control for determinants of demand such as household size; household characteristics, however, cannot be identified in the non-FSP transaction data from Chain A.

The bottom panels of Exhibits 5-18 through 5-21 show that, within FSP households, households with children purchase larger item sizes than other FSP household types. The differences by household composition, within the FSP sample, suggests that at least some portion of the FSP/non-FSP difference in purchases by commodity size may be due to differences in the distributions of household size within the FSP and non-FSP samples.

Distribution of Purchases by Nutritional Characteristics

In comparing the composition of the food baskets of FSP and non-FSP households, a final dimension of choice that we examine is the choice of nutritional quality. Recall, however, that there are over 33,000 different grocery items in our Master Item File. For obvious reasons, we do not attempt an analysis of the nutrient composition of purchased food baskets. Instead, we examine the nutritional choices that FSP households make when purchasing two specific items: milk and breakfast cereal.

A comparison of milk purchased by FSP households and non-FSP households at Chain A shows that over 70 percent of FSP redemptions for milk were for whole milk, and only 20 percent were for low-fat

or skim milk (see Exhibit 5-22). In contrast, among non-FSP purchases, only 32 percent of purchases were for whole milk, and over 60 percent were for low-fat or skim milk.

The distribution of FSP milk purchases by household type shows that households with children purchase the highest percentage of whole milk: 77 percent, versus 57 percent for households without children and 53 percent for households with elderly persons. Only 33 percent of non-FSP purchases of milk, however, are for whole milk; whereas, across all FSP households, over 50 percent of FSP redemptions for milk are for whole milk. Therefore, the FSP/non-FSP differences in milk purchase cannot be attributed totally to possible differences in the proportion of households with children in the FSP and non-FSP populations.

The second item for which we examine nutrient content is breakfast cereals. Nutritional information for individual brands of breakfast cereal are readily available from manufacturers in a concise, published format. We merged nutrient data and serving sizes (basically, the information that appears on the box) to our Master Item File.²² Because cereals are not a homogenous product—like sugar, flour, and milk—the choice of nutritional content is made simultaneously with a choice about product price. Therefore, to examine the distribution of cereal purchases by nutrition content, we did not tabulate cereal expenditures—we tabulated the number of servings purchased.

Exhibits 5-23 through 5-25 show the distribution of breakfast cereal servings purchased by FSP and non-FSP households, by sugar content, iron content, and fat content. On average, food stamp households purchase breakfast cereals with higher sugar content, lower iron content, and higher fat content, compared with non-FSP households. The difference between FSP and non-FSP purchases of cereal is most pronounced for sugar content: 79 percent of cereal servings purchased by FSP households are high sugar (greater than 10 grams of sugar per serving), compared to 57 percent of non-FSP purchases of cereal servings.

The bottom panels of Exhibits 5-23 through 5-25 show the distribution of breakfast cereal servings purchased by the three types of FSP households. There are marked differences in the breakfast cereal purchases of elderly households versus other FSP households; households with elderly persons purchase breakfast cereals that are lower in sugar, higher in iron content, and lower in fat, compared to purchases of other FSP households.

22 Nutrition information was obtained from General Mills, Kellogg's, and Quaker Oats. Chains A and B each stocked a wide variety of cereals under their store brand. Package sizes for Chain B's store brand were obtained from the host price file of a supermarket outside the study that carried the same generic brand. The store brand cereals in Chains A and B were matched to nutrient information obtained off the boxes of store-brand cereals in two store chains in the Boston area. The match of store brands was possible because the cereals at Chain A and Chain B were uniquely different sizes. For example, Chain A corn flakes was 28.4 ounces, whereas Chain B corn flakes was 31 ounces; Chain A raisin bran was 55 ounces, and Chain B raisin bran was 54 ounces. The package sizes, within a chain, matched the package sizes within Boston chains exactly.

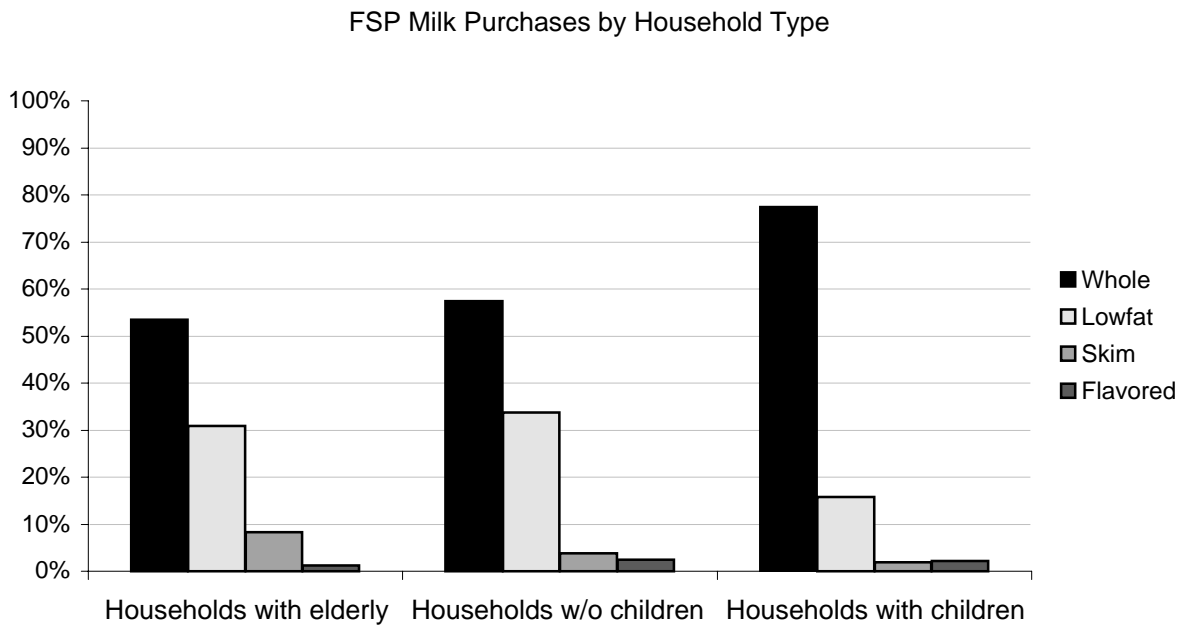
Nutrition information for individual cereal brands was matched to all occurrences of that brand in the Master Item File. A cereal will appear in the Master Item File multiple times because each package size has a unique UPC code.

Exhibit 5-22

Distribution of Milk Purchases By Milk Type: Food Stamp and Non-Food Stamp Purchases



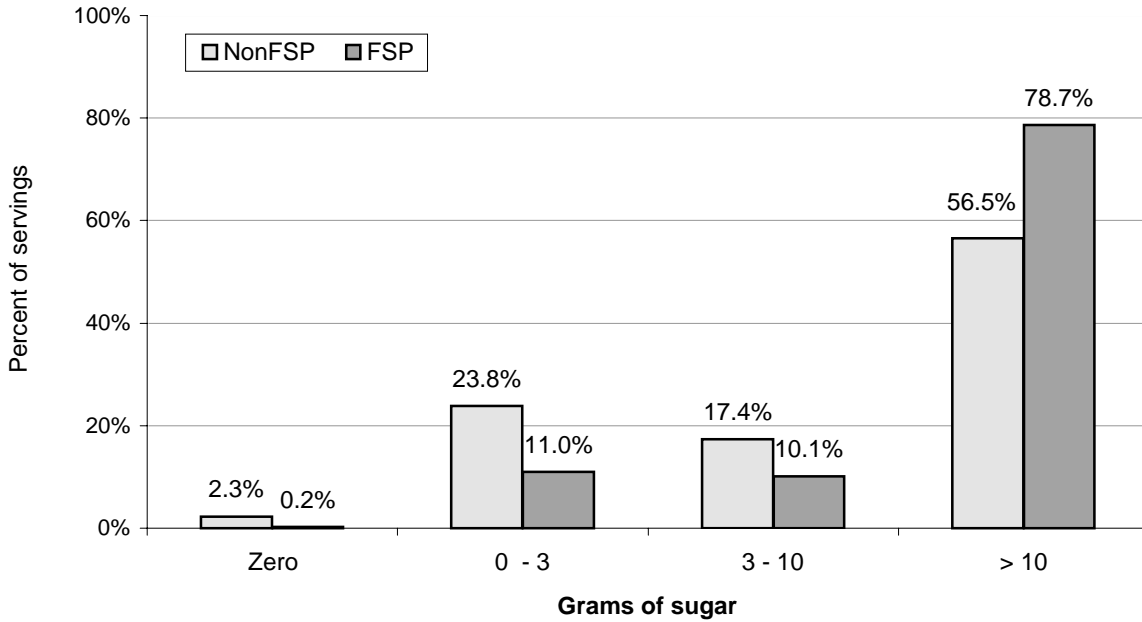
Note: Data for FSP purchases are from seven supermarkets in Georgetown County; data for non-FSP purchases are from two supermarkets in Georgetown County.



Note: Data for FSP purchases are from seven supermarkets in Georgetown County.

Exhibit 5-23

**Distribution of Purchased Breakfast Cereal Servings By SUGAR Content:
Food Stamp and Non-Food Stamp Purchases**



**Distribution of Food Stamp Purchases of Breakfast Cereal Servings,
By SUGAR Content and Household Type**

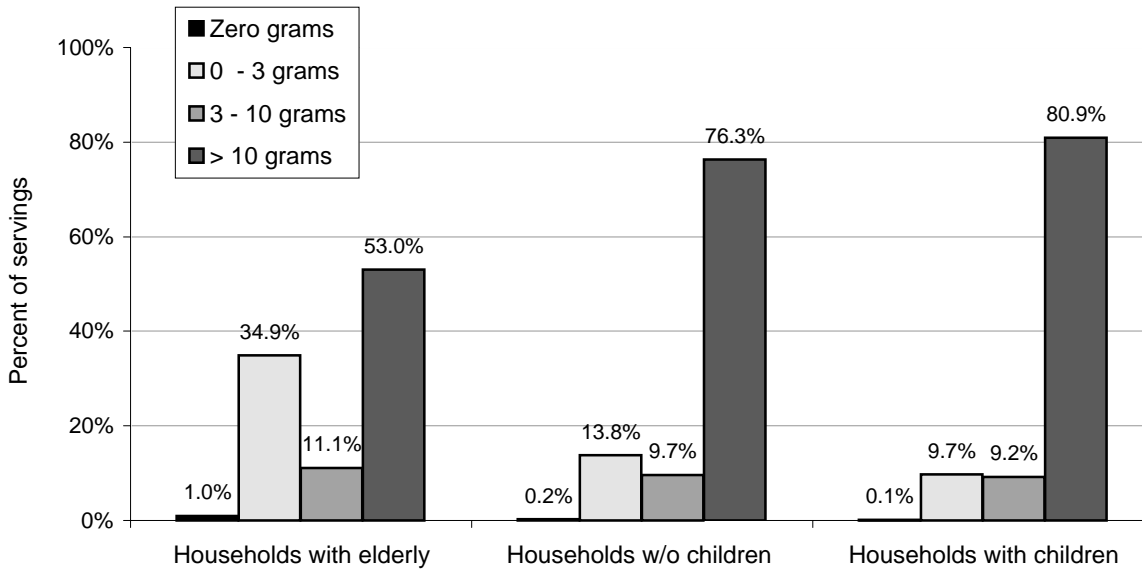
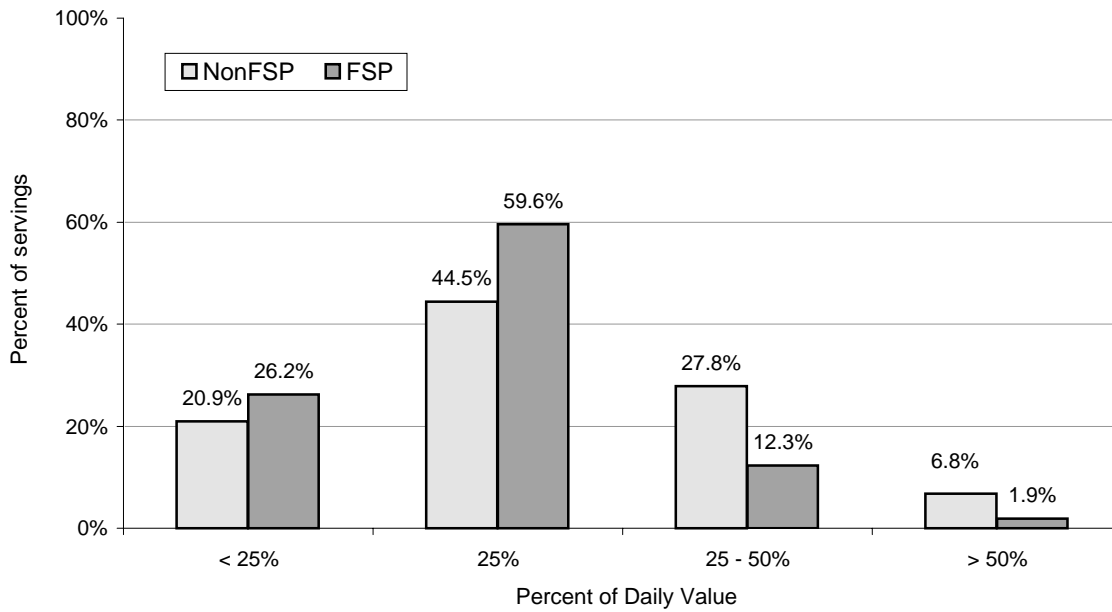


Exhibit 5-24

**Distribution of Purchased Breakfast Cereal Servings By IRON Content:
Food Stamp and Non-Food Stamp Purchases**



**Distribution of Food Stamp Purchases of Breakfast Cereal Servings,
By IRON Content and Household Type**

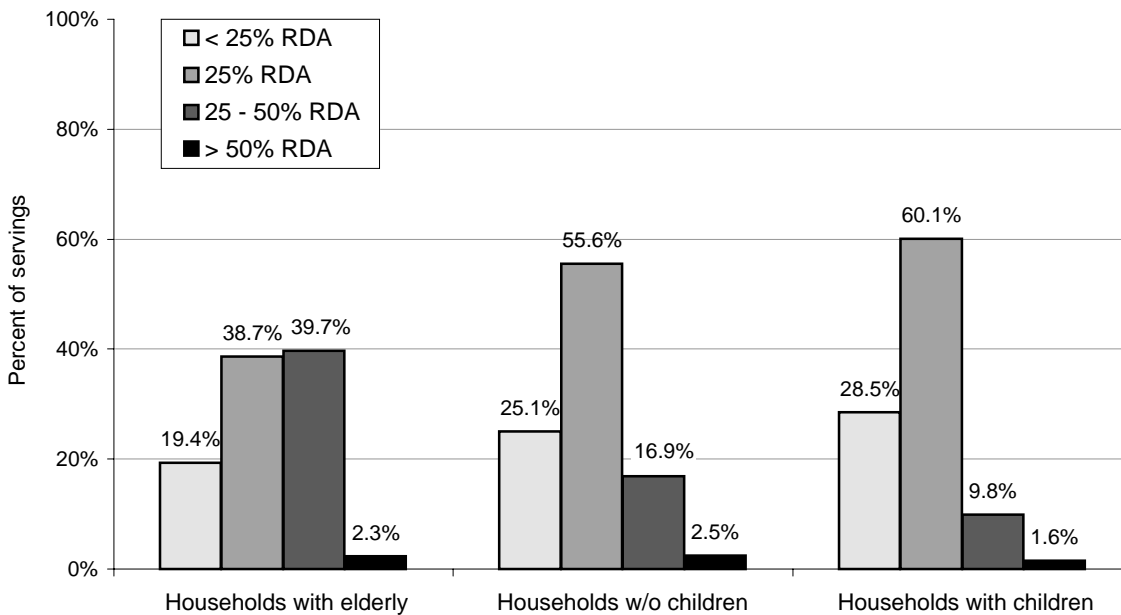
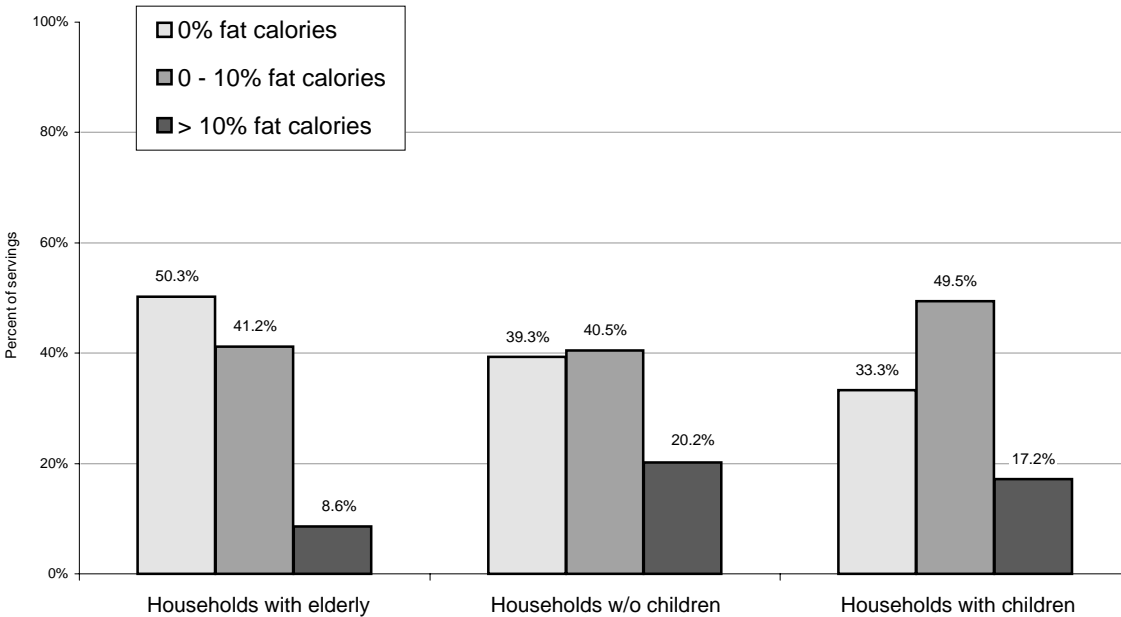


Exhibit 5-25

**Distribution of Purchased Breakfast Cereal Servings By FAT Content:
Food Stamp and Non-Food Stamp Purchases**



**Distribution of Food Stamp Purchases of Breakfast Cereal Servings,
By Percent FAT Calories and Household Type**



5.5 Characteristics of DCSU Data

The data collected by the DCSUs are so incomplete that we have not included them in the previous analyses. Store personnel in these four stores, however, did use the DCSUs occasionally. Because so little is known about what items food stamp clients buy in small, neighborhood grocery stores and convenience stores, this section presents an analysis of the limited UPC data that were collected.

During the eight-week data collection period, the DCSU file records in the four stores captured information on 3,270 transactions, of which 617 (18.9 percent) involved full or partial payment with food stamp benefits. Of these 617 food stamp purchases, only 220 (35.7 percent) included items that were scanned. For the rest of the purchases, either the scanned data were corrupted, the store clerk skipped the scanning and just entered the department code and price of the item, or none of the items purchased had UPC codes.

A total of 1,500 items were purchased in the 220 food stamp transactions with some UPC data. Nearly 59 percent of these items were scanned. Of the 882 scanned items, 298 unique items were purchased. Exhibit 5-26 presents a listing of the product categories for these items, together with the number and percentage of items purchased within each category, as well as the total dollar amount and percentage of dollars spent within each category.

Based on this incomplete sample, it appears that soft drinks and salty snacks were the most popular items purchased with food stamp benefits at these stores, representing 20.6 and 10.5 percent, respectively, of the total value of scanned items. As a group, grain products also were popular, representing 13.8 percent of total value. One must keep in mind, however, that these “findings” are suggestive at best. First, the data are too incomplete to be representative of food stamp recipient buying patterns at even these four stores. Second, some of the items (e.g., tobacco) were not necessarily purchased with food stamps, because the sample includes “combination-tender transactions,” or those in which both food stamps and another tender (e.g., cash or check) were used to pay for the entire purchase.

Exhibit 5-26
Items Purchased in FSP Transactions at DCU Stores

Product Category	Number of Items	Percent of Items	Dollar Value	Percent of Value
MILK, CHEESE, CREAM	14	1.6%	\$34.62	3.6%
Milk	9	1.0%	\$16.87	1.7%
Cheese	5	0.6%	\$17.75	1.8%
Cream	-	-	-	-
Yogurt	-	-	-	-
FRUIT AND VEGETABLES	13	1.4%	\$20.57	2.1%
Fruit	1	0.1%	\$1.39	0.1%
Juice	9	1.0%	\$12.51	1.3%
Vegetables	3	0.3%	\$6.67	0.7%
PREPARED FOODS	26	2.9%	\$49.98	5.1%
Frozen/Refrigerated	1	0.1%	\$2.29	0.2%
Deli/Salad bar	-	-	-	-
Grain Based	13	1.5%	\$20.31	2.1%
Meat Based	-	-	-	-
Soup	-	-	-	-
Diet	-	-	-	-
Baby Food	12	1.3%	\$27.38	2.8%
MEAT AND MEAT ALTERNATIVES	32	3.6%	\$55.14	5.7%
Red meat	-	-	-	-
Poultry	-	-	-	-
Fish	3	0.3%	\$6.57	0.7%
Lunch meats	10	1.1%	\$20.20	2.1%
Eggs	-	-	-	-
Legumes	4	0.5%	\$9.16	0.9%
Nuts	3	0.3%	\$3.19	0.3%
Meat, NFS*	12	1.3%	\$16.02	1.6%
GRAIN PRODUCTS	73	8.1%	\$134.59	13.8%
Breakfast cereals	13	1.5%	\$43.17	4.4%
Flour,rice,pasta	19	2.1%	\$46.39	4.8%
Bread	20	2.2%	\$33.30	3.4%
Bakery products	21	2.3%	\$11.73	1.2%
OTHER FOODS	738	82.4%	\$682.64	69.9%
Fats, oils	11	1.2%	\$22.23	2.3%
Sugar, candy	49	5.5%	\$41.03	4.2%
Seasonings	11	1.2%	\$19.70	2.0%
Soft drinks,ades	268	29.9%	\$204.12	20.9%
Coffe, tea	3	0.3%	\$5.07	0.5%
Salty snacks	160	17.9%	\$106.56	10.9%
Cookies, cakes, pies	58	6.5%	\$40.78	4.2%
Ice cream,jello,pudding	12	1.3%	\$20.78	2.1%
Condiments	15	1.7%	\$11.65	1.2%
NFS	151	16.9%	\$210.72	21.6%
Total purchases/redemptions	896	100%	\$977.54	100%

* NFS means not further specified.

Chapter 6

Lessons Learned

The project described in this report is a feasibility study, designed to address the following questions:

1. Can retailers with scanning systems be persuaded to provide their scanner transaction logs to FNS for the purposes of research into the use of food stamp benefits? At what cost? With what limitations?
2. Is it feasible to collect similar scanner transaction data in stores which do not use scanning systems? At what cost? With what limitations?
3. Is it feasible to match scanner information on food stamp purchases to characteristics of the FSP recipients making the purchases? With what limitations?
4. What analyses of food stamp recipients' purchasing behavior can be supported with the scanned food purchase data that are collected?
5. What is the feasibility and expense of expanding data collection efforts to obtain more representative data?
6. How might scanner data be used in food consumption research?

The following sections address each question in turn, focusing on what has been learned to date from this project.

6.1 Collecting Data from Retailers That Have Scanning Systems

This project has demonstrated that it is feasible to persuade some retailers with scanning systems to provide their scanner transaction logs to FNS for research purposes. Furthermore, the retailers provided the data at what seems a reasonable price of \$1,000 per chain. Although attempts were made to determine the actual costs retailers would incur to provide food purchase data, the retailers who agreed to participate in the study either could not or would not provide a cost estimate. They readily agreed, however, to the \$1,000 reimbursement proposed by Abt Associates and CCMI.

The project revealed a few technical limitations to collecting scanner data on purchased food items. Assorted problems with telecommunications systems and human error resulted in the loss of some data, and one should probably expect that similar problems would occur in any future efforts to collect scanner data. In addition, there were technical obstacles in terms of reading files prepared in different formats from the various store systems encountered. These obstacles, however, were overcome and should not pose major difficulties in any future efforts to collect scanner data from retailers.

In thinking about the feasibility of collecting data from scanning stores, one needs to remember that the two supermarket chains that participated in the study were collecting and **storing** item-level data on a transaction basis. This is important. Retailers with older POS systems may not retain data at the transaction level. Collecting transaction-level data from such retailers would be much more difficult technically, involving upgrades to the retailer's POS software and, perhaps, hardware. The feasibility of collecting data in this POS environment could not be tested during this project, although we had hoped to do so. In the end, however, no retailers needing upgrades were among the group of participating retailers. We do note, however, that the technical problems of collecting transaction-based data from retailers with older POS systems will diminish over time as more of these retailers upgrade to newer systems on their own.¹

The greatest problem in collecting data from stores is not technical; instead, it is getting the initial support of retailers. Most retailers approached during the recruiting phase of the project failed to respond to our inquiries. One had the feeling that store owners and chief executive officers simply did not want to be bothered by a government request to voluntarily participate in the proposed study. Most of the lack of support seemed to arise from indifference to the project or the competing demands for time and resources. In some instances, however, the reason was more philosophical, with retailers opposed to government use of confidential data. Indeed, the refusal of Chain B to provide non-food stamp data is a case in point. Unless a government regulation **requires** the provision of food stamp scanner data, many stores are likely to refuse to provide scanned food purchase data for research purposes.

6.2 Collecting Data from Retailers without Scanning Systems

The outlook is not very promising for collecting data on purchased items from non-scanning stores. The project did show that it is feasible to persuade some retailers to install a stand-beside data collection scanning unit, or DCSU. (The reimbursement offer of \$2,500 per store for a six-week data collection effort aided the recruiting effort.) It was difficult, however, to get the four stores with DCSUs to use them properly or consistently. After the data collection period was over, the four store owners said the biggest problem was having to double-enter information about each purchased item, first on their own cash register and then on the DCSU. This suggests that an easier-to-use DCSU, as described in the next paragraph, might be used more often and might provide usable data. Until such stores adopt their own scanning systems, however, it will be a problem to collect item-level information on what food stamp recipients are buying in these stores with their food stamp benefits. Such efforts will always be time-consuming and expensive, if only because a price file of the store's inventoried items will need to be created and updated.

Despite the above concern, it may be worth trying to collect data in non-scanning stores again, but this time with a redesigned DCSU. To eliminate the double-entry process, we would suggest combining the DCSU with cash registers already in use, if possible. Barring that, we would suggest replacing a store's current register with a full turn-key DCSU system for the duration of the data collection period. The replacement system would be a fully-integrated POS system with scanner, cash drawer, receipt printer,

1 Upgrades to newer POS systems may occur sooner than would normally be expected as retailers ensure that their POS systems are Y2K compliant (i.e., that the systems are able to read and process dates correctly beyond January 1, 2000).

and pole display.² The challenging task would be to make it easy for store employees to update and maintain the price file. In addition, because the DCSU would now become the store's primary cash register, one would need to provide very rapid service times in the event of equipment problems.

As shown in Exhibit 6-1, however, nearly 80 percent of FSP benefits are spent in supermarkets, and an estimated 88 percent of supermarkets have scanners. This means that, as shown in the last column of the exhibit, nearly 70 percent of all food stamp benefits are redeemed in supermarkets with scanning systems. Across all store types, an estimated 74 percent of all food stamp benefits are redeemed in scanning stores. So the question here is, how much time and expense should be spent trying to collect data from stores that redeem only about 26 percent of all FSP benefits? As discussed later in the chapter, the 26 percent figure is large enough to worry about the representativeness of data collected only at stores with scanning systems. There is also a policy interest in learning how these stores serve the buying needs of food stamp recipients. Therefore, even though it will clearly be difficult to collect data from non-scanning stores, it may be worth applying the lessons learned from this first test of the DCSU concept to a second effort.

Exhibit 6-1

Estimated Scanning Capabilities of FSP-Authorized Stores, by Store Type

Store Type	Number of Authorized Stores ^a	Percentage of Total Stores	Percentage of Total Redemptions	Estimated Percent with Scanners ^b	Estimated Percentage of Scanned Redemptions ^c
Supermarket	31,295	18.0	79.2	88	69.7
Military commissary	244	0.1	0.1	88	0.1
Medium to small grocery	44,691	25.7	10.9	20	2.2
Convenience store	47,665	27.4	3.0	25	0.8
Grocery/Gas station	21,792	12.5	1.1	10	0.1
Other grocery combination	4,844	2.8	1.1	60	0.7
Specialty food	15,619	9.0	3.6	10	0.4
General store	2,488	1.4	0.3	10	0.0
Grocery/Restaurant	861	0.5	0.1	5	0.0
Health/Nutrition food	1,580	0.9	0.1	10	0.0
Other stores	2,621	1.5	0.4	10	0.0
ALL STORES	173,700	100.0	100.0	32	74.0

a STARS data from January 1998

b Based on recent trade publications and judgement of independent contractor.

c The estimates assume that, within a given store type, the average FSP redemption level of scanning stores is equal to that for non-scanning stores.

Percentages may not sum to 100.0 due to rounding.

2 The existing DCSUs could be upgraded to a complete POS system for about \$912 per unit for a cash drawer, printer, and pole display. For additional units, an Omron RS6500-based system could be purchased for about \$5,700 per unit, assuming VAR (value added reseller) pricing. Retail pricing would be about 20 to 45 percent greater.

6.3 Matching Scanner Data to Recipient Characteristics

The question of whether scanner data on food stamp purchases can be matched to recipient characteristics gets the clearest affirmative answer of this feasibility study. For food stamp purchases made by Georgetown County recipients at the seven participating supermarkets, we first matched scanner data to records in the FNS ALERT file. The match rate was 96.8 percent. If one excludes the supermarket scanner transactions that were lost due to telecommunications problems, the match rate increases to over 98 percent. Our strategy of matching file records based on EBT purchase amount and the place and time of the transaction proved successful in most cases. Some problems were encountered due to differences in clock time between systems, and other problems occurred when retailer purchase amounts did not match EBT transaction log amounts.³ These problems were overcome, however, and a very high match rate was achieved.

Once we matched ALERT file records to the scanner data, we used information on EBT card numbers from the ALERT data to merge in recipient data from South Carolina's program eligibility file. No match could be found for about 12 percent of the FSP households shopping at the eleven stores. We believe, however, that a nearly 100-percent match rate would have been obtained if we had requested the state file near the **end** of the data collection period (to capture information on new entrants) and had requested that all records be included in the extract, not just records of **active** recipients. In the end, data on recipient characteristics were matched to 91 and 95 percent, respectively, of the supermarket and DCSU EBT transactions for which we had scanner data.

6.4 Analyses Supported by the Data

Because we were able to collect a large amount of scanner data on FSP and non-FSP purchases in supermarkets, the project has demonstrated the feasibility of answering, at least in part, two key policy questions: (1) what are food stamp households buying with their program benefits? and (2) how do food stamp purchases compare to non-food stamp purchases? Within these two areas of policy interest, the project has demonstrated the feasibility of looking at such detailed information as:

- How are benefits allocated across major product categories, and how does this compare to purchases by other customers?
- To what extent do food stamp recipients purchase store-brand items instead of costlier major-brand items, and how does this compare to purchases by other customers?
- To what extent do food stamp recipients purchase items in larger sizes instead of costlier (on a per-unit basis) small-size items, and how does this compare to purchases by other customers?

³ As noted in Section 4.4, retailer purchase amounts could exceed EBT transaction log amounts if a combination of FSP and other tender was used for the purchase. Also, the total dollar amount of two consecutive retailer records sometimes equalled the dollar amount of a single EBT transaction.

- What is the nutrient level of items purchased with food stamp benefits (within selected product categories), and how does this compare to purchases by other customers?

Furthermore, because we were able to link the characteristics of food stamp households to items purchased with food stamp benefits, the study has demonstrated the feasibility of looking at variations in the above purchasing patterns for subgroups of the food stamp population.

All of the above analyses, however, were limited by data problems. Most obviously, the small number of stores that participated in this phase of the study means that **the analysis results are not representative of food stamp recipients' buying patterns**. Instead, the results show what types of analysis are feasible with scanner data; they also can be used to generate hypotheses about buying behavior that could be checked in a larger study with a more representative group of stores. Another data-related problem is that the comparison of FSP and non-FSP purchases was limited to just one supermarket chain because the other chain refused to provide data on non-FSP purchases. Data difficulties also prevented analyses of how purchasing patterns vary by urban/non-urban location and item prices. Finally, due to problems collecting data from non-scanning stores, the study was unable to examine how food stamp purchasing patterns vary by store type. Especially disappointing was the inability to obtain a more complete picture of what food stamp recipients buy at convenience stores and small, neighborhood grocery stores.

The limited number of stores participating in the data collection also precluded any household-level analysis of buying patterns. That is, we were able to capture information on **all** monthly food stamp purchases for only a small, non-representative sample of recipients. Part of the problem is that, even in a county with a relatively limited number of program-authorized stores, food stamp recipients shop at multiple stores during the month. This suggests that it will often be quite difficult to capture complete household-level data unless participation rates by retailers are quite high within a given area.

Nevertheless, even with these data-related problems, the project has shown that it is possible to advance considerably our understanding of food stamp purchasing patterns. Prior to this study, the only available information on what food stamp households purchase with their benefits was survey-based. This study has shown the potential for collecting more accurate and detailed information.

6.5 Feasibility of Expanding Data Collection Efforts

This project was originally envisioned as having two phases. During the first phase, scanned food purchase data were to be collected from a small number of retailers in a single market area. Then, if the first phase demonstrated the feasibility of collecting and analyzing scanner data (and if additional research funds were available), the second phase would expand data collection in the same market area to include more retailers. This second phase of the project would have several goals. First, by seeking to include more retailers, it would further test the feasibility of recruiting retailers to participate in a study of this kind. Second, by including more retailers, it would test the feasibility of collecting, processing, and analyzing scanner data from a wider variety of retail environments. This would be especially useful in terms of gaining experience in dealing with different POS systems and scanner file formats. It might also provide the opportunity to include one or more retailers in which hardware or software upgrades would be needed before scanner data at the transaction level could be captured. Finally, by capturing a larger percentage of EBT transactions in the market area, any resulting data analyses would be more

representative of the purchasing behavior of FSP recipients within the study area. After concluding the two phases, a final objective of the study was to assess the feasibility and expense of expanding data collection efforts to obtain more nationally representative data.

The first phase of the project is now complete. Although currently available information is not sufficient to assess the feasibility and expense of obtaining nationally representative data, we can use what has been learned to date to discuss the feasibility of future data collection efforts. We can also identify the major obstacles which must be overcome before representative data could be collected.

The major impediments to collecting representative scanner data are two-fold. First, many retailers with scanning systems will not be supportive of USDA attempts to collect and use these data to study the purchasing behavior of food stamp recipients. Second, a large number of FSP-authorized retailers do not use scanning systems. We address each problem below.

Retailer Recruitment Among Scanning Stores

One cannot obtain a representative sample of scanner data, from any location, unless a reasonable number of stores with scanning systems agree to provide the needed data. The project has shown how difficult this can be. All may not be hopeless, however, because the number of retailers who refused to participate outright on ideological grounds was limited. Most retailers “refused” in a more indirect way, simply ignoring repeated requests for support for the project. It is possible that many of these retailers could be persuaded to participate, given sufficient time to build a relationship with the store’s senior staff. In addition, the fact that the first phase of this project has been completed might be viewed as setting a precedent by some retailers, making it less likely that they would refuse a future request.

Furthermore, a number of retailers we contacted during the first phase said they could not participate at that time due to a variety of reasons, including turnover in key staff and being in the midst of a POS upgrade effort. Although one cannot discount the possibility that some of these retailers were simply looking for an excuse not to participate, it is entirely possible that many of them would participate if approached at a more convenient time. What this does indicate, however, is that even among the group of retailers who would support USDA collection of scanner data in the future, some number of them at any given time will be unable to participate for various reasons.

The remaining question, of course, is what percentage of scanning stores would need to participate to provide representative scanner data for analysis? That is, what is the “reasonable” number of stores alluded to at the beginning of this section? Unfortunately, there is no straightforward answer. In thinking about the question, however, there are several factors to consider.

First, if the goal is to capture nearly complete food stamp purchasing information on a large and representative sample of FSP households, then the needed percentage of participating stores is probably high. Because most FSP recipients do not spend all of their benefits in one store, one can obtain a complete picture of buying behavior **at the household level** only if a large percentage of stores provide scanner data. If the goal instead is to learn what FSP households in general are buying with their food stamp benefits, without regard to linkage across multiple purchases, then a smaller percentage of participating stores would be acceptable.

Second, in thinking about the representativeness of data from a sample of stores, one is really more interested in the percentage of transactions and redemption dollars being captured in a data collection effort, not in the percentage of stores participating. All else equal, one is better off if the participating stores are high redeemers rather than low redeemers of food stamp benefits. Therefore, if one or two stores (or chains) that dominate a given market area agree to provide data, the data may be fairly representative even if the percentage of participating stores is low.

Third, this project has not been able to assess the extent to which buying behavior varies according to store type; both of the chains that participated were large supermarket chains. Therefore, to help ensure the collection of a representative set of scanner data, one should try to obtain participation across all types of stores (as identified by FNS store type categories) in a market area with scanning systems. This obviously would include grocery stores. In some market areas, it would also include specialty stores and convenience stores with scanning systems.

Data Collection at Non-Scanning Stores

From Exhibit 6-1, we estimate that approximately 26 percent of food stamp redemptions are spent in non-scanning stores. Thus, nothing can be learned about one-quarter of all food stamp redemptions unless a method is developed for collecting item-level bar-code data from these stores. The DCSUs tested during the project were not successful for two reasons. First, store employees were unwilling or unable to use them on a consistent basis. Clearly, processing all sales twice—once on the DCSU and once on their own cash register—was difficult for them. Second, DCSU software problems made the units unnecessarily difficult to use, and corrupted some of the data that were collected.

Keeping these problems in mind, there are four alternative approaches one could take with respect to food stamp purchases in non-scanning stores. First, one could simply ignore this component of the food stamp market. At an estimated 26 percent of all benefits redeemed, however, it is a large component to ignore. Furthermore, there is at least an *a priori* belief that items purchased in these stores are quite different from items purchased in stores with scanners, which tend to be supermarkets and large grocery stores. Thus, one risks losing information about a potentially very interesting component of food stamp purchasing behavior, and introducing bias into depictions of overall purchasing patterns.

Second, one can simply wait until more stores install scanning systems, and then try to collect scanner data from these stores. This is not much different from the first alternative, but it does point out an expectation that more and more stores will use scanners in the future. It also suggests a possible future research question for data collected from scanning stores—do items purchased with food stamp benefits in stores with recently-installed scanners differ from items purchased in stores with older scanning systems? Such an analysis might indicate whether and how recent converts to scanning differ from other scanning stores, and hence indicate the type of error that results when information from non-scanning stores cannot be obtained.

Third, one can try to encourage non-scanning stores to install and use scanners. As part of this project, CCMI was prepared to offer technical assistance to any non-scanning store that was willing to install a new POS system and participate in the project. As it turned out, no stores in Georgetown County were identified as being at this stage.

Finally, further attempts could be made to collect scanner data with a DCSU. As noted earlier in this chapter, the DCSU approach would have to be modified to improve the chance of successful data collection. At a minimum, there would need to be more extensive field testing of the units before data collection began, and more rapid inspection of collected data to verify that the units were being used and data collected properly. Finally, there should probably be greater on-site support for the retailers, particularly at the start of the data collection period. By staying on-site, project staff could help ensure that the systems were being used and that initial training had been adequate.

Cost of Collecting Scanner Data

Little prior evidence is available for estimating the cost of collecting scanner data. Although several companies collect scanner data for market research purposes, no information is available on their cost structures. From the experience to date in Georgetown County, we have the following cost information:

- The two supermarket chains that participated in the study received reimbursement of \$1,000 each for agreeing to provide six weeks of scanner data files to project staff. Ultimately, eight weeks of data were collected, but the stores did not request further compensation.
- The four non-scanning stores that participated in the study using DCSUs received \$2,500 each for agreeing to collect UPC data with the DCSUs for six weeks. As with the supermarkets, these stores did not request further compensation when the data collection period was extended to eight weeks, perhaps because the extension was needed due to inconsistencies in their use of the DCSUs.
- Seven DCSUs were used. The average cost of each DCSU was about \$4,000, although this represented a value-added reseller (VAR) pricing level. Retail prices for the same DCSUs would range between \$4,800 to \$5,800 per unit.
- The cost to contact retailers in the South Carolina study site, elicit their cooperation for the study, and negotiate agreements is estimated at roughly \$57,000.⁴
- Estimated data collection costs, excluding the above-mentioned reimbursements to stores, were approximately \$106,000. Nearly all of this cost is related to data collection at the four DCSU stores that participated in the study; it includes preparation of the DCSU software, creation of item price files for the four stores (which involved manual scanning and key entry of information about all unique items in the store's inventory), training of store personnel, and weekly trips to South Carolina to download data from the DCSUs and assist store personnel.

4 This cost estimate was derived by looking at the monthly expenditures by CCMI during the period of retailer contacts and recruitment. Prior to June 1997, CCMI was contacting retailers in 17 potential study sites. Total costs of \$55,700 during this period have been divided by 17 to estimate costs related to Georgetown County. In June, CCMI split its time contacting retailers in South Carolina and Kansas, so 50 percent of the total June cost of \$15,400 is assigned to Georgetown County, as are all costs (\$46,000) during July 1997 and thereafter. The cost estimate of \$57,000 includes travel expenses associated with a trip to Georgetown County to meet with the owners of several small, non-scanning stores to elicit their participation in the study.

- The estimated cost of processing the collected scanner files in order to build a database containing records from multiple stores was \$63,000, of which an estimated 50 percent was related to the DCSU stores. This cost does not include the expense of obtaining an extract of South Carolina’s program eligibility file, EBT transaction log records, or merging these databases with the scanner data.

Taken together, the above costs total about \$267,000, of which approximately \$188,000 is attributable to data collection at DCSU stores. The remaining \$79,000 covered retail contacts, contract negotiations, data collection, and data processing at the two supermarket chains that had POS systems in place with optical scanners. On a per-chain basis, the average cost was \$39,500. To extrapolate these costs to possible future data collection efforts, one should use a somewhat lower figure—about \$30,000 per chain. Some of the costs incurred in the current study were for initial start-up tasks. These tasks included drafting a standard data collection agreement for use with retailers and writing programs for data processing. Such tasks would be much less expensive in subsequent data collection efforts.

As noted, most of the estimated cost of \$30,000 per chain is for contacting retailers, eliciting their cooperation, assessing the technical aspects of their POS systems, entering into an agreement with them, and collecting the data through weekly file transfers. This total cost per chain generally is not subject to economies of scale, so the \$30,000 estimate is appropriate for use in projecting costs regardless of the number of chains a future data collection effort might encompass.⁵ If future data collection efforts were to cover a data collection period longer than a month or so, or involve multiple data collection periods (as in a “pre-post” research design with data being collected before and after some intervention), then estimated costs would increase somewhat to cover retailers’ ongoing costs of file transfers. As long as retailers were willing to provide data to USDA “at cost,” however, the average total cost per chain should not exceed \$35,000 or so.

To place this cost estimate in perspective, it is useful to compare it to the cost of a major recipient survey designed to collect nationally representative data on food shopping patterns, food expenditures, and household food use. The National Food Stamp Program Survey (NFSPS) was conducted in 1996-97. The total cost of the survey was about \$2.5 million, of which an estimated \$1.7 million was spent on instrument development, OMB clearance, sample preparation, interviewer training, and data collection.⁶ Thus, at \$35,000 per chain, one could collect scanner data from approximately 50 store chains for about the same price as the NFSPS. Each chain, of course, would represent many individual stores; we estimate an average of about 43 stores per supermarket chain.⁷ With this estimate, one could collect scanner data at about 2,150 supermarkets for the same price as the NFSPS. This sample would represent nearly 7

5 Economies of scale are present, however, at the store or transaction level because most costs remain the same regardless of how many stores are in the chain or how many transactions are captured.

6 Cost estimate based on January 8, 1999 telephone conversation with FNS project officer for the NFSPS. The NFSPS included a sample of eligible non-participants, which is an expensive group of households to identify. It is not known how much less expensive a survey of just FSP participants would be.

7 The figure of 43 stores per supermarket chain is based on a six-state extract from the STARS database. The states are Colorado, Kansas, Maryland, New Mexico, South Carolina, and Wyoming. Chains having fewer than ten stores have been excluded in computing the average, which skews the average upward. Because the six states are not contiguous or highly urbanized, however, the average also may be biased downward. Within the six states, supermarkets in chains with ten or more stores represent 66 percent of all supermarkets and 72 percent of all supermarket redemptions.

percent of all FSP-authorized supermarkets (or 1.2 percent of **all** FSP-authorized stores) and almost 6 percent of all FSP redemptions.⁸

6.6 Potential Research Using Scanner Data

This section presents several ideas for possible next steps in the use of scanner data for research purposes. The examples include evaluation of USDA initiatives to change shopping behavior, gaining more information on FSP expenditure patterns, and “methodological” studies designed to better assess the research value of scanner data or improve the usability of either scanner data or survey data. In evaluating these ideas, one needs to keep in mind that obtaining fully-representative scanner data on FSP expenditures is not possible at this time. Many stores do not use scanners, and stores with scanners may or may not cooperate with USDA efforts to collect scanner data for research purposes.

These coverage issues point out another difference between survey-based expenditure data and scanner data. In surveys the unit of observation is the surveyed household or individual. For scanner data the unit of observation is the food transaction. To capture all food stamp purchases for a sample of households would require nearly complete participation by scanning stores in a market area. Lacking such complete coverage and participation, it is better to view scanner data as representing a cross-section of FSP purchases for research purposes rather than an assemblage of transactions for given households.

Effects of USDA Initiatives on Food Expenditures

Large-scale national surveys of food stamp recipients have the advantage of providing nationally representative data. Their sample designs, however, do not support the collection and analysis of data within specific states or market areas. Thus, special-purpose surveys would need to be developed and fielded in order to assess the impacts of any targeted efforts by USDA to change food shopping patterns. Examples of such initiatives might include a nutrition education campaign or an effort to help food stamp recipients become more price conscious when food shopping.

Such initiatives might be better assessed through analysis of scanner data. One could use scanner data collected from area stores both before and after the initiative to measure changes in purchasing behavior. Certainly, scanner data offer a more accurate and detailed record of what recipients are buying with their program benefits than survey data.⁹ For this reason, scanner data would be more likely than survey data to detect a small treatment effect.

Although the lack of data capture in non-scanner stores would be a problem in any effort to evaluate the impacts of a USDA initiative, the magnitude of any bias arising from incomplete coverage would likely be small (assuming all scanning stores agreed to participate). Scanning stores redeem an estimated 74

8 The estimated number of program-authorized supermarkets is 31,295. If supermarket chains represent 66 percent of all supermarket stores, then an estimated 20,655 supermarkets are part of a chain. A sample of 2,150 supermarket chain stores is 10.4 percent of this universe. Supermarkets redeem 79.2 percent of FSP benefits. If supermarket chain stores represent 72 percent of all FSP redemptions, then supermarket chain stores redeem 57 percent of all FSP benefits (i.e., 0.72×79.2 percent). A 10.4 percent sample of supermarket chain stores would then redeem an estimated 5.9 percent of all FSP benefits.

9 Unlike a survey, of course, scanner data could not provide any information about whether the recipient had heard about the initiative being evaluated.

percent of FSP benefits, so any impacts of an initiative should show up in these stores. In addition, because the main cost of collecting scanner data is the up-front negotiations with stores and working out the technical details of data transfer, the marginal cost of extended data collection should be low. This means that any “decay” of impacts could be measured fairly easily by collecting scanner data some time after the initiative had ended. In comparison, three waves of a survey would be needed to capture baseline data, post-intervention data, and data on potential decay effects, with little cost savings available during multiple survey waves.

Monitoring of Long-Term Trends in Food Purchasing Behavior

In the section above we noted that the marginal cost of extended data collection should be low in stores with scanning systems. One can envision negotiating with a sample of scanning stores to collect scanner data at periodic intervals over an extended period. For instance, one might collect one month’s worth of data once or twice a year for 10-20 years. During each cycle of data collection, the scanner data would be merged with EBT transaction data and state eligibility files to match household demographic information to purchase behavior. Such a project could monitor long-term changes in the food purchase behavior of food stamp recipients. If the stores (or a subsample of stores) agreed to provide data on non-food stamp transactions as well, one could examine whether differences in purchasing patterns between food stamp recipients and the rest of the population were changing over time.

Issues of sample representativeness do not go away in such a research project, but they become somewhat less of a concern when one focuses on **changes** over time within a given group of stores. Longitudinal data from even a non-representative sample of stores should identify long-term changes in buying habits. The greater concern is how changes in the population of interest might confound the analysis, although household-level data on recipient demographics should allow one to control for the effects of such changes.

Effects of Workforce Participation on Food Purchasing Behavior

With current efforts to increase the workforce participation of food stamp recipients, one area of interest to USDA is how participation in the workforce affects diet and nutrition. For instance, if program recipients have less time to plan for and prepare meals, is there a shift toward consumption of more quick-to-prepare items like frozen dinners or hot dogs? Or, with less at-home supervision of children, is there an increase in consumption of snacks and soft drinks? Although scanner data do not directly measure what is being consumed, they could be analyzed to see whether the above hypotheses are supported by what is being purchased with FSP benefits. This could most easily be examined by comparing food expenditure patterns between FSP households with different levels of workforce participation, especially for households with and without children.

Food Stamp Expenditures at Different Store Types

One question of interest to USDA is how different types of program-authorized stores meet the needs of FSP recipients. For instance, when shopping at convenience stores with their food stamp benefits, do recipients purchase basic commodities like eggs, bread and milk, or do they buy snacks and chips? Additionally, to what extent do recipients pay a price premium for the convenience offered by small, neighborhood-based food stores? Some information on store choice and prices was collected by the NFSPS, but scanner data offer the possibility of more accurate data.

There is a problem, however, with store coverage when considering using scanner data to examine food stamp expenditures at different store types. As was shown in Exhibit 6-1, small, neighborhood stores like convenience stores appear to be among the least likely stores to use POS systems with optical scanners. Thus, at this point in time, one could not readily obtain scanner data to conduct a representative study of how food stamp expenditures vary by store type.

Although the coverage problem could be addressed with use of DCSUs in non-scanning stores, the feasibility study suggests that an easier-to-use DCSU would need to be developed before such an approach could be successful. Such a development and testing effort, including use of possible new technologies, might be an area for future research. Even if successful, however, one should expect data collection at non-scanning stores to be expensive. In the feasibility study, the total cost to collect data at four non-scanning stores was \$188,000, or \$47,000 per store. The expense of weekly trips to collect data could be avoided by electronic file transfers, but two very expensive cost components would remain in any application. First, a host price file would need to be built at the start of data collection and maintained thereafter. This would require scanning the store's inventory of products and key entering product and price information. Second, arrangements for on-site servicing of the DCSU would need to be made so retailers could be assured that any problems with the equipment would be fixed quickly.¹⁰ Although some economies of scale could be expected as more stores within a service market were added, the savings over the \$47,000 per store estimate likely would be modest.

Even without coverage in non-scanning stores, preliminary analyses of food stamp expenditures at different store types could be conducted using scanner data from the limited number of large convenience store chains that have installed scanner systems. If coupled with scanner data from one or more supermarket and grocery store chains in the same market areas, one could compare both what was purchased with FSP benefits at different store types and the prices of similar items at different stores. When conducting such a study, however, one would need to remember that FSP purchases at large, convenience store chains with scanners probably do not represent FSP purchases at convenience stores in general, even those within the same market area. This, coupled with lack of nationally representative coverage, means that the results of such a study would be suggestive at best.

The lack of national coverage would be less of a problem if USDA wanted to track **changes** over time in how FSP recipients spend their benefits in different store types. For instance, if FNS implemented an education campaign to help recipients become more aware of pricing differences among different store types, scanner data from the market area(s) involved in the campaign could be collected and analyzed to determine the effectiveness of the campaign.

More Complete Expenditure Studies

One problem with the scanner data collected during the South Carolina feasibility study is that they did not capture non-FSP expenditures at food stores by FSP households (except for food purchased with cash during mixed-tender purchases). With use of EBT cards, however, it might be possible to mimic a "loyalty card" program in selected markets. Loyalty card programs are store-based marketing programs

¹⁰ This assumes that the newly developed DCSU would serve as a **substitute** for the retailer's current cash register, not as an additional POS device. The feasibility study suggests that store employees will not use a DCSU consistently if it is a stand-beside unit, especially during times when the store is busy.

that issue special cards to voluntary participants. The cards need to have some identification (e.g., a bar code) attached so the participant can later be matched with the purchase. Usually, price discounts are offered to promote the program and continued use of the card.

The idea would be to use the EBT card as a multi-store loyalty card. Bar-coded stickers could be affixed to the card and scanned whenever a purchase was made, regardless of whether the EBT card was being used to pay for that purchase. Alternatively, the cashier could swipe the EBT card through the POS card reader so that the EBT card number would be stored on the POS transaction log.

Such a project would expand the captured scanner data to include cash purchases made by FSP recipients as well as their food stamp purchases. Implementing such a project, however, would face difficulties. It would require cooperation among state FSP officials, the EBT vendor, participating retailers, and FSP recipients. Furthermore, USDA would probably need to fund some price discounts to induce FSP recipients to offer their EBT card even when not using their program benefits. If implemented, however, a number of interesting analyses could be performed. First, a more complete record of food expenditures by FSP households would be available, so more could be learned about purchasing behavior or response to nutrition education campaigns. Second, one could better estimate aggregate food expenditures by FSP households and the extent to which FSP benefits meet food purchasing demands (including when during the benefit month FSP benefits are exhausted and recipients begin using other payment tender). Third, a direct comparison could be made between the types of food purchased with and without FSP benefits; this would be useful in determining the extent to which scanner data on FSP benefit purchases were representative of food purchased with other resources.

Methodological Studies

The South Carolina feasibility study developed a small database containing nutrient data for selected products, based on their UPC codes. If USDA decided to expand its research into the use of scanner data, it might want to sponsor the creation of a more complete database. There would be many challenges in such a project. The first would be searching for the requisite data in electronic form. If manufacturers could not supply nutrient data based on UPC codes in electronic form, then the project would incur the expense of locating and manually key entering nutrient data for tens of thousands of products. One also would have to determine whether it was feasible to assign nutrient values to variable-weight PLU items. Existing nutrient databases probably could be used to assign values for, say, a pound of apples, but some PLU items might have such generic descriptors that no meaningful nutrient values could be assigned.

Finally, an additional challenge is maintaining such a nutrient database. Thousands of new food products are introduced each year, and the nutrient data for these products would need to be added to the database. Maintenance would also require some spot checking of nutrient values of existing products to see whether their nutrient content had changed over time.

Despite these difficulties, it should be easier and less expensive to build and maintain a nutrient database based on UPC and PLU codes than similar databases based on food use or intake surveys. A UPC-based database would have standard products with known (and often labeled) nutrient contents. Compare this to the problems faced by databases that need to assess the nutrient value of food as served (e.g., a portion

of meat loaf). Thus, the real question is whether it is worth building a nutrient database for **food expenditure** data, given the gap that often exists between what is purchased and what is eaten.

Another possible use of scanner data is to assess the accuracy of survey-based food expenditure data. Unless survey respondents keep receipts of their food purchases, such surveys must rely on respondents' recall ability. To assess the accuracy of survey expenditure data, one would need to negotiate agreements to collect scanner data from as many stores as possible within a market area, and then field a food expenditure survey within that same market. The survey should ask where food items were bought as well as their cost. The survey would have to limit its questions to food purchased with food stamp benefits. The survey data and scanner data could then be matched (based on EBT card or case identification number) and compared. Such a study could evaluate the extent of recall problems both for items purchased and their price.

Finally, given the concern about coverage of scanner data, another research option would be to more formally assess the potential representativeness of a national sample of scanner data. Toward this end a survey of program-authorized stores could be conducted to determine the characteristics of stores with and without scanning systems. For those stores with scanning systems, the survey could ask details about the manner in which scanner data are collected and stored (e.g., are scanner data retained at the transaction level?). For all stores, the survey could ask about future plans to install or upgrade scanning systems. Together with information on FSP redemption levels within the sampled stores and separate EBT data on the number and dollar value of FSP transactions across all store types, the survey information would support an assessment of how representative scanner data might be. The final issues to consider in such a study are whether stores with scanning systems would be willing to provide scanner data to USDA for research purposes, and at what cost. Although somewhat difficult to assess without actually trying to negotiate a data collection agreement, the survey should be able to provide a preliminary assessment of likely cooperation with efforts to collect scanner data.

Appendix A

FSP Research on Food Consumption

Research questions regarding food consumption often distinguish among food **expenditures**, food **use**, and food **intake**.¹ In surveys, food expenditures are usually measured at the household level for a specified period (e.g., one week). Data on food expenditures usually include food eaten away from home and meals ordered in, as well as food items purchased at grocery stores. The expenditure data may be collected either in aggregate or disaggregate form. In the latter approach, respondents are asked both about what they purchased and how much it cost, as opposed to a simpler question regarding total (aggregate) food expenditures over some period.

Food use surveys, in contrast, focus on the types, quantities, and prices of all foods used (rather than purchased) during a specified time period, whether eaten at home or away. Such data are usually preferred to food expenditure data because they include non-purchased food items (e.g., home-grown food and gifts) and they avoid problems of how to handle the “food storage” problem when evaluating diets. The latter problem arises because, with food expenditure data, one does not know when food purchased this week will be eaten. Similarly, part of this week’s food use may be taken from storage. Data on food use bypass the problems introduced by the temporal mismatch between food purchase and food use.

Food intake data represent the types and quantities of food actually eaten during a specified time frame. They differ from food use data in that they are usually collected at the individual rather than household level. They also exclude food used but not consumed (e.g., baking a cake for a friend, preparing a casserole for next week’s meals).

Both food use and food intake data are often converted to measures of nutrient value using databases that match nutrients to specified food products. In assessing diet or nutrition levels, food intake data are generally considered superior to food use data because they more directly measure what is consumed. Collecting food intake data, however, is typically more expensive than collecting food use data, and it imposes more burden on respondents. Similarly, when assessing diet or nutrition, food use data are generally viewed as superior to food expenditure data. Again, however, the more valuable food use data are typically more expensive and burdensome to collect than expenditure data.

The following sections provide brief summaries of some of the major surveys providing information on food consumption.

1 These distinctions are elaborated upon in Thomas Fraker and Sheena McConnell, “Recommendations on Sources of Survey Data on Food Consumption to Meet FNS’ Research Needs,” Washington, DC: Mathematica Policy Research, Inc., August 28, 1992.

Nationwide Food Consumption Survey (NFCS)

A major survey source of information on food consumption has been USDA's Nationwide Food Consumption Survey (NFCS). This survey, which has been conducted by the Agricultural Research Service (ARS)² approximately every ten years, was last conducted in 1987-88. The previous survey was conducted in 1977-78. The NFCS is designed to collect information about the types, nutrient content,³ and money value of food used in US households and about the food intake of individuals while at home and away from home. As such, it is one of the few major surveys to collect data on all three measures of food consumption.

The 1987-88 NFCS consists of two samples. The basic sample is designed to represent all households in the 48 contiguous states; the second sample represents low-income households (i.e., total income less than or equal to 130 percent of the poverty level). The planned sample sizes for the 1987-88 NFCS were 6,000 for the basic sample and about 3,500 for the low-income sample. Response rates to the survey, however, were very low.

Continuing Survey of Food Intake by Individuals (CSFII) and the Diet and Health Knowledge Survey (DHKS)

In between the years of the decennial NFCS, the USDA's Agricultural Research Service sponsors the Continuing Survey of Food Intake by Individuals (CSFII) and the Diet and Health Knowledge Survey (DHKS). These surveys were most recently conducted in 1994 through 1996.⁴ The CSFII includes a nationally representative sample of individuals who, during in-person 24-hour recall interviews, provide aggregate information on household food expenditures and detailed information on individual food intakes for two nonconsecutive days. The DHKS, which is a telephone interview administered about two weeks after the CSFII interview, collects information about knowledge and attitudes toward dietary guidance and health. Both surveys oversample the low-income population. The annual survey sizes are roughly 5,000 individuals for the CSFII and 2,000 adults for the DHKS.

Third National Health and Nutrition Examination Survey (NHANES III)

Between 1988 and 1994, the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention conducted the Third National Health and Nutrition Examination Survey (NHANES III).⁵ This survey was conducted on a nationally-representative sample of nearly 34,000 individuals age two months and older. It was designed to obtain information on the health and nutritional

2 Formerly the Human Nutrition Information Service (HNIS).

3 Data on nutrient availability and intake are derived from quantities of food used at home and food intake using the HNIS National Nutrient Data Bank.

4 The CSFII and DHKS also were conducted in 1989 through 1991. The CSFII was conducted in 1985 and 1986 as well.

5 Earlier NHANES surveys were conducted in 1971-74 and 1976-80.

status of the U.S. population through interviews and direct physical examinations. The survey collected food consumption data by asking how often in the month before the interview respondents consumed food in 60 different groups. Detailed food intake data were collected for the prior 24 hours for all respondents. Additional food intake data were collected at two later points for respondents aged 50 or older.

Integrated National Food and Nutrition Survey for Year 2000

Plans are now in place to combine the CSFII and NHANES samples and to integrate the two surveys in the future. The Integrated National Food and Nutrition Survey for Year 2000 will complete interviews with a nationally-representative sample of 11,000 persons of all ages.

Cash-Out Studies

The USDA has sponsored a number of evaluations to assess the impacts of cashing out FSP benefits. The first study examined a demonstration in which benefits for elderly recipients and recipients receiving Special Supplemental Income (SSI) were cashed out in nine sites in 1980. The second study looked at cash-out in Puerto Rico when the Commonwealth replaced the FSP with the Nutritional Assistance Program (NAP) in 1982. In the mid- to late 1980s, cash-out demonstrations took place in Washington State, San Diego, and Alabama.

Each evaluation of cash-out included a recipient survey. All six evaluations (two demonstrations in Alabama were studied) looked at the effects of cash-out on food expenditures, and all but the SSI evaluation and one of the Alabama evaluations looked at impacts on food use. The Puerto Rico study examined the effects of cash-out on dietary intake. Finally, one of the Alabama studies and the Washington State and San Diego evaluations assessed impacts on nutrient availability. Sample sizes were generally in the 1,200 to 2,400 range.

National Food Stamp Program Survey (NFSPS)

USDA recently sponsored the National Food Stamp Program Survey (NFSPS), the first nationally representative survey of FSP recipients. Conducted in 1996-97, the NFSPS includes three sample components: in-person interviews with 1,100 FSP households, telephone interviews with another 1,000 FSP households, and 1,100 telephone interviews with low-income households. The survey collected information on program access for FSP participants and other low-income households, stigma and customer service, access to stores, and food security and nutrient availability. Follow-up interviews with 1,100 FSP households collected detailed data on food shopping patterns, food expenditures, and household food use. This included a detailed enumeration of types, quantities, source and prices of foods used in the previous seven days.

Appendix B

Sample Recruitment Letter to Retailers

Dear (store owner):

The United States Department of Agriculture (USDA) is sponsoring a research study of the purchasing habits of food stamp recipients. The study will be conducted within a single, medium sized market area with a variety of different retailer types, from supermarkets to convenience stores. Your company has been identified as a retail food chain operating in one or more of the potential market areas under consideration for the study. We would therefore like to talk to you about the study and to gain your support if a site where your stores are located is selected.

The study will explore the feasibility of using scanner technology to collect data on food items purchased with food stamp benefits. Consumer Card Marketing, Incorporated (CCMI), of Braintree, Massachusetts, in conjunction with Abt Associates Inc., the prime contractor for this study, will conduct the field research and assist in preparing the analysis files.

There are two primary reasons why the Food and Consumer Service (FCS) of USDA is undertaking this effort. First, improving the nutritional status of low-income households is a mission of the Food Stamp Program. Second, information on what food stamp households buy is very useful in measuring the program's effectiveness in meeting its mission and in designing more effective nutrition assistance programs. In the past special household surveys were used to collect information. Today, however, scan data are deemed to be more reliable and timely.

This study will be conducted over a six-week period this summer. During the study period, we will seek to collect transaction-based scan data for all food stamp purchases and a sample of other purchases. In stores with scanner systems, we will seek a copy of the store's transaction log for analysis. In other stores, we wish to explore the possibility of installing a stand-beside data collection scanning unit (DCSU) for the six-week study.

All information will be kept in strict confidence. Once the transaction data are combined with the food stamp households' demographic information (using information from the EBT system and the state's program files), all store and household identifiers will be stripped from the database. No store or household-specific results or data will be reported or provided to FCS.

We ask for your support and participation in this important pilot study. If a site where you operate is selected and you agree to participate in the study, Abt Associates Inc. will reimburse you for the direct costs involved in preparing a copy of your transaction log. In addition, we at CCMI (Consumer Card Marketing, Inc.) will provide any technical assistance required in generating a log from your current POS (Point of Sale) system.

I will call you next week to schedule a convenient time to speak with you and/or the appropriate person on your management staff, and explain the study in further detail. I will also be able to address any questions you may have.

Thank you for your support. We look forward to working with you.

Sincerely,

(Signed)
Senior Vice President

Appendix C

Sample Data Collection Agreements

The first agreement that follows was used for the two supermarket chains that participated in the study. The second agreement was used for the four stores that used data collection scanning units (DCSUs).

**Data Collection Agreement Between
ABT ASSOCIATES INC.
and
[SUPERMARKET CHAIN]**

Abt Associates Inc. and [chain] agree as follows:

1. Abt Associates Inc., a Massachusetts corporation headquartered in Cambridge, Massachusetts, has entered into a contract with the United States Department of Agriculture, Food and Consumer Service (FCS) to study the feasibility of using scanner technology to collect data on food items purchased with and without food stamp benefits; and
2. Abt Associates Inc. has obtained the services of Consumer Card Marketing, Inc. (CCMI) of Braintree, Massachusetts, as a subcontractor to conduct the field research and assist in collecting the scanner data and preparing the analysis files; and
3. Georgetown County, South Carolina, has been selected as the project's study site; and
4. [Chain] agrees to participate in the above-named study and to provide item-level scanner data to CCMI for the period starting on or about August 20, 1997 and ending on or about October 10, 1997. The scanner data will be for all purchases made during this period at the following stores:

[store
address
address]

and

5. The initial evaluation of [chain's] operating system's environment indicates that [chain] can provide scanner data to CCMI without the need for additional hardware or software modifications at the point of sale (POS). Ten blank computer tapes and mailers will be needed, however, in order to provide scanner data. These materials will be provided to [chain] by CCMI; and
6. Abt Associates Inc. will pay [chain] the fixed sum of \$1,000 at the conclusion of the data collection period to compensate for efforts involved in gathering the aforementioned scanner data. [Chain] will

provide the scanner data to CCMI in a mutually agreed format on a weekly basis for the time period involved; and

7. The purpose of the study is to determine participant food purchasing patterns and not the compliance of food retail stores with federal or state Food Stamp Program regulations; and
8. All parties agree that the data used for this study should remain confidential; and
9. All data provided by [chain] to Abt Associates Inc. or its subcontractor, CCMI, shall be used for the purposes of this study only and shall not be used for any other purposes without the written permission of [chain]; and
10. All original data tapes, once processed and used to create the database, shall be returned to [chain]; and
11. Although the study shall describe the purchasing patterns of participants based both upon participant and store demographics, individual stores shall not be referenced by name or address. Once the transaction data are combined with the food stamp households' demographic information (using information from the EBT system and the state's program files), all store and household identifiers will be stripped from the database. No store or household specific results or information will be reported or otherwise provided to any state or federal government agency; and
12. Data tapes are not a deliverable product required for submission by FCS, and no original data tapes will be provided to FCS at the conclusion of this study. CCMI shall retain all data necessary to comply with federal audit requirements pertaining to activities conducted under this study.

The above is agreed to by the parties as signed below:

[Chain]

Abt Associates Inc.

Date: _____

Date: _____

Name: _____

Name: _____

Title: _____

Title: _____

Signature: _____

Signature: _____

**Data Collection Agreement Between
ABT ASSOCIATES INC.
and
[DCSU Store]**

Abt Associates Inc. and [store] agree as follows:

1. Abt Associates Inc., a Massachusetts corporation headquartered in Cambridge, Massachusetts, has entered into a contract with the United States Department of Agriculture, Food and Consumer Service (FCS) to study the feasibility of using scanner technology to collect data on food items purchased with and without food stamp benefits; and
2. Abt Associates Inc. has obtained the services of Consumer Card Marketing, Inc. (CCMI) of Braintree, Massachusetts, as a subcontractor to conduct the field research and assist in collecting the scanner data and preparing the analysis files; and
3. Georgetown County, South Carolina, has been selected as the project's study site; and
4. [Store] agrees to participate in the above-named study and to provide item-level scanner data to CCMI for the period starting on or about August 21, 1997, and ending on or about October 10, 1997; and
5. CCMI will provide one data collection unit and scanner to [store] for purposes of capturing scanner data during the study period. During the week of August 18, 1997, CCMI will visit your store to do the following: install the data collection unit in a convenient and safe location jointly determined by CCMI and [store]; create an item data file from the current store inventory; and conduct a training session on how to use the data collection unit. Store staff will use the data collection unit during the study period to record item-level information and payment information about all consumer purchases. CCMI will periodically collect the recorded data during the study period. CCMI will retrieve the data collection unit and scanner on or about October 10, 1997; and
6. Upon return of the data collection unit, Abt Associates Inc. will pay [store] the fixed sum of \$2,500 to compensate for efforts involved in gathering the aforementioned scanner data; and
7. The purpose of the study is to determine participant food purchasing patterns and not the compliance of food retail stores with federal or state Food Stamp Program regulations; and
8. All parties agree that the data used for this study should remain confidential; and
9. All data provided by [store] to Abt Associates Inc. or its subcontractor, CCMI, shall be used for the purposes of this study only and shall not be used for any other purposes without the written permission of [store]; and
10. Although the study shall describe the purchasing patterns of participants based both upon participant and store demographics, individual stores shall not be referenced by name or address. Once the data

files are processed, all store and household identifiers will be stripped from the database. No store or household specific results or information will be reported or otherwise provided to any state or federal government agency; and

11. Data tapes are not a deliverable product required for submission by FCS, and no original data will be provided to FCS at the conclusion of this study. CCMI shall retain all data necessary to comply with federal audit requirements pertaining to activities conducted under this study.

The above is agreed to by the parties as signed below:

[Store]

Abt Associates Inc.

Date: _____

Date: _____

Name: _____

Name: _____

Title: _____

Title: _____

Signature: _____

Signature: _____

Appendix D

Installation Guide and End-User Manual for Data Collection Scanning Unit (DCSU)

To support the installation and retailer use of the DCSU, CCMI had to create and document a standardized set of procedures. These procedural steps and programming elements are formalized here into an installation guide and end-user manual.

UPC Codes: Data Processing Steps to Load Additional UPC Codes from Hand-Held Scanner

Plug the hand-held data scanner into its base/charger.

Attach scanner base/charger to notebook computer through serial cable provided.

Make sure the hand-held unit is on.

On Notebook Computer

1. Retrieve UPC_ADD.DAT file from first data collection unit into laptop:
 - Load Apex Program Generator
 - Open USDA2.MAP
 - File, Retrieve, Application 3, OK — this retrieves the upc_add.dat from HH to notebook.

2. Append the UPC_ADD.DAT file to the current master UPC file on the notebook:
 - Load Foxpro Visual 3.0
 - Type in the following commands:
DO ABT
USE UPCADD
APPEND FROM C:\APEX\UPC_ADD.DAT SDF

3. Retrieve UPC_ADD.DAT file from second data collection unit into laptop following instructions in step 1, then:
APPEND FROM C:\APEX\UPC_ADD.DAT SDF

4. Run the UPCADD.PRG program within Foxpro:
DO UPCADD

This will perform the following:

- Remove duplicate records from the UPCADD table
- Update department description and multi-link department number in UPCADD table from DEPTFLE table
- Copy original UPC number to UPC field

- Replace short with “Y” if the UPC number is not eleven digits
 - Expand compressed UPC codes
 - Update newdesc (new UPC item description) from UPCMAST table if UPC description is available
 - Update size with size description from the UPCMAST table if UPC size is available
 - Update the UPCFLE (hand held lookup table) with new UPC codes and descriptions.
5. Download application 2 (UPC Verify) into each of the hand-held units.
 - Open Apex
 - File, Download, Application 2, OK — downloads upcfle to HH from Notebook
 - If asked, set handheld to Run Mode=Yes

UPC Codes: Data Processing Steps to Load Price File Codes from Hand-Held Scanner

Plug the hand-held data scanner into its base/charger.

Attach scanner base/charger to notebook computer through serial cable provided.

Make sure the hand-held unit is on.

On Notebook Computer

1. Load Foxpro Visual 3.0
2. Delete all records in the current UPCUPDT table with the following Foxpro 3.0 command:
USE UPCUPDT
ZAP
3. Download UPC_UPDT.DAT file from first hand-held unit into laptop:
 - Load Apex Program Generator
 - Open USDA2.MAP
 - File, Retrieve, Application 2, OK (this downloads the upc_updt.dat file to the notebook)
4. Append the UPC_UPDT.DAT to the UPCUPDT table on the notebook:
APPEND FROM C:\APEX\UPC_UPDT.DAT SDF
5. Go back to Apex software and retrieve file from second data collection unit into laptop following step 3 instructions:
 - File, Retrieve, Application 2, OK (this downloads the upc_updt.dat to the notebook)
6. Go back to Foxpro Visual 3.0
7. Append the UPC_UPDT.DAT to the UPCUPDT table on the notebook:
APPEND FROM C:\APEX\UPC_UPDT.DAT SDF
8. Run the UPCUPDT.PRG program within Foxpro:
DO UPCUPDT

This will perform the following:

- delete duplicate UPC codes in the UPCUPDT table,
- create a link on oldupc between the UPCUPDT table and UPCADD table,
- update/append records into ADDREC from UPCUPDT table,
- create data file APIDATA.DAT for loading into the POS units.

9. Copy C:\APEX\APIDATA.DAT to a:.

Adding Items to Inventory

*** You **must** have a PS/2 Style Keyboard (with NO ADAPTERS) and **should** have a PS/2 Mouse to execute the following sets of instructions. ***

From the DOS prompt:

1. **CD\APS**
2. **APS**
3. Put CAPS LOCK ON (all entries in the following screens should be done in all caps)
4. Code: 2 (tab)
Password: 2 (tab) (Enter)
5. **I**nventory → Inventory Master File

Here you can search for an inventory item by item number, description, or category. (You can use the tab key to move around the screen to select an item, or if you have a PS/2 style mouse, click on the item you want to select.)

6. Once you find the item you want to change, press **Alt - C** to change an item's UPC code.
7. If you do not find the item in the database, press **Alt - I** to insert the item.
8. Once in the item detail screen, the following fields are important:

- **Item #:** This is the item number, not the description or the UPC #
- **Taxes:**
- **Link SKU:** This is used to link items that will always be “sold” together, like bottle deposits.
- **SKU #**
- **Conv. UM = 1**

then press **Alt-N** (for next page)

- **Food Stamp:** This field will show a “Y” if the item is foodstamp eligible. Otherwise, it will be blank.

then press **Alt-P** to return to previous page

then press **Alt-O**, for OK. This creates the item, or adds it to inventory file.

9. Exit from the Inventory Master File
10. Go to Inventory → Enter receipt/price changes (see also p. 85)
 - Insert
 - Enter item number
 - Tab to **New Retail** and type in the price of the item.
 - Tab to OK and press Enter.
 - Exit from this section.

*** Note: You may also want to make any pricing changes while still in this area. See page entitled “Changing the Price of an Item.” ***

11. Go to Inventory → Post Receipts/Changes
12. Type the following:

Print Price Labels	N
Print Receiving Report	N
Print Shelf Labels	N
Print Show Cards	N
Update Inventory	Y (once to move it to the Yes area, and once again, Y, to select <u>Y</u> es)
Change Register Price	N

(tab) to **OK** - **DO NOT PRESS ENTER YET.**

13. Tab through all fields to fill in the spaces and confirm that it filled in the items the way you wanted them, then tab to **OK** and press Enter. It will then “post” these items to the correct databases.
14. Just to make sure, go into the Inventory Master file and search for the items you just changed/inserted to make sure your changes stuck.
15. Remove CAPS LOCK. You do not want caps lock on for POSALR program. It will lock up the RIVA keyboard.
16. Exit from the APS software to DOS. At a DOS prompt, run the INV.BAT file. This copies all the files you changed to the POSALR directory so you can use them through POS software and automatically re-enters you to the POS Program.

Changing the Price of an Item

*** You **must** have a PS/2 Style Keyboard (with NO ADAPTERS) and **should** have a PS/2 Mouse to execute the following sets of instructions. ***

From the DOS prompt:

1. **CD\APS**
2. **APS**
3. Put CAPS LOCK ON (all entries in the following screens should be done in all caps)
4. Code: 2 (tab)
Password: 2 (tab) (Enter)
5. *(Note: The following steps are identical to Steps 10+ on the Adding items to Inventory directions.)*

Go to Inventory → Enter receipt/price changes
Insert
Enter item number
Tab to **New Retail** and type in the price of the item.
Tab to OK and press Enter.
Exit from this section.
6. Go to Inventory → Post Receipts/Changes
7. Type the following:

Print Price Labels **N**
Print Receiving Report **N**
Print Shelf Labels **N**
Print Show Cards **N**
Update Inventory **Y** (once to move it to the Yes area, and once again, **Y**, to select Yes)
Change Register Price **N**
(tab) to **OK** — ***DO NOT PRESS ENTER YET.***
8. Tab through all fields to fill in the spaces and confirm that it filled in the items the way you wanted them, then tab to OK and press Enter. It will then “post” these items to the correct databases.
9. Just to make sure, go into the Inventory Master file and search for the items you just changed/inserted to make sure your changes stuck.

10. Remove CAPS LOCK. You do not want caps lock on for POSALR program. It will lock up the RIVA keyboard.

Cashier Tutorial

Signing On

- When prompted for clerk #, press 1.
- When prompted for password, press 1. System will blank out and re-enter the program.
- When prompted for password again, press Clear.

Ring a Sale

- Scan item or items

Ring a Sale on Open Department (If Item Does Not Scan)

- Key in price of item. Always key in amounts without decimal point (i.e. 500 = \$5.00).
- Press the corresponding department key of item.

Using Quan. Key to Ring in Multiple Items

- Key in number of items.
- Press Quan. key
- Scan item or ring under open department.

Using the At/for Key

Enter the Number of items being purchased

- Press the Quan. key
- Enter the "AT" quantity and press At/For key
- Enter the price and press department key
- Example (to ring 2 of an item that sells 3 for a dollar in meat department):
 - 2 Quan
 - 3 At/For
 - 100 meat

Voiding a Single Item (can only be done while in current transaction)

- Use arrow keys to scroll up or down to the item you wish to void
- When item is highlighted, press the Void key.

Voiding an Entire Sale Transaction

- Press Order Cancel key.

Note: This will void all items currently being rung, but will not clear the food stamp totals. You will still be in the same transaction number. It is ok to ring your next sale on this transaction number. The food stamp totals will re-adjust after you enter the next item.

Tendering a Sale with Food Stamps Only

**** NOTE: Total amount in food stamps section is the amount allowed for food stamps.*

- Key in total amount of Food Stamps tendered. Always key in amounts without decimal point (i.e. 500 = \$5.00).
- Press Food Stamp Key.

Tendering a Sale with Other Tender Only

- Key in total amount of Other (cash, credit card) tendered. Always key in amount without decimal point (i.e. 500=\$5.00).
- Press Other Tendered Key.

Tendering a Sale with Food Stamps and Other Tender

**** NOTE: Total amount in food stamps section is the amount allowed for food stamps.*

- Key in total amount of Food Stamps tendered. Always key in amounts without decimal point (i.e. 500 = \$5.00).
- Press Food Stamp Key.
- Key in total amount of Other (cash, credit card) tendered. Always key in amount without decimal point (i.e. 500=\$5.00).
- Press Other Tendered Key.

Other Tips

- If you find you are at a plain blank screen, with just C:\POSALR> at the top of the screen, you have exited the program. Just type in the number 1 and press Enter to return to the program.
- The scanner unit will go into “sleep” mode when it has not been used for awhile. If this happens, you can “wake it up” by passing your hand in front of it. If you find that the scanner still will not scan, check to make sure that it has not been turned off. There should be a little green light on the top of the base unit. If this light is not on, press the button just below the light to turn the unit back on.

Cashier Data Changes Form

The cashier changes form on the next page was designed to enable each store to document specific information when adding inventory items or changing pricing.

Two of the stores using the DCSUs did use the form for updates; the other two stores did not.