

Contract No. 53-3198-3-120

MPR Project Number 7553-600

Author: Pat Doyle

POTENTIAL OF STATE AUTOMATED
ELIGIBILITY SYSTEMS FOR PROGRAM ANALYSIS

March 29, 1985

Prepared For:

U.S. Department of Agriculture
Food and Nutrition Service
3101 Park Center Drive
10th Floor
Alexandria, Virginia 22302

Prepared By:

Mathematica Policy Research
600 Maryland Avenue, S.W.
Suite 550
Washington, D.C. 20024

CONTENTS

<u>Chapter</u>	<u>Page</u>
I. INTRODUCTION.....	1
Organization of the Report.....	3
II. OVERVIEW OF INFORMATION NEEDS.....	5
ANALYSIS OF PROGRAM CHARACTERISTICS.....	6
Economic Status and Program Targeting.....	6
Impact Analysis of Program Changes.....	7
Prospective Analyses of Proposed Program Change.....	8
Effect of Interactions With Other Transfer Programs.....	8
Effect of Changing Economic Conditions.....	9
ANALYSIS OF BEHAVIORAL CHARACTERISTICS.....	10
Program Dynamics.....	11
Welfare Dependency.....	12
Participation.....	12
III. OVERVIEW OF THE STATE DATA SYSTEMS.....	14
State Data System.....	14
Potential Uses of SDS.....	15
Advantages and Disadvantages.....	17
IV. COMPARISON OF STATE SYSTEMS WITH ALTERNATIVE DATA SOURCES FOR FOOD STAMP PROGRAM POLICY ANALYSIS.....	20
POINTS OF COMPARISON.....	21
ANALYSIS OF PROGRAM CHARACTERISTICS.....	22
Characteristics of Participants.....	25
Income And Asset Eligibility - What If?.....	27
Income And Asset Eligibility - What Happened?.....	30
Level And Distribution Of Benefits - What If?.....	32
Level And Distribution Of Benefits - What Happened?.....	34
Interaction With Other Programs.....	36
Effects Of Changing Economic Conditions.....	38
Analysis of Gainers and Losers.....	40
Summary.....	42
ANALYSIS OF PROGRAM BEHAVIOR.....	43
Turnover In The Food Stamp Program.....	46
Program Entry and Exit Rates.....	47
Welfare Dependency.....	49
Determinants Of Participation.....	50
Summary.....	51
START UP COSTS ASSOCIATED WITH EACH SURVEY.....	51

Table of Contents cont'd...

V.	KEY CONSIDERATIONS IN DEVELOPING SDS.....	54
	Frequency of Data Collection.....	54
	State Case Numbering System.....	55
	Sampling.....	58
	Sample Size.....	61
	APPENDIX.....	65
	CURRENT POPULATION SURVEY.....	66
	INTEGRATED QUALITY CONTROL SYSTEM.....	71
	MICHIGAN PANEL STUDY ON INCOME DYNAMICS.....	76
	THE NATIONWIDE FOOD CONSUMPTION SURVEY.....	80
	SURVEY OF INCOME AND PROGRAM PARTICIPATION.....	84

ACKNOWLEDGEMENTS

The author wishes to thank Harold Beebout, Ted Cavin, and Audrey Mirsky for their assistance in the preparation of this report. Dr. Beebout was the principal architect of the two tables on which Chapter IV is based. Along with Ms. Mirsky he also provided excellent editorial assistance. Dr. Cavin provided consultation on the determination of sample sizes in Chapter V.

The author also wishes to thank Donna Leiss and Lucia Wesley for typing the manuscript.

I. INTRODUCTION

Analysis of the Food Stamp program caseload characteristics and the impacts of legislative reform are currently conducted using a variety of techniques and data sources. General purpose national surveys of households are used to simulate the impact of changes in eligibility and benefit formulas particularly when the changes are likely to increase the pool of potentially eligible households. Interactions with other means tested programs are often measured with these nationally representative surveys as well. Data from the Integrated Quality Control Survey (IQCS) are used to analyze caseload characteristics and to measure changes in the composition of the caseload over time. These data are also used to simulate the impact of reforms to the program which are not expected to add new participant households. Impact studies of earlier program changes such as OBRA are often conducted using specially created samples of administrative records created through manual case record abstraction. The IQCS and manual case record abstraction files are the principal source of data used to analyze the results of the implementation of reforms to the Food Stamp Program and other programs with which it interacts.

Program dynamics and dependency issues have been studied using the subset of general purpose surveys which are longitudinal, that is, they follow units over time and observe actual changes in household (or person) behavior. Manual case record abstraction has also been used to produce data for longitudinal analysis of the food stamp population over short periods of time. This method has been used to measure changes in food stamp entry and exit rates due to program reforms.

General purpose surveys have been used to develop models of food stamp participation behavior. These models were built to explain the behavior patterns of potential food stamp participants as well as to determine the outcome of a participation decision in the simulation of food stamp costs and caseloads. Studies of food consumption and expenditures have been conducted using the limited number of general purpose surveys which collect food expenditures and nutrition information.

This current stock of data used for program analysis, although extensive, is not comprehensive in terms of meeting the needs of all research areas of concern to FNS. Some surveys provide the right data elements but the universe is too restrictive for a particular study. Others provide the right universe but not a complete set of data elements. Still others suffer from either not being targeted to the time frame of the analysis or are very expensive to use.

The focus of this paper is the feasibility of enhancing the current stock of data for program analysis in order 1) to more accurately answer a set of questions that existing data are not well suited to address and 2) to reduce the need for new data collection and its associated cost. In particular, the paper identifies program analysis areas for which existing data are not well suited and there is potential for filling those gaps with routine collection and processing of administrative data from states with automated caseload processing systems. This data collection effort differs from the administrative data collection activities mentioned earlier in several ways. The content will be similar to that of the manual case record abstraction used for earlier studies but once the system of acquisition is in place, the cost of creating analysis files for a

particular study may be significantly reduced. The new collection effort would be different from the IQCS in that the system could be designed to support longitudinal analysis (IQCS is essentially a series of cross section samples of food stamp case records), sample sizes can be large enough to observe rare events or units with rare characteristics, and there is potential to expand the set of data elements.

This new data collection effort would not replace the need for other data sets currently in use. For example, it cannot fulfill the data needs of participation studies. Furthermore, it will not provide national estimates of caseload composition and hence would not eliminate the need for the IQCS.

Organization of the Report

The following chapter summarizes the types of policy analysis relevant to the food stamp program and attempts to categorize the data needs for each of these. This categorization facilitates the comparison of data sources in terms of their suitability for each of the analytic areas. The third chapter of the report presents an overview of the new data collection effort, its potential uses for program analysis and its advantages and disadvantages. The final chapter demonstrates that the new data source will more accurately answer a set of questions for which existing data are not well suited and, hence, reduce the need for manual case record abstraction. This is accomplished by comparing the suitability of the new data collection effort to other sources for each of the analytic areas summarized in Chapter II.

The existing data sources to which the new data collection effort are compared are summarized in an appendix. The report concludes with a chapter discussing key features of the new data collection effort which should be taken into consideration when the system is designed.

II. OVERVIEW OF INFORMATION NEEDS

Food Stamp Program policy analysis needs can be classified into two major groups. The first deals with issues such as who the program is now serving, how the group being served relates to the target population, the level and distribution of benefits and how program participation, costs and the distribution of benefits would vary if program rules were changed. The second group of analysis needs deals with how households react to the program in terms of deciding to participate, changing their work and food purchase behavior, and their incentive and ability to leave the program. This categorization of analysis needs is particularly convenient for this evaluation because it is broadly parallel to the type of data required for studies of the issues. These two classifications of analyses around which this chapter is organized are:

- o Analysis of Program Characteristics. This area is inclusive of studies affecting the program eligibility requirements, the level and distribution of benefits, the interaction of the program with other federal programs and with the changing economic environment, and the characteristics of participants either in total or just those affected by program change.
- o Analysis of Behavioral Characteristics. This area covers studies of program entry and exit rates, turnover, duration of participation, welfare dependency and the decision to participate in the program

The data requirements for each of these research areas are outlined below.

ANALYSIS OF PROGRAM CHARACTERISTICS

These issues are addressed with both prospective studies i.e., "What would happen if..." and retrospective studies, i.e., "What happened

when..." The former usually employ simulation models of program reform whereas the latter require observation of an event after it has occurred. Both of these two types of studies are accompanied with an analysis of the characteristics of households participating in the food stamp program.

Economic Status and Program Targeting

One of the most fundamental needs of program policy is to describe the characteristics of households authorized to receive food stamps either under existing regulations, other alternative regulations or under alternative economic circumstances. This need may be to describe the total caseload or a particular subset and can occur in conjunction with every other analysis area discussed in this chapter. The data requirements are principally cross sectional and the degree to which caseload can be characterized depends on the following criteria

- o Content. The characteristics by which participants can be described are lengthy, including determinants of eligibility and benefits, adequacy of dietary intake, demographic characteristics of the unit and its members, participation of one or more household members in other programs, the ratio of income to need measured in various ways, and labor force activity of unit members. The need for one or more of these elements varies considerably depending on the objectives of the study.
- o Sample. Analysis of the characteristics of participants only requires observation of participants unless the objective of the study includes contrasting participants with either eligible non participants or the population in general. The latter require larger universes. Examination of participants with rare characteristics requires high sampling ratios.
- o Useability. The data should be targeted to the point in time that characteristics are to be observed. The data must permit the analysis to be conducted within time and budgetary constraints.

Impact Analysis of Program Changes

In order to observe what happened as a result of a change in the program, accurate data is required on households authorized to receive food stamps in the months before and after the change occurred. The data can either consist of pre- and post-implementation cross sections or a longitudinal case history file depending on the statistical techniques to be employed and the goals of the study. Regardless of the nature of the data (cross sectional versus longitudinal) the success of the study of program impacts depends on several criteria.

- o Content. This refers to the accuracy of reported level of benefits both before and after the change, the ability to replicate the eligibility and benefit determination process, the ability to distinguish program impacts from other exogenous events, and the ability to describe the demographic and economic characteristics of the units both before and after the program reform.
- o Sample. In addition to requiring that the data include participants before and after the change, the study requires reliable representation of units directly affected by the program change. Some studies may require representation of the national program caseload whereas others may require the analysis of specific population subgroups and still others may require both.
- o Useability. The more targeted the reference period of the data base to the time in which the impact occurred, the more definitive the study can be. The data must be selected to allow the study to be conducted within budget and time constraints.

Prospective Analyses of Proposed Program Change

Simulation of the outcome of a proposed reform requires cross sectional data on current participants as well as all non participating households potentially affected by the proposed change. The success of the simulation study depends on the following criteria.

- o Content. Simulation studies require the ability to accurately measure the relative change in monthly caseload and benefits more so than perfect replication of the before and after benefits. These studies also need to describe the demographic and economic characteristics of the units potentially affected by the reform.
- o Sample. To measure the total impact of any reform the sample must include eligible units under both the current program and the proposed alternatives. To study the impact on current caseload without regard to behavioral response, only observation of participants under the current program is required.
- o Useability. The reference period of the data base should reflect the economic circumstances expected to be in effect when the proposed reform would be implemented. The data must permit the study to be conducted within time and budgetary constraints.

Effect of Interactions With Other Transfer Program

Studies of program interactions have a requirement for more content than the study of actual or proposed food stamp program impacts which neglects these interactions. These studies can require either cross sectional or longitudinal data depending on the statistical techniques to be employed and the objectives of the study. Program interactions are more adequately measured with monthly data than with annual. The data requirements for the study of program interactions are

- o Content. At a minimum the data must include benefits from the program with which the food stamp program interacts. However, studies of this interaction are enhanced if the components of the eligibility and benefit formulas for these other programs are included. In addition to Food Stamp program information requirements, demographic and economic data on the observations enhances the ability to disaggregate the population according to characteristics of participants in one or more of the programs and to distinguish program impacts from other events.

- o Sample. The universe must include participants in all programs of interest. If the study is focusing on the impact of an actual change in another program on the food stamp program, the study requires participants in all programs before and after the change. The simulation of the expected outcome of such a reform on the food stamp program requires units potentially eligible in all programs under either scenario.
- o Useability. To capture the full effect of changes in other programs the reference period of the data must be sufficiently long to permit the observation of behavioral responses to the change. A series of cross sections suitably far apart in time will support the observation of the net impact of program changes inclusive of behavioral responses, if there are no major changes in the economy occurring simultaneously. Longitudinal data are preferable, however, to measure the gross impact of program changes and to net out program changes from other exogenous events. To simulate effects on the food stamp program of such external events, the data need to reflect the economic conditions expected to be in effect when the event occurs. The data must permit the study to be carried out with time and budgetary constraints.

Effect of Changing Economic Conditions

Analyzing the effect of changing economic conditions on the food stamp program requires a measure of this change as well as its outcome in terms of food stamp program eligibility and benefits. Because of the expected lag between the observation of a changing economy at the macro level and its ultimate effect on individual participants, longitudinal data are most suited to this analysis area. The criteria for studies of economic effect are

- o Content. As stated above the reference period for each observation should cover a relatively long period of time. The data must include a measure of the impact of a changing economy on monthly income, employment, assets, and living arrangements as well as the components of the Food Stamp eligibility and determination process. Demographic data enhance the ability to target the analysis to individuals or units

expected to be most severely affected by swings in the economy.

- o Sample. The universe for a study of the total impact must include eligibles and participants during the time it takes for the full effect to be felt. The universe for studies restricted to the impact on caseload in existence at the beginning of the cycle can be limited accordingly.
- o Useability. The reference period must cover the calendar period during which the events took place and the household sector responded. In the absence of a change in the food stamp program or another program with which it interacts, a series of suitably timed cross sections will be adequate to describe the net effect of economic changes. However, it is often the case that program reforms do occur during these periods making longitudinal data more appropriate.

ANALYSIS OF BEHAVIORAL CHARACTERISTICS

Households react differently to the Food Stamp Program and reforms to the program. Analyses of these behavioral responses are needed to evaluate how well the program is serving the needed population, to assess how units change their economic circumstances or food consumption response to program reform, and to measure the effectiveness of incentives to reduce the demand for food stamps. Results of these analysis can be used to develop the behavioral components of simulation models designed to address the "what if" questions described previously.

Program Dynamics

Program dynamics refers to the rate at which units enter and leave the food stamp program and the turnover in aggregate caseload in a given year. The study of program dynamics requires longitudinal data. Some analysis may be restricted entirely to participants whereas others may be concerned with eligible units as well. The data requirements for the study

of program dynamics are listed below.

- o Content. The length of the reference period for each observation varies depending on whether or not the analysis is long term but it always covers more than one month. The basic information requirements are food stamp participation and benefits as well as economic and demographic characteristics correlated with food stamp participation. The latter are needed to describe events that trigger a change in participation status and to enhance the ability to target the analysis. For studies of eligible units the data must also include the components of the eligibility and benefit determination process. Monthly data are preferred.
- o Sample. At a minimum, studies of turnover and exit rates can be conducted using a sample of units participating at least one month in a specified calendar period which varies according to the goals of the study. Calculation of entry rates and studies of the dynamics of the eligible population require a larger universe.
- o Useability. Cost is of particular concern because the requirement that the data be longitudinal and the desire that they be arranged by month necessarily imply that the data sets are large and expensive to compile and access. This extensiveness of the data often means that they may not be available very quickly and that the software necessary to conduct the analysis is time consuming to develop.

Welfare Dependency

Studies of welfare dependency analyze the duration of program participation taking into account the difference between long- and short-term participants. As is true for program dynamics, these studies require longitudinal data. However, the length of the needed reference period is often very long, particularly when the objective is to study chronic

participants. Monthly data are essential to study duration for units with weak attachment to the program, but less essential for the longer term

- o Content. For studies targeted to chronic participants the reference period should be very lengthy. For studies concentrating on transitory participants monthly information over one or two calendar years will suffice. The information requirements are, minimally, program participation and benefits and measures of correlated characteristics.
- o Sample. The universe consists of households authorized to receive food stamps for at least one month during the study period. The larger the sample size, the more flexibility in targeting the analysis to the population of interest.
- o Useability. As is true for program dynamics, studies of long term dependency could be very expensive and time consuming, particularly if monthly data are used.

Participation

This analysis area includes the description of why some eligible units participate and others do not, how units vary work patterns to maximize benefits under the program and how units alter their food consumption behavior according to the availability of food stamps. These studies are essential to capture the behavioral component in the simulation of the outcome of all "what if" questions. Developing models of the behavior patterns of program participants or even just a simple estimate of the overall participation rate requires the observation of all units eligible for the program. Longitudinal data are not essential for this type of analysis but they are useful. The data requirements for studies of the determinants of participation are

- o Content. The elements necessary are the components of the eligibility and benefit determination process as well as an account of actual program participation and benefits. Demographic and economic data correlated with the decision to participate are desirable but the extent of the need for these elements depends on the complexity of the behavioral model. Food consumption and expenditures are needed if the analysis focuses on dietary intake.

- o Sample. The universe must contain all eligible units. The sample size requirements vary along with the need for detail on demographic and economic characteristics.

- o Useability. The reference period should be inclusive of some period when the current program is in effect. Longitudinal data covering periods of program reform or changing economic conditions permit the behavioral model to capture events that may trigger a change in participation status. However, the latter are only necessary for more sophisticated multivariate models. Cost and time constraints affect the degree of the complexity of the model and hence the selection of the appropriate data.

III. OVERVIEW OF THE STATE DATA SYSTEMS

As noted in the introduction, this paper focuses on how a potential new data source can enhance the current stock of data for the analysis described in the preceding chapter. This new data source referred to as the State Data System (SDS) is actually a continuing series of data extracted from administrative records of the Food Stamp Program. It is therefore a potentially useful data source for studies which require detailed information on the components of the program eligibility and benefit determination process. An overview of this system is provided followed by a description of the analytic areas for which it is a potentially useful data source and its advantages and disadvantages for these studies.

State Data System

This system represents the routine acquisition, archival, and use of data from the master case record files of households authorized to participate in the FSP. An arrangement with several states would be made to extract case record files at regular intervals. Once received, the data would be validated and individual case identifiers would be deleted to maintain confidentiality, if the states had not already done so. Then analysis files would be created by restructuring the data elements to establish a degree of uniformity across states and to select a sample of the caseload from each state. The analysis files would then be integrated both across states and across time within a data access system. Once completed the data would be accessed as needed for program analysis.

The ultimate products of the SDS would be analysis files of food stamp participants containing a complete set of information on the components of the eligibility and benefit determination process, the food stamp allotment, and demographic characteristics to enable the identification of population subgroups of interest to a particular study. The reference period for each observation would be expanding as the system continues to operate with repeated acquisitions so that eventually a complete case history could be established. At this time it is envisioned that the analysis files will contain a representative subset of cases from each state. Sample size determination would be based on the need to examine rare subgroups and impacts of program change on these small groups. Furthermore, appropriate sampling techniques will be employed to permit linking of data for each food stamp unit across time as well as to maintain a representative sample over time. Since the system will be restricted to a subset of states chosen to facilitate the process, the pooled samples will not be representative of the nation. However, it may be possible to select the states such that the distribution of food stamp units on key characteristics will be roughly typical of the national caseload.

Potential Uses of SDS

There are two types of analysis files which can be generated from SDS: cross sectional files of Food Stamp cases and longitudinal case history files. Hence, the data are potentially useful for a broad range of applications. Furthermore, due to the continuing nature of the system, these files will be available for use soon after an analysis area is identified.

The series of cross sectional files will be appropriate for examining changes in the distribution of caseload over time or the impact of a legislative reform to the program. SDS allows the examination of the changes in distribution shortly after the changes occur and it will permit the time frame of the analysis to be keyed directly to the implementation schedule within each state.

The system is particularly well suited to respond to queries about population subgroups served by the program. This data system will provide statistically reliable estimates of small subgroups within each of the states selected (even a census rather than a sample is feasible) and hence, will permit the simulation of program reforms and the observation of the actual impact of reforms affecting these small groups. The system also allows examination of the fluctuations in the distribution of participants in response to external forces such as changes in local unemployment rates or changes to other programs with which the food stamp program interacts. Again, with the increase in reliability of small group estimation, the analysis of the effects of these external conditions can be targeted more directly to specific sets of participants.

In addition to its usefulness in providing a series of cross-section data bases for the analysis of caseload composition, SDS will yield longitudinal case record files with which program exit rates can be measured as well as patterns of program participation and turnover. Eventually the system will permit measurement of duration of food stamp receipt for all but the most chronic participants. However, this feature is a function of the length of time the system is supported. The system will provide an opportunity to link changing economic conditions with

changing caseload dynamics and to measure the impact of program reform either to the FSP or to other interrelated programs on exit rates.

The system will support measurement of change at the case level which can then be aggregated over units experiencing the change to examine the total impact of program reform. This represents a more powerful methodology than a series of cross-section measures where net change is modeled as a function of pre- and post-measures.

Advantages and Disadvantages

SDS originates from administrative data on food stamp cases. Hence, it offers several inherent advantages over the general purpose surveys to which it is compared in the next chapter. Survey data in general do not provide all components of the eligibility and benefit determination process as SDS and the other administrative sources do. Furthermore, respondents to general purpose surveys do not always provide the same information to the interviewer as they provide to the case worker. SDS contains the latter. Regardless of which responses are actually correct, the administrative data may be better for analyzing certain food stamp policy issues such as the cost of program changes. Respondents in general purpose surveys can also refuse to answer some or even all questions. This missing information is a source of error even though the producer or the user of the data compensates through techniques such as imputation. SDS and other administrative sources have this problem to a much lesser extent since applicants who do not respond fully cannot participate. Another inherent advantage of the case record data is that the universe is better targeted to program participants. While administrative practices in some states may make it difficult to access

information on cases with very short certification periods, general purpose survey data entirely omit observations outside their universe.

SDS offers some advantages over the other administrative sources in addition to its inherent advantages over general purpose survey data. It potentially provides a more comprehensive set of data elements than the current IQCS. Furthermore, the system can cost effectively provide statistically reliable estimates of small population groups. IQCS is not well suited for this purpose and manual case record abstraction can only support these estimates at a very high cost.

Other advantages of SDS include the ability to effectively target the reference period according to the needs of the analysis. In particular SDS can be used to produce a series of cross section snapshots of the population over a precisely defined time period. Over time the system also provides longitudinal data covering reference periods suitable for all but very long term studies.

There are three major disadvantages of SDS, two are inherent because the data originate with administrative case records and the third results from the design of the system. The two disadvantages common to all administrative sources are the lack of representation of the non participant population and the limited detail on unit characteristics not directly needed to determine program eligibility and benefits. The third disadvantage is that the sample is restricted to a subset of states which are not randomly selected and therefore pooling the state samples does not yield nationally representative estimates.

Another disadvantage is that SDS, as we envision the initial system, does not provide complete information on other programs with which the Food Stamp Program interacts. The information SDS does supply would likely be restricted to benefits carried in the system from those programs by units who participate in the Food Stamp program. This restriction could eventually be relaxed for states with one unified system containing the case record data for other programs as well as food stamps. IQCS has the same disadvantages which could also be overcome.

IV. COMPARISON OF STATE SYSTEMS WITH
ALTERNATIVE DATA SOURCES FOR FOOD STAMP PROGRAM
POLICY ANALYSIS

As described in Chapter III, SDS is a potentially useful tool for analysis of Food Stamp Policy. This chapter compares the utility of SDS with existing data sets currently used by FNS for research in the areas of program and behavioral characteristics. Existing data sets are grouped into two classes: general purpose surveys of households, and information extracted from Food Stamp case records. The first class is extremely broad including numerous nationally representative surveys conducted by the Census Bureau and other institutions through government contracts or grants. For purposes of this paper, however, the discussion is limited to a few surveys which have been used or planned to be used by FNS for program analysis. These are

The Survey of Income and Program Participation (SIPP)¹

The March Current Population Survey (CPS)

The Michigan Panel Study on Income Dynamics (PSID)

National Food Consumption Survey (NFCS)

For purposes of this discussion the second class of data includes the IQCS and other administrative data potentially available through manual case record abstraction. Surveys of administrative data on other programs such

¹The original scope of this report included comparison to the 1979 Income Survey Development Program Research Panel (ISDP). This was omitted since ISDP was a precursor to SIPP and the potential uses of SIPP are broader.

as AFDC are not candidates for inclusion here because they exclude at least some portion of the food stamp caseload. Existing data sets which are also excluded from this discussion are those collected for previous studies, pertained to some historical time period, and were in most instances specific examples of the general class of data sets referred to as manual case record abstraction. Each of these existing data sets is summarized in the appendix.

This chapter is organized into four sections. The first describes the manner in which SDS and other data sets are compared in terms of their suitability for use for each series of topics. The second two sections provide the actual comparison and are divided according to the classification of analyses into the two broad areas outlined in Chapter II. The final section addresses start up costs associated with the use of each of the data sets discussed in this chapter.

POINTS OF COMPARISON

In each of the next two sections, SDS is compared with other surveys on the basis of three characteristics. The first characteristic is content which refers to the list of data elements available for use and their appropriateness for each analysis area. The second characteristic is sample which refers to the number of observations in the entire sample, the representativeness in terms of producing national estimates, and the appropriateness of the sample design for analyzing various population groups. The final area is useability. This area represents an assessment of both the cost of using the data and its timeliness. Cost is measured in terms of using the data for a typical study and is presented in relative terms. Case record abstraction costs discussed here are inclusive of

collection and preparation principally because the collection effort is typically geared to the issue at hand. For the other surveys, collection and initial data preparation costs incurred are external to particular analyses and, therefore, are not counted in the discussion of relative cost. Instead they are discussed in the fourth section of this chapter. Timeliness refers to how quickly the data can be made available for a particular study as well as how appropriate the time frame of the data are to the study in question.

ANALYSIS OF PROGRAM CHARACTERISTICS

As described in Chapter II this area is inclusive of studies of the effects of changing the program eligibility criteria, the level and distribution of benefits, the interaction of the program with other federal programs and with the changing economic environment, and characteristics of participants affected by program change. For most of these issues, the studies conducted can be generally classified into two types. One type raises the basic question - "What would happen if" and the other type asks "What did happen as a result of ...". The first class of studies generally involves the simulation of an event and the analysis of the simulated outcome. The second class of studies analyzes the observed outcome after the event took place. All of these studies require data which permit the analysis of the characteristics of Food Stamp households. Table 1 compares SDS to five other data sources in terms of its suitability for analyzing these program characteristics. The comparisons are elaborated below for each analysis area.

TABLE 1: COMPARISON OF STATE SYSTEMS WITH ALTERNATIVE DATA SOURCES FOR ANALYSIS OF PROGRAM CHARACTERISTICS

	STATE SYSTEMS			INT. QUALITY CONTROL			CASE RECORD TRANSC.		
	Content	Sample	Useability	Content	Sample	Useability	Content	Sample	Useability
A. Characteristics of Participants	Good, Limited demographics	Excellent size, Not representative	Very Timely, Moderate cost	Good, Very limited demographics	Adequate size, Representative	Less timely, Low cost	Good, Limited demographics	Needed size, May be representative	Very timely, High cost
B. Program Eligibility									
1. Income Eligibility									
-What if questions	Limited to current participants	Excellent size, Not representative	Very timely, Moderate cost	Limited to current participants	Adequate size, Representative	Timely, Low cost	Limited to current participants	Needed size, May be representative	Very timely, Very high cost
-What happened questions	Excellent, Longitudinal cap.	Excellent size, Not representative	Very timely, Moderate cost	Good for cross section comparisons	Adequate size, Representative	Less timely, Low cost	Excellent, Longitudinal cap.	Needed size, May be representative	Very timely, Very high cost
2. Asset Eligibility									
-What if questions	Limited to current participants	Excellent size, Not representative	Very timely, Moderate cost	Limited to current participants and countable assets	Adequate size, Representative	Timely, Low cost	Limited to current participants	Needed size, Limited representation	Limited data, Very costly, Very timely
-What happened questions	Excellent, Longitudinal cap.	Excellent size, Not representative	Very timely, Moderate cost	Good for cross section comparisons	Adequate size, Representative	Less timely, Low cost	Excellent, Longitudinal cap.	Needed size, May be representative	Very timely, Very high cost
C. Level and Distribution of FSP Benefits									
1. Allotment Standards									
-What if questions	Limited to current participants	Excellent size, Not representative	Very timely, Moderate cost	Limited to current participants	Adequate size, Representative	Timely, Low cost	Limited to current participants	Needed size, May be representative	Very timely, Very high cost
-What happened questions	Good, No food exp., Longitudinal cap.	Excellent size, Not representative	Very timely, Moderate cost	Good for cross section comparisons No food exp.	Adequate size, Representative	Less timely, Low cost	Good, No food exp., Longitudinal cap.	Needed size, May be representative	Timely, Very high cost
2. Benefit Formula									
-What if questions	Limited to current participants	Ample size, Not representative	Very timely, Moderate cost	Limited to current participants	Adequate size, Representative	Timely, Low cost	Excellent, Limited to cur. partic.	Needed size, May be representative	Very timely, Very high cost
-What happened questions	Excellent, Longitudinal cap.	Excellent size, Not representative	Very timely, Moderate cost	Good for cross section comparisons	Adequate size, Representative	Less timely, Low cost	Excellent, Longitudinal cap.	Needed size, May be representative	Very timely, Very high cost
D. Interaction with other programs, primarily AFDC	A few state systems have integrated longitudinal info.	Excellent size, Not representative	Very timely, Moderate cost	Poor, data not integrated ???	Sample design problem ???	Limited use, Low cost	Excellent, Tracking prob. in some sta.	Needed size, May be representative	Very timely, Very high cost
E. Effect of changing economic conditions	Excellent, Longitudinal cap.	Excellent size, Not representative	Very timely, Moderate cost	Poor, not longitudinal	Adequate size, Representative	Limited use, Low cost	Excellent	Needed size, May be representative	Very timely, Very high cost
F. FSP Cost & Analysis of Gainers/Losers;									
-What if questions	Limited to current participants	Ample size, Not representative	Awkward, High moderate cost	Limited to current participants	Adequate size, Representative	Less timely, Low cost	Excellent	Needed size, May be representative	Timely, Very high cost
-What happened questions	Excellent	Excellent size, Not representative	Very timely, Moderate cost	Poor, not longitudinal	Sample design problem??	Limited use, Low cost	Excellent	Needed size, May be representative	Very timely, Very high cost

	SIIP			CPS			NFCS		
	Content	Sample	Useability	Content	Sample	Useability	Content	Sample	Useability
A. Characteristics of Participants	Good, Limited expenses	Small size, Representative	Timely, Cost varies	Good, no expenses	Good size, Representative	Timely, Moderate cost	Good, has food consumption	Small size, Representative	Not timely, Costly to use
B. Program Eligibility									
1. Income Eligibility									
-What if questions	Excellent detail, Missing few deduct.	Moderate size, Representative	Timely, Cost varies with use	Good, Missing most deductions	Good size, Representative	Timely, Moderate cost	Limited, Income info. not detailed	Small size (3000), Representative	Not timely, Costly to use
-What happened questions	Good, Limited longitudinal cap.	Small size, Representative	Less Timely, Cost varies	Good for cross section comparisons	Good size, Representative	Less Timely, Moderate cost	Limited, No survey since 1979	Small size, Representative	Not timely, Costly to use
2. Asset Eligibility									
-What if questions	Good, Asset data quality fair	Moderate size, Representative	Timely, Cost varies	Limited, Missing asset variables	Good size, Representative	Timely, Moderate cost	Limited, Asset info. not detailed	Small size, Representative	Not timely, Costly to use
-What happened questions	Good, Limited Longitudinal cap.	Small size, Representative	Less Timely, Cost varies	Very limited	Good size, Representative	Timely, Moderate cost	Limited, No survey since 1979	Small size, Representative	Not timely, Costly to use
C. Level and Distribution of FSP Benefits									
1. Allotment Standards									
-What if questions	Very good, Survey data differ from ad.	Moderate size, Representative	Timely, Cost varies	Very good, Survey data differ from ad.	Good size, Representative	Timely, Moderate cost	Good, Potential for est. nutrition imp.	Small size, Representative	Not timely, Costly to use
-What happened questions	Good, No food exp., Longitudinal cap.	Small size, Representative	Less Timely, Cost varies	Good for cross section comparisons No food exp.	Good size, Representative	Less timely, moderate cost	Limited, No survey since 1979, Info. on nutrition impacts	Small size, Representative	Not timely, Costly to use
2. Benefit Formula									
-What if questions	Excellent	Moderate size, Representative	Timely, Cost varies	Very good.	Good size, Representative	Timely, Moderate cost	Limited, Detail lacking	Small size, Representative	Not timely, Costly to use
-What happened questions	Very good, Limited Longitudinal cap.	Small size, Representative	Less Timely, Cost varies	Good for cross section comparisons	Good size, Representative	Less timely, Moderate cost	Limited, Info. on nutrition impacts	Small size, Representative	Not timely, Costly to use
D. Interaction with other programs, primarily AFDC	Excellent for both "what if" and "what happened"	Small size, Representative	Timely, Cost varies	Good for "what if" questions	Good size, Representative	Timely, Moderate cost	Limited info. on other programs	Small size, Representative	Not timely, Costly to use
E. Effects of changing economic conditions	Excellent for "what if"	Moderate size, Representative	Timely, Cost varies	Good for "what if"	Good size, Representative	Timely, expensive for "what if"	Poor	Not applicable	Not applicable
F. FSP Cost & Analysis of Gainers/Losers;									
-What if questions	Excellent detail, Longitudinal cap.	Moderate size, Representative	Timely, Cost varies	Very good participants	Good size, Representative	Timely, Moderate cost	Limited detail	Small size, Representative	Not timely, Costly to use
-What happened questions	Good, Longitudinal cap.	Small size, Representative	Timely, Cost varies	Poor	Not applicable	Not applicable	Poor	Not applicable	Not applicable

Characteristics Of Participants

The most fundamental need for data on the food stamp caseload is to describe the characteristics of units participating in the program. This is of general interest as well as necessary in the discussion of actual and simulated reforms to the food stamp program and other programs with which it interacts. This section discusses SDS and other surveys in terms of the ability to generate tables showing the distribution of food stamp units by various economic and demographic characteristics. The principal conclusion is that all the surveys are good in at least one respect but none can support the examination of food stamp recipients in all characteristics of interest.

The content of administrative surveys includes excellent detail on deductible expenses, countable income and countable assets of units participating in the program. This permits caseload to be arrayed by important eligibility determinants such as receipt of earnings, level of assets or the cost of shelter. However, the sample universe for these surveys lacks representation of non participants in order to assess how well the program is targeted. SDS cannot support national statistics but can support more finely arrayed tables than the other administrative sources because of its high sampling ratios. All of the administrative sources are limited in their identification of unit characteristics other than the components of the eligibility determination process. For example, they permit examination of the caseload by race of head and number of children but do not permit the analysis of caseload by level of food consumption. SDS and case record abstraction offer more information with which to array Food Stamp units than the IQCS but the additional data

elements are not uniform across states and hence are somewhat cumbersome to use.

General purpose surveys, on the other hand, offer less in terms of countable income and assets and deductible expenses (either less accurate in the case of SIPP or fewer elements in the case of CPS and NFCS) but all of these offer the ability to contrast the Food Stamp population with the rest of the population or even with the rest of the potentially eligible non participant population. Each of the three discussed here has a unique qualification that the other surveys do not. NFCS offers the only data on which Food Stamp units can be arrayed according to their dietary intake and expenditures on food. SIPP permits in depth analysis of persons in households receiving food stamps including those individuals not belonging to the food stamp unit. CPS permits more finely stratified tables of such individuals than SIPP due to a higher sampling ratio but these tables are less accurate than those produced for SIPP because the demographic data do not necessarily pertain to the calendar period during which the unit participated in food stamps.

The cost of preparing descriptive tables of food stamp participants in each of the surveys (except case record abstraction) varies more according to the nature of the tables than to the data set. Most applications using IQCS will be very inexpensive in part because the survey does not support finely stratified tables. Tables for SDS will be equivalent to IQCS on a per record basis unless the additional non uniform data elements are used. In the latter case tables must be tailored to each state thus increasing the development cost.

Simple tables of persons participating in Food Stamps from one wave of SIPP can be done quite easily and cheaply from the public use files. However, most requests for information from SIPP are likely not to be simple and many of them are likely to require data from more than one file. More than one file is required to integrate deductible expenses with program participation as well as to produce annual estimates.

The cost of generating tables of participants from CPS can also be low. In fact, for a comparable table CPS might actually be a cheaper source than SDS because it has fewer observations. However, most of the use of the CPS to produce characteristics tables has been at the culmination of a simulation study which is quite expensive in total and the cost of characteristics tables are not separately measured due to its relative insignificance.

The NCFHS is relatively costly to use when the full wealth of nutritional information is accessed. However, it is only considered when no other suitable data set is available. Therefore in some sense it is cheap, at least relative to the only other alternative which is new data collection.

Income And Asset Eligibility - What If?

In the area of program eligibility, SDS as well as the other administrative data sources is somewhat limited in its usefulness for simulating the impacts of changes in income and asset eligibility regulations. This is due to the lack of representation of the non-participating population thus restricting its use to analyzing the impact of reforms on current caseload. SDS is limited more so than other administrative sources in the ability to address these what if questions because

it is not necessarily representative of the nation. However, SDS does have one advantage over the other administrative sources in this area because it provides sufficient detail to analyze impacts on small groups. Of course, case record abstraction files could be designed to address issues affecting small groups but this is potentially very costly.

In terms of content SDS will provide all of the elements needed to replicate the current eligibility formulas as is true for IQCS and case record abstraction. All these are suitable for addressing "what if" questions which only require knowledge of income, expenses or assets currently itemized in the eligibility regulations. They are limited, however, in supporting analysis that requires the identification of items not currently counted such as vehicles used to produce income.

SDS has the potential to permit increased flexibility over the IQCS in describing characteristics of the population affected by "what if" questions. This is due to the collection of more demographic data on Food Stamp cases by many states. However, these data will not be uniform across states and hence will be somewhat more cumbersome to use in generating reports.

In the area of the cost of a particular "what if" analysis, SDS is expected to be equivalent on a per record basis to IQCS for cross sectional

studies. The total cost of SDS will be higher due to the increased number of observations and the potential to use the additional demographic detail. Case record abstraction is a much more expensive data source than the other two. Both SDS and case record abstraction can provide timely

The surveys most often used to address "what if" questions regarding income and asset eligibility are the larger nationally representative household surveys such as SIPP and the CPS. Due to the representation of potentially eligible but non participating units these surveys are most appropriate when the program reform may induce increased participation in all or particular subgroups of the population.

The CPS and SIPP are generally more expensive to use than SDS and IQCS. However, in the case of the CPS this is due more to the potential for increased complexity of the simulation than to the nature of the data base. The cost of using SIPP to simulate the impact of income and asset eligibility formulas varies depending on the extent to which the available information is used. An extract representing one month of data similar to that which is currently developed from the CPS for quick response analysis is not likely to be any more expensive to use than the CPS extract itself. However, SIPP provides the opportunity to use more data even with a monthly extract and it is likely that analyses will be designed to take advantage of this and therefore be more costly.

In terms of content, both SIPP and CPS are less appropriate than any of the administrative sources for replicating the current income and asset formulas. The CPS lacks deduction and asset information entirely and the income detail covers an annual period when monthly data are more appropriate. SIPP includes asset and deduction data but they are supplemental questions and therefore more expensive and time consuming to access.¹ SIPP also provides monthly income detail although it may differ

¹Note all comments on the content of SIPP refer to the original design rather than proposed alternatives in the scope.

from the amounts reported to the welfare agency and, hence, available from case records. On the other hand, SIPP offers some advantage over administrative data sources in assessing definitional changes in the income and asset eligibility formulas that result in the allowance of a new deduction or the exclusion of all or part of a new unearned source or the addition of a new countable asset. Furthermore, SIPP and CPS offer increased flexibility over administrative sources in describing characteristics of the population affected by program reform due to the increased demographic and economic detail on each observation. Both SIPP and CPS provide information as quickly as IQCS but less quickly than SDS and case record abstraction.

In general SDS will be the preferred data set when the "what if" questions can be addressed with current participants and the issues are centered around small population groups. IQCS will be the preferred data set when the "what if" questions can be addressed with current participants, the results need to represent the national caseload and the results need not be very finely disaggregated. CPS and SIPP will be the preferred choice when other than current participants are needed.

Income And Asset Eligibility - What Happened?

In the area of program eligibility SDS is well suited to the observation and measurement of the impact of a change in eligibility soon after its implementation. Its only major drawback in this regard is the lack of a nationally representative sample. SDS is more timely than all of the other data sources. Furthermore, both SDS and manual case record abstraction provide the opportunity to target an analysis to the period of time most relevant to the implementation of a reform. None of the other

surveys permit this when the reform is implemented at different times across the country.

Like SIPP and manual case record abstraction, SDS has sufficient content to support both pre- and post cross sectional studies of the impacts of reform on caseload composition as well as longitudinal studies of the impacts based on following individual cases. IQCS and CPS only support cross sectional studies. All three administrative data sources contain better information on the components of the eligibility determination process than the publicly funded surveys. For SDS, this fact combined with its longitudinal capability allow studies to more accurately measure the cause of a change in eligibility status and to determine the relationship between the status change and the program change if one exists. As is true for the "what if" questions, SIPP and CPS provide the most demographic and economic data to describe the food stamp caseload. SDS and case record abstraction provide more demographic data than IQCS.

In terms of sample size, SDS permits the study of the impact of eligibility changes on rare groups such as households with income above poverty but does not provide national estimates. IQCS, CPS, and SIPP, on the other hand, can support generation of estimates for the nation and major groups of food stamp cases but their samples are too thin to look at impacts on small segments, particularly with SIPP. Manual case record abstraction is generally geared to the question at hand and can be targeted to particular groups as desired within the budget constraints.

SDS compares favorably with other surveys in terms of the cost of the impact analysis. Cross sectional studies will be more expensive with SDS than the IQCS but only in proportion to the increased number of

observations. Cross sectional studies are likely to be less expensive with SDS than the CPS and SIPP. Longitudinal studies based on SDS will be considerably less expensive than those based on SIPP due to the reduced number of variables available for analysis. Case record abstraction is the most expensive way to observe what happened to the food stamp caseload in response to a change in eligibility determination.

In general SDS will be the preferred data set for analysis of the impact of reforms on the food stamp program because it offers a cost effective way to target the analysis to the most appropriate period. The only data source equally well suited is manual case record abstraction which is very expensive. IQCS and CPS will be useful when the analysis does not need to be so precisely targeted in terms of time frame or when national estimates are required.¹ SIPP provides an alternative when longitudinal data are needed but it is not very timely and may prove expensive to use.

Level And Distribution Of Benefits - What If?

As is true for simulations of the impact of changes in eligibility regulations, the administrative sources including SDS are restricted to studies of changes in the benefit formula and allotment standards that do not increase participation. For studies which do, SIPP and CPS are the principal data sources. SDS and the other administrative surveys provide sufficient content to analyze the effect of reductions in allotments and

¹This comment regarding the IQCS system refers to the collection of data which is typically analyzed in preparing the reports characteristics of participants, rather than data which could potentially be used to analyze the impact of program reform.

benefits. SDS offers somewhat more demographic detail for this purpose but as noted earlier these data elements will not be uniform across states. SIPP and CPS provide the opportunity to simulate the effects of either increasing or decreasing benefits.

Similarly, all types of changes in the benefit formula can be studied using CPS and SIPP but the administrative data sources are restricted to situations where benefits are reduced or are not expected to induce increased participation in the program.

One area in which none of the aforementioned surveys, including SDS, are suitable in terms of content, is the assessment of the impact of changes in benefits on the level and quality of household food consumption. These surveys lack the necessary data on food expenditures and intake. This information is provided on the NFCS. Hence NFCS is data source for the analysis of the nutritional impacts of food stamp program reform.

The comparisons of the sample and the useability of SDS to other surveys for use in simulating benefit formula changes have similar outcomes to the comparisons described previously for simulation of changes in eligibility requirements. Namely, SDS provides increased capability to examine small population groups over the other sources but lacks a nationally representative sample. The cost of using SDS will be more than using IQCS due to a larger number of observations and more data elements but will be less than using the other sources. Finally, SDS is one of the most timely sources of data.

The NFCS provides a representative sample and extensive data on food expenditures and intake. However the survey is becoming out-of-date (the last survey was 1979) and the sample size is relatively small. Also, it is expensive to access the full wealth of information on nutrition.

In short, SDS is the preferred choice when the analysis is focused on the arithmetic affects of food stamp reform and the reform is expected to impact small groups in diverse ways. For nutritional impacts, NFCS is the only alternative until the new Continuing Survey of Food Consumption and Intake is available. For national estimates IQCS, SIPP, and CPS are the preferred choices with IQCS restricted to analysis of the direct effects on current participants. SIPP and CPS are appropriate when the program reform will induce program participation. SIPP will have fewer observations than the CPS but more information to use in simulating the reform.

Level And Distribution Of Benefits - What Happened?

SDS offers an excellent data source to examine the actual changes in the distribution of the food stamp caseload and in program dynamics as a result of the implementation of changes in benefits available to the eligible population. Again, it does not support national estimates but it does permit the analysis to be more targeted to specific groups and highly targeted to the time frame immediately surrounding the change. SDS, like all other surveys except NFCS, does not contain a measure of food consumption so nutritional impacts of changes in benefits cannot be measured.

SDS, SIPP, and manual case record abstraction are the only data sets of the six discussed here which permit true longitudinal analysis of program impacts.¹ For comparable reference periods i.e., a study of what happened over a period of two years or less, SDS will be cheaper to use than SIPP primarily because the limited universe and data elements of SDS will restrict the complexity of its use. Manual case record abstraction is potentially very costly for this purpose although the cost varies with the length of the reference period and the number of observations. SDS and case record abstraction offer an advantage over SIPP in longitudinal studies because they allow the analyst to examine the actual cause of a particular outcome and therefore to distinguish true program impacts from other factors such as income fluctuations. Although technically speaking SIPP could support this as well, Food Stamp participants do not always report the same information on income and assets in a household survey that they do when applying for food stamps. Hence, some noise is introduced when SIPP is used to measure the occurrence and cause of changes in benefits.

There is a trade off among these three sources in their use for longitudinal studies. SDS is less expensive, more accurate but less representative than the other two. Case record abstraction is more expensive, but potentially more representative than SDS. SIPP is more

¹The Michigan Panel Study on Income Dynamics permits longitudinal analysis as well but it is more suited to the topics discussed in the next section.

expensive, less accurate, and more representative than SDS. SIPP allows a more in depth analysis due to the richness of the data, but it is expected to have too few Food Stamp participants to examine small groups.

In cross sectional studies, the comparison of SDS to other surveys in terms of cost and useability has the same outcome as described in preceding sections and hence will not be repeated here.

Interaction With Other Programs

The Food Stamp Program is designed to increase food purchasing power in low income households to improve dietary intake. Many low income households or groups of individuals within these households are also eligible for assistance through other government programs designed to provide monetary support to selected groups of individuals such as needy children, disabled or elderly adults, or unemployed individuals. Therefore, the food stamp program eligibility and benefit determination process has been designed to take into account the income received by household members from these other programs. This results in a situation where changes in benefits received through these other program affects the size and distribution of food stamp caseload in the absence of any changes made directly to the program. This section reviews the potential utility of SDS in analyzing the impacts of changes in other programs, principally AFDC, on the Food Stamp Program.

In terms of content the analysis of program interactions requires at a minimum information on the benefits received from the relevant programs. All surveys provide this data in one form or another. However, SDS and manual case record abstraction are the only two sources which could have potentially sufficient additional information needed to distinguish

program affects such as changes in AFDC guarantees from non-program effects such as a change in the level of deductible expenses. Presumably the IQCS could be designed to provide this information at least for the AFDC program. However, at this time the data available for Food Stamp Program analysis is limited. SIPP permits the observation of changes in other program benefits concurrent with Food Stamp Program benefits but lacks sufficient detail to isolate the precise cause of the changes. The CPS and NFCS are very limited in their ability to examine program interactions.

Interactions between means tested transfer programs can be simulated to analyze the potential effect on the Food Stamp program of a change in the other programs. SDS and manual case record abstraction are appropriate for this purpose when the changes do not increase Food Stamp program participation. However, in the absence of an offsetting change to the FSP, most changes to the AFDC program might induce some increase in FSP participation. For example, a reduction in AFDC benefits, would cause a reduction in countable Food Stamp income resulting in higher food stamp benefits which could induce a marginal increase in Food Stamp participation among the few AFDC cases not receiving Food Stamps. Similarly an increase in AFDC benefits would induce higher participation in AFDC which would likely increase Food Stamp participation.

SIPP is a good data base to support the simulation of program interactions since its universe includes potentially eligible non-participating units and the accounting period is monthly. The CPS is less appropriate because the accounting period is annual.

The evaluation of these surveys on their sample frame and size is similar to previous evaluations. SDS is not representative of the nation

but supports small group estimation. For other surveys the reverse is true. However, there is potential for the universe of SDS and case record abstraction to be expanded in states with integrated systems to include households not participating in Food Stamps which contain individuals participating in other means tested programs. For the study of issues like cashing out Foods Stamps for SSI participants, this would be a valuable addition to the scope of SDS.¹

In terms of useability SDS is very timely for studies of what happened as a result of a change to other government programs and the cheapest to use among the appropriate data sources. Manual case record abstraction is very timely but expensive. SIPP is timely for use in simulating program interactions but is likely to be expensive due to the potential for extremely complex analysis. CPS is less expensive than SIPP but less appropriate for this topic as well. SDS would be the cheapest source for the simulation of program interactions but as noted above, it is likely not to have the appropriate universe.

Effects of Changing Economic Conditions

Since the Food Stamp Program and the other programs with which it interacts are designed to serve the low income population, the size and distribution of the caseload are directly affected by changes in the unemployment and inflation rates. Economic conditions can vary widely within the country, resulting in some areas experiencing severe recessions while other areas are experiencing moderate upswings. This situation results in diverse program impacts across the nation with perhaps no

¹This expansion of the SDS universe is not currently planned.

significant impact on total caseload. If inflation rates vary significantly within the continental U.S., the program may be observed to contain inequities since the income cutoffs are tied to an average overall inflation rate. Periods of changing economic conditions are often coincidental with reforms to the food stamp and/or other programs. Hence, to appropriately analyze program impacts, it is essential to be able to identify the true cause of benefit fluctuations and program turnover.

As indicated earlier, SDS permits the analysis of benefit fluctuations and changes in program dynamics that result from variations in income which are measures of the outcome of changes in the economy. Hence it is the most suitable data set (other than case record abstraction) for the analysis of the effect of changing economic conditions on the program and the separation of program impacts from other exogenous events. One aspect of SDS and case record abstraction which make them more suitable for this analysis than IQCS or the CPS, is their longitudinal nature. Cross-sectional surveys other than time series data do not have sufficient reference periods to capture the delay in the effect of economic swings on program caseload.

Like SDS, SIPP will also permit the analyses of the effect of changing economic conditions on the Food Stamp program but the information reported is potentially less accurate than administrative data thus introducing some noise into the results. However, its sample size may be limiting.

SDS and case record abstraction are not well suited for the simulation of changing economic conditions because they lack detailed employment data and observations of all households potentially affected by

fluctuations in the economy. SIPP is considered to be potentially useful for this simulation. To date the CPS is the only data source used for this type of "what if" question but it has proved to be ineffective.

Analysis of Gainers and Losers

In the analysis of program impacts discussed earlier, there are essentially two types of statistics to analyze, one is the net impact on total caseload and the other is the gross impact. Net impact refers to the change in the total number of participating units and total benefits paid. The gross impact separates units positively affected by a program change from those units negatively affected by the change, i.e. gainers and losers. The first set of statistics reveals the budgetary implications of program reform whereas the latter set describes the distributional implications of a reform.

To analyze gainers and losers it is essential to be able to array program participants according to the comparison of the benefits before and after the reform. This is in contrast to the net effects analysis which does not require comparison of benefits in two time periods at the case record level before tabulation.

To address "what if" questions, most data sources will support the analysis of gainers and losers because the process is to simulate the result rather than observe it. Hence, the "after" benefit can be simulated and recorded for each case along with the "before" benefit. What happened questions about gainers and losers are restricted to longitudinal data sets.

The suitability of SDS compared to other data sets for observing or simulating the program impacts have been discussed at length previously. This section concentrates on the ability to generate gainer/loser statistics once the program reform has been observed or simulated. For "what if" questions that SDS can address i.e. those restricted to current participants, SDS provides the ability to identify gainers and losers, some demographic characteristics with which to array these units and details on the economic characteristics directly counted in determining benefits. IQCS has the same capability of identification but fewer demographic elements. Manual case record abstraction potentially provides the same capabilities as SDS.

SIPP and CPS are richer data sources for the analysis of gainers and losers under proposed program changes because they support a broader class of simulations, allow the identification of gainers and losers and provide a wealth of demographic and economic information on these units. However, the reported variables (particularly on the CPS but possibly true on SIPP) do not capture all aspects of the benefit determination process in contrast to the administrative sources.

For the analysis of gainers and losers after a program reform has been implemented, the SDS and case record abstractions are the most suitable data sets. As noted earlier, this analysis requires longitudinal data thus eliminating all general purpose surveys except SIPP. However, with SIPP the accuracy of the identification of gainers and losers and of the measurement of the extent to which each of these cases was affected by the program reform is reduced.

In terms of sample, while SDS may not be representative, the ability to produce statistically reliable estimates of small groups is particularly important in the analysis of gainers and losers. In terms of cost, SDS is expected to be more expensive than IQCS but less expensive than other sources in the generation of gainer loser tables resulting from simulations. Among the data sets suitable for describing gross impacts resulting from legislated reforms. SDS will be the least expensive. SDS and case record abstraction remain the most timely of the entire set. SIPP will be relatively expensive to use for this analysis and is expected to be restrictive as a result of the relatively small sample of Food Stamp participation.

Summary

This section of the report discussed the suitability of SDS for the analysis of eligibility and benefit formula issues and contrasted that with other surveys considered suitable for these studies. The report identified several studies for which SDS is the preferred data set. The most important benefit of SDS is its cost effectiveness in providing accurate and timely estimates of the impact of legislative changes to the Food Stamp program. SDS is also an excellent choice for the analysis of the interaction of the Food Stamp Program with other government transfer programs as well as with changing economic conditions. In effect SDS can be considered a cost effective alternative when the study might otherwise require the expense of manual case record abstraction. The only time SDS cannot effectively substitute for case record abstraction is when representative national estimates are required.

ANALYSIS OF PROGRAM BEHAVIOR

Due to the limited data sources available for longitudinal studies of the low income population, little is known about Food Stamp entry and exit rates, turnover in the program within a year, and duration of program participation. Also, due to limited data sources, little is known about the determinants of program participation. In recent years studies have been conducted in most of these areas using the 1979 Income Survey Development Program Research Test Panel (ISDP) which was the precursor to SIPP, the Michigan Panel Study on Income Dynamics, and some special nonrepresentative surveys. While these research projects have been valuable and provided useful information in the areas of program dynamics and the determinants of participation, they have contained numerous qualifications as a result of the limitations in the data.

In this section of the report, SDS is evaluated in terms of its usefulness in expanding the knowledge base in these important areas. The data sets with which it is compared are

Michigan Panel Survey on Income Dynamics (PSID)

Manual Case Record Abstraction

SIPP

These surveys are summarized in the appendix.

Table 2 presents an overview of how the four surveys compare as data sets for the analysis of program dynamics and dependency issues. As

TABLE 2: COMPARISON OF STATE SYSTEMS WITH ALTERNATIVE DATA SOURCES FOR ANALYSIS OF BEHAVIORAL CHARACTERISTICS

	STATE SYSTEMS			MICHIGAN PANEL STUDY		
	Content	Sample	Useability	Content	Sample	Useability
A. Turnover in the Program						
1. How many annual recipients	May need to link longitudinally	Excellent size Not representative	Timely, Moderate cost	Data not applicable	Small size, Representative	Not applicable
2. How does turnover vary across characteristics of population	Excellent, but demo. data limited	Excellent size, Not representative	Timely, Moderate cost	Annual longitudinal observations	Small size, Representative	Timely, Low cost
B. Program Entry and Exit Dynamics						
1. What trigger events precede FSP entry	No information	Excellent size, Not representative	Not very useful	Good for long term view	Small size, Representative	Timely, Low cost
2. What trigger event precede FSP exit	Limited information	Excellent size, Not representative	Timely, Moderate cost	Good for long term view	Small size, Representative	Timely, Low cost
C. Welfare Dependency						
1. Can recipients be divided into transitory and long term depend.	Depends in how long the system is supported	Excellent size, Not representative	Timely, Moderate cost	Good for long term view	Small size, Representative	Timely, Low cost
2. How sensitive are recipients to work incentives	Good potential, Work data limited	Excellent size, Not representative	Timely, Moderate cost	Good for long term view	Small size, Representative	Timely, Low cost
3. Duration of receipt	Depends on how long the system is supported	Excellent size, Not representative	Timely, Moderate cost	Good for long term view	Small size, Representative	Timely, Low cost
D. Decision to Participate in FSP*						
1. Determinants of participation	Poor, Little info. on prior status	Excellent size, Not representative	Timely, Moderate cost	Good, but other sources are better	Small size, Representative	Timely, Low cost
2. Variation in participation rate	Very poor, No info. on eligible nonpart.	Excellent size, Not representative	Not very useful	Good, but other sources are better	Small size, Representative	Timely, Low cost

TABLE 2 (Continued)

	CASE RECORD TRANSC.			SIPP		
	Content	Sample	Useability	Content	Sample	Useability
A. Turnover in the Program						
1. How many annual recipients	Good	Needed size, May be representative	Timely, Very high cost	Excellent	Small size, Representative	Timely, Expensive
2. How does turnover vary across characteristics of population	Good, but limited demo. data	Needed size, May be representative	Timely, Very high cost	Excellent	Small size, Representative	Timely, Expensive
B. Program Entry and Exit Dynamics						
1. What trigger events precede FSP entry	Very poor, No data	Needed size, May be representative	Not useful	Excellent	Small size, Representative	Timely, Expensive
2. What trigger event precede FSP exit	Good, Data may lack detail	Needed size, May be representative	Timely, Very high cost	Excellent	Small size, Representative	Timely, Expensive
C. Welfare Dependency						
1. Can recipients be divided into transitory and long term depend.	Poor, Difficult to track long enough	Needed size, May be representative	Timely, Very high cost	Poor	Small size, Representative	Timely, Expensive
2. How sensitive are recipients to work incentives	Good, Work detail may be limited	Needed size, May be representative	Timely, Very high cost	Poor	Small size, Representative	Timely, Expensive
3. Duration of receipt	Poor, difficult to track long enough	Needed size, May be representative	Timely, Very high cost	Good for short spells but spells and truncated		
D. Decision to Participate in FSP *						
1. Determinants of participation	Poor, No info. on prior status	Needed size, May be representative	Not very useful	Excellent	Small size Representative	Timely, cost varies
2. Variation in participation rate	Very poor, No info. on elig. nonpart.	Needed size, May be representative	Not useful	Excellent	Small size Representative	Timely, cost varies

* The CPS is also a potential candidate to study the determinants of participation. The content is good but not excellent principally because it uses an annual accounting period. The sample is good and representative. The cost is moderate.

is true in the preceding section, the surveys are ranked according to three characteristics: content, sample, and useability. Each analysis area is discussed in turn.

Turnover In The Food Stamp Program

Since the Food Stamp Program is administered at the state or local level and eligibility and benefits are determined on a month by month basis, no administrative statistics are maintained at the national level on the total number of households serviced by the program in any given year. Hence, to determine the number of cases participating in a twelve month

household sector. All surveys in Table 2 except the PSID will have sufficient content to provide estimates of this measure of turnover in the program. SDS and case record abstraction provide the most accurate data for the estimates but they are both limited in the demographic information needed to analyze variations in turnover among the participant population. They are also completely inflexible in defining what

SDS and case record abstraction are the only two surveys which can offer sample sizes large enough for many studies of turnover. SIPP is likely to end up with 4,000 or fewer households with food stamps in 1985, the first time the survey is at its maximum in terms of size. SDS is expected to have over 100,000 observations pooled across the state samples. With case record abstraction, of course, the size is determined by the study requirements and budgetary constraints.

The cost of turnover studies will vary significantly across the three relevant surveys. SDS will certainly be the cheapest but it offers no flexibility in terms of defining a household longitudinally. SIPP will be expensive, particularly in the next few years when the research community is learning how to use it. Manual case record abstraction will be more expensive than the other two but the magnitude of the difference in cost depends on the sample size. All three are timely, but SDS and manual case record abstraction can be made available for analysis more quickly than SIPP.

Program Entry And Exit Rates

Program dynamics refer to the reasons why and the rate at which households enter and leave the program. Program exits can be observed and studied on all four surveys but program entries can only be explained using SIPP and PSID. The latter is limited for the study of the causes of program entry because the precise timing of the event cannot be measured nor can trigger events be pinpointed to the few months prior to entry. SIPP is therefore the survey of choice to study what events trigger a program entry. It will have some drawbacks similar to those experienced with the ISDP. However, they are not expected to be as severe. One

example of a SIPP disadvantage is the potential to produce biased program entry rates because the SIPP universe does not include all persons who could enter the program. In particular persons who leave institutions or move into the US from another country after the initial wave are excluded.

The study of program exits can be conducted using the administrative sources in addition to SIPP and PSID because the time period of interest is the time units are on the program. SDS and case record abstraction have the potential to yield artificially high exit rates depending on the case numbering system in the states selected for inclusion. One of the objectives in the state selection process for SDS will be to minimize this bias but it cannot be eliminated entirely. One situation where an artificial exit will be generated for both SDS and case record abstraction is when a participant moves out of a state or the area covered by a centralized system but stays in the program. SIPP will also have some problems when exit rates are measured because some observations drop out of the sample for reasons other than death.

While program exits can be measured relatively accurately with SDS and case record abstraction, the determinants of the exit cannot always be completely measured. Furthermore, the administrative sources provide very little information on cases after they leave the rolls.

The PSID provides good information for a longer term study of program exits, but exits are not pinpointed to a particular month. The survey cannot provide estimates of changes in circumstances immediately preceding a program exit. SIPP on the other hand, does not permit an analysis of exits over a period longer than two years but does permit the exits and the trigger events to be pinpointed more precisely than the

PSID. SIPP and PSID provide more demographic and economic information than the administrative surveys which can be used to determine the cause of a program exit. In particular they provide information on the characteristics of participants after they exit.

The analysis of relative sample size, and useability has essentially the same outcome as the analysis presented previously for the study of turnover in the program and hence will not be repeated here.

Welfare Dependency

Measuring the determinants of duration on the food stamp program is important for the study of poverty as well as for studies or program design. If efforts such as work incentive programs are made to reduce dependence on Food Stamps it is critical to know whether they are effective. If the caseload can be divided into transitory and long term recipients, these efforts can be targeted more effectively to the group with the highest probability of being helped i.e., the transitory group. Similarly, if these two groups can be distinguished, the food stamp program incentives can be more effectively targeted.

To date the only nationally representative data source which permits a study of duration in the program is the PSID. However, its principal use is to examine long periods of attachment to the food stamp program. Furthermore, its sample is very small which restricts the amount of information which can be obtained on the recipient population.

SIPP has one major draw back for the study of duration on the program which is spell truncation. Attempts to identify cases of spell truncation in SIPP are being made (although they are subject to elimination

with the recent round of budget cuts) which will facilitate the use of that survey to study cases with short periods of participation (less than two years). However, the survey is wholly inappropriate to study long term dependency on the program.

SDS has the potential of becoming a useful source for studies of duration and dependency. However, to be effective the system will need to be in operation for a number of years. It is potentially a good source of information to study the effect of programs designed to reduce the need for food stamps to the extent that states incorporate appropriate data elements in their system e.g., participation in the incentive programs. Case record abstraction has similar advantages in the study of incentive programs but is entirely too expensive as an alternative to study long term welfare dependency.

Determinants Of Participation

Studies of the determinants of participation are useful to explain why some units choose to apply for benefits and are important in developing behavioral models of participation needed to accurately simulate the impact of program reform on costs and caseload. The surveys which are potentially useful for the study of participation are the general purpose surveys of households, particularly SIPP, PSID, and CPS. PSID and CPS are good but their lack of monthly income detail results in relatively inaccurate determination of eligibility which is a key component of the participation studies. They also require the imputation of many components of the eligibility and benefit determination process which further affects the accuracy of the eligibility simulation. PSID is further limited in its usefulness because of its small sample size.

SIPP is perceived to be the survey of choice for future studies of participation. Its monthly accounting period and (hopefully) collection of countable assets and deductible expenses offer promise to greatly improve the identification of eligible units and therefore the measurement of participation rates. Its wealth of demographic and economic data will also permit the analysis of variation in participation rates among the eligible population.

Summary

This section of the report identified the utility of SDS for the analysis of program dynamics and dependency issues. The principal conclusions are that it is excellent for studies of turnover, good for a limited analysis of program exit rates and potentially useful to study program dependency. As is true for all studies in which SDS is considered, the results cannot be generalized to the national level.

START UP COSTS ASSOCIATED WITH EACH SURVEY

The preceding sections of this report discussed the advantages of SDS over other surveys according to the relative cost of typical applications and ignores the cost of the development tasks necessary to produce general purpose analysis files. It is pertinent therefore to compare the cost of this development across the surveys. Each is discussed below in terms of its cost relative to the start up costs associated with developing an analysis file from one month of data from IQCS.

IQCS. The initial data base development required to use these data include consistency and range checks, screening bad cases, reconciliation of inconsistent data, new variable construction, and development of sample

weights. The cost of this effort used as a benchmark excludes the cost of FNS involvement in the task which is relevant but includes the cost of generating SAS tables for the reports on characteristics of food stamp households which is not relevant.

CPS. The initial cost to prepare the CPS for generation of descriptive statistics is the cost of purchasing the data (under \$200). The cost of preparing it for use as a simulation data base, however, is considerably more. This preparation typically includes converting the data to MATHtm format, aging it to a target year specified by FNS and simulating the current law tax and means tested transfer programs complete with behavioral models of program participation. This development effort usually costs about two and a half times as much as IQCS exclusive of major efforts to modify the simulation system.

PSID. The longitudinal data in this panel survey is distributed in a format designed to facilitate its use. Therefore, the start up costs are essentially restricted to the effort required to understand the content. This is about one tenth the cost of developing an IQCS analysis file.

NFCS. The initial effort and cost required to use the NFCS depends greatly on the nature and complexity of the analysis. Simple tabulations based on the household level records are very inexpensive. However, if the analysis requires a more complete file development effort including income imputation and the use of the individual food intake data, the cost of dealing with the complex hierarchical file can require several weeks of effort and substantial specialized knowledge.

SIPP. The expected cost to prepare an analysis file from one wave of SIPP is on average about half the cost of IQCS with the first wave being somewhat more expensive than the subsequent waves. This activity essentially involves tailoring the complex array of information to the needs of FNS. The resulting cross section files will be useful for descriptive analysis of food stamp caseload and development of behavioral models of food stamp participation.

The costs to develop longitudinal files from SIPP for the study of program dynamics is not known at this time. This cost is highly dependent on what longitudinal products the Census Bureau distributes and how suitable they are for FNS research. Even less is known about the potential cost of developing SIPP as a microsimulation model data base since the models themselves have not been designed.

SDS. Since this system is not yet in place, the development costs are divided into two components. The first component consists of the design phase, the establishment of agreements with the states to participate and the initial acquisition of data and the development of the first set of analysis files. The second component consists of the expected average cost of obtaining an additional round of data and producing analysis files. The latter includes merging each new wave of data to data in previous submissions to produce longitudinal files as well as producing a set of cross sectional files. The estimate cost of the first component is about two and a quarter times expensive than IQCS and the cost of the second is about one and a half times as expensive as IQCS.

V. KEY CONSIDERATIONS IN DEVELOPING SDS

The evaluation of the suitability of SDS for program analysis incorporated a number of assumptions about its design. This chapter clarifies those assumptions and discusses the features of the system necessary to ensure that SDS has the capabilities discussed earlier.

As discussed in Chapter II and III, two principal uses of the SDS are in areas for which existing data are insufficient, namely, small group estimation and examination of caseload dynamics. Hence the system must incorporate the necessary features to permit these uses. Sample sizes should be large enough to support statistically reliable estimates of subgroups of less than 10% of the total population. The sample selection process must permit linking of case records across time with minimal problems of attrition. Finally, the frequency of the data collection must be high enough to identify cases closed and the reason for the case closing.

Other areas which are not so critical but should be taken into consideration when designing the system are the length of time the state maintains an active record subsequent to the closing of a case, the number and types of variables in the system beyond those needed to determine eligibility and benefits, and the ability to select non food stamp cases participating in other means tested programs.

Frequency of Data Collection

The frequency of data collection needed to insure complete representation of case closings varies depending on the length of time the states maintain a record for a case after it becomes inactive. If a state

waits at least six months before purging records of inactive cases, then collecting data from the states every six months minimizes the collection cost while insuring complete representation.

The frequency of data collection needed to insure the identification of and reasons for a case closing varies depending on the nature of the information retained in the record of an inactive case. Suppose a state keeps inactive records for six months but does not record sufficient details to determine the reason for the closing. In that situation, collection every six months insures all cases are represented but incomplete records exist for cases closed between two collection cycles. This would severely restrict studies of the impact of program reform because it would not be possible to determine whether or not the case closing was a direct result of the program reform. On the other hand, if the state maintains sufficient detail on the case closing, collection every six months would not jeopardize the ability to determine the impact of the program reform.

State Case Numbering System

In order to conduct studies of caseload dynamics and to measure changes in benefits over time at the case level it is necessary to construct longitudinal files of food stamp cases. This is particularly true of states that do not have extensive case histories (which many states do not) or do not maintain much data on inactive cases. The simplest

method of establishing longitudinal files is to match data from successive collections using the assigned case number or a scrambled version of the original case number. The only other alternative is to match cases across time using lists of members which is cumbersome and expensive.¹ Using the case number as an identifier, however, assumes that the state does not have separate processing systems for each welfare program. For example, if a unit receives both Food Stamps and AFDC in one month but only Food Stamps in the next month, it is essential that the state identify this as the same unit in the absence of any other changes such as marital formation or dissolution. Another example of the artificial generation of caseload turnover is when temporary case numbers are assigned upon receipt of an initial application and later replaced with a permanent identifier. For example, some states use a case numbering system that is a function of the social security number of the authorized recipient. When an applicant does not already have a social security number, temporary case numbers are assigned until one can be obtained. In order to successfully construct longitudinal records for applicants in these states it is essential to be able to link the permanent case number with the temporary case number. Often this can be easily accomplished by the retention of the temporary identification on the case record.

It is useful to note at this point that developing longitudinal units based on constant case numbers can adversely affect the study of

¹This option is not available in all states.

program dynamics. The issue of what constitutes a longitudinal unit over time has recently been subject to debate in light of the need to construct longitudinal families and households on surveys such as SIPP. In the course of that debate it has become clear that there is no one rule simple or otherwise which is universally appropriate for all longitudinal studies. Furthermore, it is clear that the choice of a particular definition affects measures of program dynamics. As an illustration consider a two parent household with children in January who subsequently experience a divorce generating two households in February. Depending on the nature of the longitudinal unit definition, there could either be two or three different units in existence in the two month period. Furthermore, not all definitions which result in the count of two describe the same two units. One definition is to say that two units observed at different points in time are the same if the head (in this example, the husband) stays the same. (This definition is quite similar to defining a food stamp unit based on the continuity of the authorized recipient). If the children stayed with the wife after the divorce then one household is the father which existed two months but shrank from size four to one. The second household existed only one month and consisted of the mother and two children. For obvious reasons, this definition is not preferred for longitudinal analysis. Another definition, which similarly counts two households in the two month period, is based on where the majority of the members go. In this case one household would be the father only and it would just exist in the second month. The other household would exist two months and shrinks from size four to three.

This discussion of the longitudinal unit definition is not intended to be comprehensive so all possible treatments of all possible household transformation will not be discussed. The simplistic description, however, is intended to indicate how limiting the case number can be in defining a continuous food stamp unit.

Sampling

The SDS as it is currently envisioned includes the collection of complete case records from selected state systems on a regular schedule when constructing both cross-sectional and longitudinal data sets. It is feasible, therefore, to consider performing analysis on 100% of all the cases in each state. However, in most states it is not necessary to process 100% of the cases in order to produce statistically reliable estimates of even relatively small populations. (The states in which close to 100% of the caseload would need to be processed are the states with small caseloads). Hence, one cost saving feature of SDS could be the incorporation of sampling techniques to reduce the size of the analysis files ultimately used for research.

Two important aspects of sampling are the technique used to draw the sample and the number of units selected. There is a tradeoff between these two aspects in the achievement of the goal of reliable estimation of rare events on small subgroups. The issue of technique will be addressed first.

Sampling Method. There are numerous strategies which can be used to draw a statistically valid sample of cases from complete state files. One is a simple random sample where cases are literally randomly selected from the complete set. Traditionally, the precision of all other sampling

techniques (i.e., the variance of the estimate around its true value), is discussed relative to this method.

A second general process is stratified random sampling where the units are arrayed according to one or more characteristics and then within each cell a simple random sample is drawn. Stratified sampling is an alternative often chosen to reduce sample size over simple random sampling while keeping the same standard error of certain selected estimates. However, the degree to which this precision is retained (or even prevented from decreasing) with smaller sample sizes depends heavily on the manner in which the data are arranged into cells and how the stratifiers in that arrangement are correlated with the statistics being estimated.

One similar technique which effectively arrays the data into cells prior to selection is systematic sampling. In this case one unit of the first k units on the file is chosen and then every k th unit thereafter is selected. The first unit can either be randomly chosen from the first set or specified in advance of sample selection (at say the mid point). This method differs from stratified random sampling because every observation selected is in the same relative position within a cell (which is determined by the sort sequence of the case records rather than by explicitly arraying the data) rather than distributed randomly within each cell. This method is relatively easy to implement and can yield more precise estimates than simple random sampling. However, there are conditions under which systematic sampling yields higher standard errors of the estimates. This occurs when the data are sorted in such a way as to yield a sample where there is less variation among the observations than exists in the universe.

Other more sophisticated techniques exist for sampling but will not be addressed here. These techniques are needed in situations where the cost of sampling is a main factor. For SDS, the marginal cost of sampling one observation once the system is established is quite small.

In the selection of the appropriate technique it is important to consider whether using a process other than simple random sampling provides a significant increase in precision for the same cost or a significant decrease in cost for the same level of precision. Since the marginal cost of increasing the sample size in SDS is minimal, the latter is not likely to occur. The former is difficult to explicitly address within the scope of SDS because the objective is to produce general purpose analysis files from which a variety of statistics are generated.

Another issue to consider in designing the sampling technique for SDS is the need to construct longitudinal case records. Hence it is essential to ensure that once a case is selected for inclusion in the sample in one round of data collection, it continues to be selected in all subsequent rounds of data collection until it is dropped from the state's administrative records. In contrast to this, however, the sampling method must give a non zero probability of selection to all new entrants to the food stamp program in the round of data collection immediately following the entrance. Furthermore, once a case has been eligible for sampling either in the initial data collection or as a new entrant, then for all subsequent data collections the probability of selection must be zero. This latter condition prevents the creation of truncated case histories for selected observations other than those sampled in the initial collection.

There are essentially two ways in which the goal of producing longitudinal case records can be achieved. One is to design a simple random sample which selects cases as a function of the last digit (or digits) of a random number which is continuously associated with the unit. The last four digits of a social security number exemplify this type of random number. These four digits are assigned sequentially to individuals applying for social security numbers within certain areas. Another example is a case number with a component which is assigned sequentially to units as they apply for Food Stamps.

The second approach would be to select a sample during the initial data collection using any valid procedure. In subsequent submissions, new cases would be separated from existing cases, and part of the sample would be extracted from the old cases through matching of identifiers with the previous submission. The remainder of the sample would be selected from the new cases using an appropriate sampling technique.

The first approach is clearly easier and cheaper to implement than the second since it involves fewer steps. Combined with the fact that procedures other than simple random sampling do not offer significant advantages in terms of cost or precision, the first approach is recommended for SDS.

Sample Size.

The sample size determination for SDS is based on the assumption that the sample should be sufficiently large to permit detection of small impacts of a program reform in the total Food Stamp Caseload or reliable estimates of the impact of program reform on small groups of the population. Hence the objective can be described as controlling the

probability of incorrectly rejecting a true hypothesis because of small sample sizes. The analysis files to be produced from SDS will be used for such a broad number of topics that the complete set of hypothesis tests cannot be explicitly taken into account in determining sample sizes. Instead, an approximation is made based on one hypothesis that is in some sense typical of the analysis of small subsets of food stamp cases. The hypothesis is

Is there a significant impact on elderly households if the current shelter deduction provision is replaced with a more generalized standard deduction?

In particular this analysis would be focused in the small porportion of Food Stamp households containing an elderly individual where the unit's shelter deduction exceeds the current cap. This study is typical of the expected cross sectional applications for which SDS will be used in that it measures the impact of a change in the program and it requires a sufficiently large sample size to permit making statistically valid inferences for groups less than 10% of the population in each state. The impact would be measured in terms of the change in food stamp benefits before and after program reform. Note that this could be before and after actual implementation of the reform or before and after the simulation of the reform.

There are six factors to consider in assessing the approximate sample size needed to test the hypothesis described above:

- o Confidence level of the test. For this purpose a 95% confidence level on a one tailed test and a 90% confidence level on a two tailed test are assumed.

- o Power of the test. This refers to the described probability of detecting statistically significant impacts when they exist. A 90 percent power level is assumed.
- o Size of the detectable impacts. The size of the impacts that are expected affects the sample size in that smaller impacts can be detected with larger samples.
- o Variance of the outcome measures. The larger the variance of the outcome measure, the more difficult it becomes to attribute observed (or simulated) differences in average benefits to the program reform rather than to chance sample variance. The variance measures used in the sample size determination below were calculated from the August 1982 IQCS survey.
- o Level of Disaggregation in the Analysis. Because one of the objectives of SDS is to permit finely stratified tables, the overall sample size must be large. In order to ensure sufficient observations size calculations are first made for the smallest subgroups of interest (i.e., elderly households with deductions exceeding the cap). The overall sample size is computed by multiplying by the ratio of total food stamp cases to the number of cases in that subgroup as observed in the August 1982 IQCS file.

Normally in calculating sample sizes the relative cost of data collection is considered as an important factor along with the above because there is a trade off between the confidence to be placed in the results and the additional costs of increased sample sizes. However in the case of SDS, the marginal cost of adding extra observations is extremely low (estimated to be about 50 cents for adding one observation to the cross sectional files). Hence, it is not factored in the calculations below.

It is estimated that a sample size of 25000 from each state is sufficient to detect impacts of about 10% in the average benefits of elderly households whose shelter costs exceeds the current cap. This

assessment is based on the following formula:

$$\delta = \sigma_T (t + \phi^{-1}(\gamma)) \sqrt{2/n}$$

where:

δ = Minimum detectable difference in average benefits

σ_T = standard deviation of the average benefit (estimated to be 61.1 for average food stamp benefits for elderly households whose shelter deduction exceeds the current cap)

t = 90% significance level for two tailed test (1.645)

γ = Desired power of the test, i.e., .9 ($\phi^{-1}(.9) = 1.282$)

n = Number of elderly households with deductions exceeding the cap in a state given a random sample of the total state caseload of size 20000 (estimated to be 800 since the August 1982 IQCS file should that 4% for the national caseload falls into this category)

This calculation results in a value of the minimum deductible difference to be 8.90 which is approximately 11% of the reported average benefit in August 1982 for this group (\$78).

The above formula can be restructured to determine the sample size needed to detect a impact of 10% in the average benefit (7.82). This calculation results in a minimum sample of 1043 of the households of interest on an over all sample size of 26067.

APPENDIX

This appendix provides an overview of five different surveys referenced in the body of the report. The surveys are:

Current Population Survey
Integrated Quality Control System
Michigan Panel Study on Income Dynamics
National Food Consumption Survey
Survey of Income and Program Participation

CURRENT POPULATION SURVEY

The Current Population Survey (CPS) is a monthly household sample survey of labor force activity and economic status.

Nature of the Survey

The CPS is administered by the Bureau of the Census to about 60000 households representing the civilian non-institutionalized in all 50 states and the District of Columbia. Sample households are interviewed for four consecutive months, are skipped for eight months, and are finally included in the sample for the corresponding four months one year after the first set of interviews.

The original, and still central, focus of the CPS is on estimating the size and characteristics of the U.S. labor force. The survey, in addition to measuring employment and unemployment, now provides estimates of characteristics of the population as a whole. In addition to the regular monthly survey questions, the March survey, known as the Annual Demographic File, collects additional information on income, work experience, and migration. These questions, covering income and transfer receipt (including food stamps), work experience, and migration, refer to the previous calendar year.

Time of Survey and Intervals Between Samples. The survey is conducted monthly, during the week containing the 19th day of the month. The reference period for the standard monthly survey questions on labor force activity is the preceding week--the week containing the 12th of the month.

Labor force information from the regular monthly surveys is available from the Census Bureau and Labor Department relatively quickly following the survey. The unemployment rate is typically announced at the beginning of the next month, and published reports are available shortly thereafter. Supplemental information from the March CPS is not available as quickly, however. Preliminary reports, such as those on poverty are released about four months after the survey. Public use microdata computer tapes of the Annual Demographic Supplement are available from six to ten months after the March survey is conducted.

Data Elements. The regular monthly CPS survey provides information for each household on household size, composition, relationships, and type of living quarters. For each adult, labor force activity information is obtained, including type of work, hours worked, reasons for not working, and a variety of related information. In the Annual Demographic Supplement (the March CPS), a much more extensive list of questions is asked. Table A.1 presents an abbreviated list of the person, family, and household-level variables available on the March CPS. Of particular interest in the context of this report are the food stamp receipt variables, together with income items and demographic and socioeconomic characteristics.

Potential Utility of the Data Base

The March CPS is often used to compare both participant and non-participant populations before and after the implementation of legislative changes and to simulate the affect of legislative changes.

TABLE A.1

Information Contained in Current Population Survey

Regular CPS--Labor Forces Status in Reference Week	March Supplement--Economic situation in Previous Calendar Year
Age	Tenure (own/rent)
Sex	School lunch
Race	Public Housing
Family status	Food Stamps
Marital status	Work Experience
Veteran status	Unemployment Weeks
Highest grade	Reasons for working
Relationship	Weeks worked
Activity in reference week	Layoff
Work	Number of jobs
Hours of work	Usual hours
Time lost from work	Part-time work
Overtime work	Longest job (type)
Reasons for part-time work	Class of worker
Layoff	Earnings and self-employment income
Other reasons for missing work	Unemployment compensation
Paid time off	Worker's compensation
Full time work	Migration
Industry	Social Security
Occupation	SSI
Unemployment	Othe survivors benefits
Job search	Disability status
Reasons for unemployment	Disability income
Weeks of unemployment	Pension/retirement income
Last worked	Income from financial assets
Type of work sought	Education
Reasons for not seeking work	Educational Assistance
Usual hours worked	Public Assistance (AFDC, GA)
Hourly or weekly wage	Veteran's payments
	Child support
	Alimony
	Other financial assistance
	Pension/retirement coverage
	Health insurance
	Medicare
	Military health care

Analysis of Cross Section Before and After Legislative Changes.

With regard to the before and after analysis, the March CPS offers several distinct advantages. The CPS makes available a wealth of current information including receipt of food stamps, participation in other transfer programs, income and demographic characteristics. It is thus a potential source for comparing the characteristics of the food stamp and other low income populations, before and after program changes. However, for the variables of interest to FNS (income, food stamps, other transfers), the Demographic Supplement provides calendar year information that will not coincide exactly with the timing of implementation of legislative changes.

The CPS has information on nonparticipants as well as participants, unlike the QC data or other administrative data. It has a large sample size and contains a wealth of non-food-stamp information that is useful in assessing the separate effects of legislative change and macroeconomic conditions. It also has relative limitations, however.

The CPS has less detail on programmatic aspects of participants. For example, it lacks information on work registration requirements and deductions. The information related to food stamp eligibility is also somewhat less detailed on the CPS than on the SIPP (although these items are routinely imputed by MPR on CPS-based MATH files). The CPS provides successive cross sections rather than longitudinal observations. Finally, the food stamp information on the CPS is subject to underreporting, although supplemental data are available with which to assess the magnitude and any resulting bias from such underreporting.

Simulation Analysis. As noted above, the second way in which the March CPS is potentially valuable for program analysis is its potential as a data base for simulating the effects of program change. This potential is best exemplified by the extensive use of CPS-based microsimulation analyses of actual and proposed program changes over the last 10 years. Many of the strengths of the CPS that are described above (e.g., the richness of detail on income and demographic characteristics) are applicable here. In particular, the level of detail of sociodemographic data on households is attractive in that it facilitates the modeling of behavioral response (e.g., altered labor supply behavior) as well as the arithmetic effects of program changes.

INTEGRATED QUALITY CONTROL SYSTEM

The basic purpose of the Integrated Quality Control System (IQCS) is to estimate the amount of food stamp and AFDC benefits issued in error on a state by state basis. As a byproduct, the IQCS provides a source of detailed data on the characteristics of participants in the Food Stamp Program.

Nature of the Data Base

The Integrated Quality Control System is based on reviews of a sample of 150 to 1,200 cases in each state over a six month period on a continuing basis. With the appropriate sample weights, cross section analysis extracts can be developed from these case reviews for a single month or the full semiannual state sample that are representative of each state and of the national caseload. The full sample consists of approximately 45,000 cases.

For the past several years, analysis files have been created for one or two months of each year. Each file is a nationally representative cross-section survey of food stamp households. These analysis files include information on demographic characteristics, sources and amounts of income, assets, deductions, and other program-specific variables.

The IQCS data provide detailed recurring cross-sectional information on the characteristics of the actual food stamp caseload. Other data sources generally provide far less detailed information on food stamp participants. For example, data reported in the Food Stamp Program Statistical Summary of Operations provide only a very restricted set of variables on total households and persons participating and total benefits

paid. The Current Population Survey does not provide the information on shelter, medical, and dependent care expenses necessary to determine food stamp deduction amounts; nor does it provide the information on household assets necessary to determine eligibility for food stamp benefits.

Information contained in the QC analysis files is drawn from the IQCS database, extracting all cases for which QC reviews were completed in a particular month. Although the integrated database includes AFDC and Medicaid reviews as well, only food stamp recipients records are routinely extracted for these FNS analysis files. These files contain a sample of approximately 7000 food stamp recipient households in all 50 states and the District of Columbia. A sample of cases from outlying areas is also available. The review sheet provides detailed demographic and socioeconomic data, including:--

- o Ethnic classification of the household: black, white, Hispanic, American Indian, other
- o Special status classification of the household: alien, migrant, striker
- o Whether expedited service was received
- o Size of the household certified for program participation
- o Household member characteristics: age, sex, relationship to head of household, employment/work registration status, and sources of income
- o Details on income sources: earned income, OASDI/pensions/other such benefits, receipt of AFDC or general assistance, and receipt of SSI
- o Detail on assets: amount of liquid resources, value of real property, value of nonexcluded vehicles, and other nonliquid resources

- o Amounts allowed as deductions: earned income deductions, dependent care expense, shelter expense, and medical expenses

In addition, data on the amount of the coupon allotment received, the status of the case after review, and the amount of coupon allotment issued in error (if any) are available.

The Utility of the QC Data

There are at least five different ways in which the IQCS data could be used to study the impact of recent changes in the Food Stamp program:

- o Descriptive Tabular Analysis of Participant Impacts. Tabulation of QC data collected before and after a legislated change in the program would provide useful descriptive summaries of the characteristics of the food stamp recipient population in the two periods and how those characteristics have changed. The approximate impact of the reduction in the earned income deduction could be illustrated with tables showing the average earned income deduction and benefit amount across categories of earned income. These tables could be disaggregated by type of household--elderly/nonelderly, with children/without children, etc.--to determine which households were most affected by the change. This application of the IQCS data could provide important descriptive statistics, but could be of limited value in identifying and quantifying the causal factors behind changes in the economic status of food stamp recipients.
- o Descriptive Multivariate Analysis of Participant Impacts. The cross-section surveys could be pooled and state level variables added for each time period, reflecting the external environment (such as the unemployment rate and per capita personal income). Multivariate models could then be estimated to describe the changes in recipient characteristics, controlling for factors such as the economy and time trends. These descriptive analyses are likely to be very important in providing information on how participant characteristics such as average food stamp benefits have changed, but they cannot address many behavioral changes as only participants are observed.

- o Microsimulation of Participant Impacts. Under contract to FNS, MPR has developed a small-scale microsimulation model that operates directly on a QC data base. It has provided FNS with low-cost, fast turnaround estimates of the effects of proposed changes in food stamp regulations. Because it operates on a sample of food stamp participant households, it cannot be used to simulate the effects of a relaxation in eligibility requirements. One limitation of this analysis technique is its neglect of behavioral responses to changes in food stamp regulations. For example, in simulating the effects of more stringent income eligibility requirements, this model would not make adjustments for reductions in market labor that might occur as households seek to remain eligible for food stamp benefits. On the positive side, this technique does identify and quantify specific causal factors underlying changes in program costs and changes in the characteristics of the food stamp caseload.

- o Other Analyses. MPR has developed a single equation model that disaggregates changes over time in the average per capita food stamp benefit into dollar amounts attributable to changes in six factors (MPR, May 1982, and December 1982). It can operate on aggregate QC data from any two distinct months. Some disaggregation by demographic group or the availability of specific deductions or types of income is also possible. An important limitation of this analysis technique is its inability to account for changes in food stamp eligibility and participation.

In summary, the QC data are an excellent source of information on participants before and after legislated change in the program and, hence, as a measure of the net change that occurred. They also provide a good source of data on participants at intermediate points. They constitute a good data base for simulation analyses of the impact on recipient households of selected changes, in the eligibility and benefit formulas such as the gross income limit, under the assumption that the impacts on eligibility and benefit amounts do not induce participation or other behavioral change.

However, the QC data do have two serious limitations. First, the gross before and after picture is too broad--including all the environmental factors which confound the analysis. Although multivariate analysis making use of variation in economic conditions across states, variation in what was implemented, and time trends, may be able to sort out some of these factors, the likelihood is that they cannot be adequately disentangled without longitudinal data that allow the analyst to follow all the transitions. Second, simulation of program changes are incomplete because they do not allow for households to respond by changing work effort or participation behavior.

MICHIGAN PANEL STUDY ON INCOME DYNAMICS

The Michigan Panel Study on Income Dynamics (PSID) is another data base that may be of use for the analysis of food stamp behavior.

Nature of the Data Base

The PSID is a longitudinal data base of a sample of the entire U.S. population that was begun in 1968. In that year, a sample of 5,000 families was selected and interviewed by the staff of the Survey Research Center at the Institute for Social Research of the University of Michigan. The individuals were reinterviewed the following year and each subsequent year thereafter. The purpose of the PSID was to provide general information on the income patterns of U.S. families over time, with a focus on the low-income population. To focus on this group, a structured sample was drawn with heavy overrepresentation of low income households. The initial sample combined a random sample of the U.S. population with a sample from the 1967 Survey of Economic Opportunity, which oversampled poor families. Initial funding for the project was provided by the Office of Economic Opportunity and subsequent funding by HEW-HHS. Questions were asked each year on the income, employment, and transfer benefits of the individuals in the family. Questions relating to food stamps were also asked. In the early few years of the survey these were rather crude, but since 1975 a generally stable set of food stamp questions has been asked. Information includes the amount received in the survey period; and, for more recent surveys, the amount received in the month prior to the survey.

Potential Utility of the Data Base

Like most of the data sets discussed in this report, the PSID has both advantages and disadvantages relative to other data sets. Its chief advantages are: (1) it is a longitudinal panel; (2) it covers a relatively long period of time; and (3) it covers the entire population. Its disadvantages are: (1) it is only annual; (2) the Census-type questions regarding income received in the prior year are quite crude relative to either SIPP or food stamp administrative data; (3) the benefit and participation data are subject to survey response error; and (4) sample sizes of recipients are not as large as in administrative data.

In light of all of these considerations, the PSID is considered potentially useful for a longitudinal analysis of the long term effects of the program on participant households.

Longitudinal Analyses of Program Impacts. The panel nature of the data enable the analyst to go beyond simple before and after analysis and track individual households longitudinally. The panel nature of the data and their length cannot be duplicated with any other existing survey in the country. This makes the PSID of potential use in studying macroeconomic effects as well, for the time period covers several complete business cycles. By contrast, although the CPS covers long periods of time, in the CPS sampling scheme only waves close together in time are of a panel nature (i.e., contain at least some of the same households); and panel data sets such as SIPP do not cover very long periods.

Behavioral Models. The PSID can be (and, indeed, has been) used to estimate multivariate models that predict the probability that a household will receive food stamps. Only with a sample of the total population can complete food stamp participation models be estimated. The main disadvantage of all food stamp administrative data is that they only contain data on recipients, and only for the periods of participation. Thus, information on nonparticipating eligibles is not available. Nor is information available on what former recipients are doing after they leave the rolls or what present recipients were doing before they came onto the rolls--which is necessary to estimate complete turnover models. Again, however, whereas other data sets such as the SIPP and the CPS also have data on nonrecipients, they lack the long period and panel nature of the PSID.

The disadvantage of annual data such as the PSID is that detailed intrayear turnover analyses cannot be done. Essentially, one only has "snapshots" of data every 12 months. The retrospective nature of some of the food stamp questions likewise do not provide information on intrayear turnover, a disadvantage for programs with monthly accounting periods. But the newly added questions on "last month's" food stamp benefit give a figure that can be linked up with a specific calendar month--although again, it is only a snapshot. In addition, the sample sizes on the PSID are not large because it is a sample of the entire population. Its overrepresentation of low-income families however, makes the subsamples of the food stamp recipients somewhat larger than would otherwise be the case. Participation rates among unweighted sample households for food stamps in the PSID have ranged from 8 percent to 10 percent so that, other

things being equal, with a 5,000-family sample one would expect 400-500 food stamp recipients. This is a maximum estimate, however, because there has been attrition from the panel, which has been offset to some extent (but not completely) by the new families created when families split up or when subgroups of individuals split off to form their own families (these new families are continued in the survey). In any case, the sample size does appear large enough to perform simple forms of multivariate analysis of food stamp participation and turnover.

THE NATIONWIDE FOOD CONSUMPTION SURVEY

The Nationwide Food Consumption Survey (NFCS) is part of a continuing series of surveys on food consumption.

Nature of the Data

The 77-78 NFCS consists of seven different surveys each with two parts - household food use and individual intake (Rizek, 1978). These surveys are

- o **The 1977-78 National Food Consumption Survey (NFCS).** The basic nationwide survey is a selfweighting area probability sample-a representative sample of about 15,000 households in the 48 conterminous States and the District of Columbia. Interviews were completed throughout a year's time-April 1, 1977 through March 31, 1978. In each of the four quarters about 3,750 interviews were conducted.

Individual intake data were collected for all family members in the April-June quarter. In the other three quarters, all members of the household 18 years and under were interviewed, but only half of those 19 years old and older were interviewed.
- o **The "bridging" survey.** During April-June 1977, about 1,500 households were surveyed by the 1965-66 survey procedures. This will permit evaluations of differences between results from the 1965-66 and 1977-78 surveys that are associated with changes in methodology.
- o **Alaska.** Data were obtained from about 1,200 households and all members (urban only) during a 3-month period (January-March 1978).
- o **Hawaii.** Data were obtained from about 1,200 households and all members statewide during January-March 1978.
- o **Puerto Rico.** Data were obtained from about 3,000 households territorywide and all members during a 6-month period (July-December 1977).

- o **Supplemental survey of the elderly.** Household data were collected from about 5,000 households in the 48 conterminous States and the District of Columbia, with one or more members 65 years or older. Individual intake data were obtained through 24-hour recalls from all household members.
- o **Supplemental survey of low-income households.** Data collected from about 5,000 low-income households during the period November 1977 through March 1978. Low-income households were defined as either receiving food stamps or eligible to receive food stamps.

The Survey of Food Consumption in Low Income Households 1979-80 (SFC-LI) was very similar to the supplemental survey of low-income households conducted in 1977-78 as mentioned above. This survey, designed to measure the effect of the elimination of the purchase requirement and other Food Stamp Program changes implemented in early 1979, consists of a sample of about 3,000 households eligible for food stamps in the contiguous United States.

Information Contained in Food Consumption Survey

These surveys provide detailed information on household food use. Household food use refers to food and beverages used from household food supplies during the seven days preceding the survey interview. Food used includes food and beverages consumed at home, carried from the home, discarded, or fed to pets. Food purchased with cash, credit, or food stamps and food that was home-produced, received as a gift or payment for work, or received through other programs are all included in the measure of household food use. Data were also collected on the number and type of meals eaten from household food supplies by household members and others, on the snacks and refreshments eaten by guests, and on meals eaten away from home by household members. In addition to the data on food use,

information was obtained on household characteristics presumed to be related to food use and dietary quality, such as participation in the FSP or NAP, participation in other food assistance programs (School Lunch, School Breakfast, WIC, or programs for the elderly), household composition, income, education and employment of the household heads, urbanization, tenancy, and food-buying practices.

Data on household food energy (calories) and nutrient availability are also calculated from the quantities of each food item used by the household. Calories and 14 different nutritive values for each food item are calculated from tables of the nutritive value of foods. Total household caloric availability is derived by summing the calories of the individual food items, and, similarly, the household availability of the 14 nutrients is obtained by summing the nutritive values of the individual food items. Nutritive values pertain to the edible portion of the food used from household food supplies, with adjustments for losses during preparation.

Potential Uses of the Data

NFCS is used by USDA for a number of different activities. Those which pertain most directly to analysis of the Food Stamp Program are

- o Provide measures of the adequacy of food intake of Food Stamp Participating and nonparticipating households.
- o Update the Thrifty Food Plan.
- o Describe the food consumption patterns of the U.S. population.

NFCS is the data set of choice for analyzing the impact of the Food Stamp Program and other nutrition programs on food expenditures and dietary intake. In fact, the continuing series of food consumption surveys of which NFCS is a part, is the primary source available to measure this impact.

NFCS has the appropriate universe to support the analysis of a broader set of questions, related to the Food Stamp Program since it includes potentially eligible participating units. However, its use in this area is limited since the most recent survey was fielded in 1977-80.

SURVEY OF INCOME AND PROGRAM PARTICIPATION

The Survey of Income and Program Participation (SIPP) is a new survey which has great potential for the analysis of Food Stamp Policy. The summary which follows is based on its original design which is currently being reevaluated with the objective of reducing response burden and cost.

Nature of the Data Base

SIPP is a longitudinal survey designed to follow individuals over a 2 and 1/2 year period. The initial sample (1984 panel) consists of 26000 housing units selected to represent the noninstitutionalized population of the United States of which about 21000 were occupied and eligible for interview when the survey was initiated. A new panel of slightly smaller size is introduced each year increasing the combined sample size to about 35000. The adult individuals in each sampled household are interviewed initially and then reinterviewed every 4 months for a period of 2 1/2 years. The reference period for each round of interviewing is the four months preceding the interviewing date. Each sample panel is divided into four equal size subsamples or rotation groups which are interviewed on a staggered schedule. One wave (or complete round of interviewing) therefore takes four consecutive months to complete.

Along with the adult individuals in the initial sample for each panel, all other individuals with whom they reside at the time of the interview (including children) are included in the sample. These individuals remain in the sample for as long as they reside with the adult individuals interviewed initially in Wave 1. There are some instances when

the original sample members are not reinterviewed in one or more waves
~~after the first round of interviews. This occurs when an individual moves~~

out of the country or to Alaska, enters an institution or military barracks,
moves too far from a sample area to make interviewing cost effective,
refuses to participate, or dies.

Data Elements. The survey instruments are modular in design
consisting of a control card, three core modules and a series of topical
modules. The control card and three core modules are administered in each
of the waves and cover the following topics:

- o Household and family composition and characteristics
- o Characteristics of adults in the household, labor force activity and income reciprocity for the four months prior to the interview date
- o Earnings for the four months prior to the interview
- o Unearned income for 50 different sources for the four months prior to the interview
- o Receipt of inkind benefits such as Food Stamps, participation in health insurance, and asset ownership
- o Characteristics of children.

The series of topical modules is currently being refined. However, it is anticipated that the following topics of interest to FNS will be covered.

Health and Disability
Work History
Assets and liabilities
Housing costs
Child care arrangements and financing
Duration of welfare
Work related expenses
Annual Income and taxes
Energy usage

Potential Uses of the Data

The SIPP data are expected to be a superior cross-section and longitudinal information source for examining a wide variety of policy issues related to the economic situation of families and individuals and the impact of government transfer programs.

The cross section aspects of SIPP allows the same types of applications that FNS and others have been doing based on the CPS, but with richer and better designed data sets, including questions like the following:

- 1) What combination of benefits from government programs are received by income class and demographic group? With over 50 sources of cash and noncash income identified in SIPP, this question can be answered in much greater detail than has been possible with the CPS.
- 2) How does the level and distribution of poverty vary as the definition of this concept is altered? In analyzing these questions, the SIPP data offer several significant advantages over the CPS, including better reporting of income, elimination of the income/composition mismatch problem, and more detailed data.
- 3) How do program costs and the distribution of benefits vary as eligibility requirements or other program parameters are changed? The key potential advantage of SIPP is that far more of the detailed information used in the eligibility and benefit formulas is available, thereby reducing the inevitable dependence on the assumptions and estimates embedded in the current models.

The longitudinal aspect of SIPP provides the capability for examining the dynamics of intrayear household behavior, which should provide important new insights. With the exception of the small 1979 ISDP data set, this type of information has not been previously available. A few examples of applications in this area are

- 1) What trigger events precede families starting and stopping participation in transfer programs such as food stamps?
- 2) Poverty is currently measured based on a particular annual accounting period, income, and unit definition, while the key transfer programs base eligibility and benefits on monthly accounting periods for income, and unit composition. These differences obviously make it difficult to assess the effectiveness of transfer programs in reducing poverty. SIPP allows the measurement of the extent to which the differences in definition account for the seemingly low target efficiency of income maintenance programs. With at least 15 months of information, the CPS poverty measure, as well as the transfer program provisions, can be constructed on the SIPP data with reasonable accuracy.
- 3) What events are associated with families entering or leaving poverty?