Statistical Policy
Working Paper 13

Federal Longitudinal Surveys

Prepared by
Subcommittee on Federal Longitudinal Surveys
Federal Committee on Statistical Methodology

Statistical Policy Office
Office of Information and Regulatory Affairs
Office of Management and Budget

May 1986

MEMBERS OF THE FEDERAL COMMITTEE ON
STATISTICAL METHODOLOGY

(November 1985)

Maria Elena Gonzalez (Chair)
Office of Information and
Regulatory Affairs (OMB)

Daniel Kasprzyk
Bureau of the Census
(Commerce)

Barbara A. Bailar Bureau of the Census (Commerce)	William E. Kibler Statistical Reporting Service (Agriculture)
Yvonne M. Bishop Energy Information Administration (Energy)	David Pierce Federal Reserve Board
Edwin J. Coleman Bureau of Economic Analysis (Commerce)	Thomas Plewes Bureau of Labor Statistics (Labor)
John E. Cremeans Business Analysis (Commerce)	Jane Ross Social Security Administration (Health and Human Services)
Zahava D. Doering	Fritz Scheuren

Defense Manpower Data Center	Internal Revenue Service
(Defense)	(Treasury)
Daniel H. Carnick	Monroe G. Sirken
Bureau of Economic Analysis	National Center for Health
(Commerce)	Statistics (Health and
	Human Services)
Terry Ireland	Thomas G. Staple
National Security Agency	Social Security Administration
(Defense)	(Health and Human Services)
Charles D. Jones	Robert D. Tortora
Bureau of the Census	Statistical Reporting Service
(Commerce)	(Agriculture)

PREFACE

The Federal Committee on Statistical Methodology was organized by OMB in 1975 to investigate methodological issues in Federal statistics. Members of the committee, selected by OMB on the basis of their individual expertise and interest in statistical methods, serve in their personal capacity rather than as agency representative. The committee carries out its work through subcommittees that are organized to study particular issues and that are open to any federal employees who wish to participate in the studies. Working papers are prepared by the subcommittee members and reflect only their individual and collective views.

This working paper of the Subcommittee on Federal Longitudinal Surveys discusses the goals, management, operations, sample designs, estimation methods, and analysis of longitudinal surveys.

Conclusions are drawn about where to use longitudinal surveys, and the need to have an evaluation component in these surveys. The Appendices contain twelve case studies of recent longitudinal surveys. The report is intended primarily to be useful to Federal agencies in choosing to do, and then in designing, carrying out, and analyzing data from longitudinal surveys. The Federal Committee on Statistical Methodology intends to organize seminars to discuss the report with interested Federal agency staff members.

The Subcommittee on Federal Longitudinal Surveys was co-chaired by Barbara A. Bailar and Daniel Kasprzyk, Bureau of Census, Department of Commerce.

MEMBERS OF THE SUBCOMMITTEE ON FEDERAL LONGITUDINAL SURVEYS

Barbara A. Bailar* (Co-chair)	Lawrence Ernst
Bureau of the Census (Commerce)	Bureau of the Census (Commerce)
Daniel Kasprzyk* (Co-chair)	Marie E. Gonzalez* (ex officio)
Bureau of the Census (Commerce)	Office of the Information and Regulatory Affairs (OMB)
Barry Bye	Catherine Hines
Social Security Administration (Health and Human Services)	Bureau of the Census (Commerce)
Dennis Carroll	Curtis Jacobs
Center for Statistics (Education)	Bureau of Labor Statistics (Labor)

Robert Casady
National Center for Health
Statistics
(Health and Human Services)

Inderjit Kundra
Energy Information
Administration
(Energy)

Steven B. Cohen
National Center for Health
Services Research (Health
and Human Services)

Bruce Taylor
Bureau of Justice Statistics
(Justice)

ADDITIONAL CONTRIBUTOR TO THE REPORT

Lawrence Corder
Research Triangle Institute
(Previously National Center
for Health Statistics)

*Member, Federal Committee on Statistical Methodology

ACKNOWLEDGEMENTS

This report is the result of collective work and many meetings of the Subcommittee on Federal Longitudinal Surveys. Each chapter had a principal author (or authors), as noted below, but the final report, particularly the introduction and summary sections, reflects contributions from all of the Subcommittee

Many useful suggestions on content and organization were made by Maria Gonzales, chairperson of the Federal Committee on Methodology (FCSM).

Barbara Bailar, Co-Chair of the Subcommittee, prepared the Introduction and the concluding Chapter, which embody the discussions held by the whole Subcommittee.

All of the FCSM members reviewed several drafts and made many important suggestions. The Subcommittee in particular wishes to recognize the valuable contributions made by the primary reviewers: Zahava Doering, Fritz Scheuren and especially Monroe Sirken, who read and commented on two drafts of the complete report.

The principal authors of each chapter of the report are:

Chapter One	Catherine Hines
Chapter Two	Lawrence Corder
Chapter Three	Bruce Taylor
Chapter Four	Daniel Kasprzyk and Lawrence Ernst
Chapter Five	Barry V. Bye

The Subcommittee thanks also the following persons who were responsible for preparing the Case Studies that appear in the Appendix: Edith McArthur (SIPP), Curtis Jacobs (CPI), Steve Kaufman (ECI), Dennis Carroll (NLS-72, HS&B;), Catherine Hines (NLS), Barry V. Bye (RHS, WIE), Stephen B. Cohen (NMCES), Robert Casady (NMCUES), James L. Monahan (LED), John DiPaolo, Robert Wilson, and Peter J. Sailer (SOI).

Catherine Hines edited the report. Joanne Watson (Bureau of the Census) prepared each of the drafts, and the Subcommittee

thanks her for her patience and accuracy.

GLOSSARY OF ABBREVIATIONS

AHS American Housing Survey (Formerly Annual Housing Survey)

CPI Consumer Price Index

CPS Current Population Survey

ECI Employment Cost Index

HCFA Health Care Financing Administration

HS&B; Longitudinal Survey of High School and Beyond

ISDP Income Survey Development Program

ISR Institute for Social Research (University of Michigan)

NCES National Center for Education Statistics

NCHS National Center for Health Statistics

NCS National Crime Survey

NLS National Longitudinal Surveys of Labor Market Experience

NLS-72	National Longitudinal Study of the High School Class of 1972
NMCES	National Medical Care Expenditure Survey
NMCUES	National Medical Care Utilization and Expenditure Survey
OSIRIS	Statistical Analysis software, Survey Research Center, U. Michigan
PSID	Panel Survey on Income Dynamics
RAMIS	Data base management system, Mathematical Research Inc., Princeton, N.J.
RAPID	Data base management system, Statistics Canada, Ottawa

RHS	Retirement History Study
SAS	Data base management system, SAS Institute, Cary, N.C.
SSA	Social Security Administration
SIPP	Survey of Income and Program Participation
SIR	Data base management system, SIR, Inc., Evanston, IL
SOL	Statistics of Income Program, IRS
WIE	Work Incentive Experiment, SSA

TABLE OF CONTENTS

INTRODUCTION

Since the 1960's, the Federal government has sponsored an increasing number of longitudinal surveys as vehicles for research on administrative and policy issues. The goal of the Federal Committee on Statistical Methodology's subcommittee on Federal Longitudinal Surveys is to identify the strengths and limitations of longitudinal surveys, and to propose some guidelines for using them most effectively.

Beginning its work, the subcommittee found that there were multiple definitions of a longitudinal survey, so our first task was to define what this report would mean by the term. The difficulty arises because there are two facets to the definition, design and analysis. To be absolutely clear, one must distinguish between a longitudinally designed survey and a survey with longitudinal analysis. We have elected to put these components together in our definition. The distinguishing features of a longitudinal survey are:

- repeated data collection for a sample of observational units over time;
- the linkage of data records for different time periods to create a longitudinal record for each observational unit; and

- the analysis is based on the longitudinal microdata and refers to data collected over time.

The essential feature is that, from the beginning, there is a plan to elicit data from the future for each observational unit.

This definition excludes some surveys with longitudinal elements, such as the Current Population Survey (CPS). The Survey of Income and Program Participation (SIPP) is included here as a longitudinal survey, although there are as yet no longitudinal analyses of SIPP. Federal agencies also conduct surveys of establishments that have longitudinal elements but these are not yet true longitudinal surveys either. There is an effort to create a longitudinal file for manufacturing firms at the Bureau of the Census. We included this program as a case study in this report because, although it does not meet our definition, it may be of interest to readers. Similarly, Federal agencies maintain longitudinal files of administrative records that do not meet our definition. Yet they may be used in ways that are similar to the analysis of longitudinal surveys, so we have included an example, the Statistics of Income Data Program, as a case study.

Rotating panel surveys* are often described as longitudinal surveys. They are not, but they may share many sampling, estimation, and analysis characteristics with longitudinal surveys. In addition, there is a tendency for ongoing rotating panel surveys to be changed to make longitudinal analysis possible. The National Crime Survey (NCS) is currently considering such a transition, and one possible result of the current redesign activities will be to create a longitudinal NCS data file if the cost is not prohibitive.

There is interest in moving in the same direction with both CPS and the American Housing Survey (AHS, formerly the Annual Housing Survey). We should anticipate that eventually more rotating panel surveys will be modified, or designed from the beginning, to make longitudinal analysis possible. At this time, however, many rotating panels lack longitudinal data files, and many longitudinal surveys are designed without rotating panels.

The subcommittee members examined in detail 12 recent longitudinal surveys sponsored by the Federal Government, as examples and illustrations. These are: (1) the Survey of Income and Program Participation (SIPP); (2) the Consumer Price Index (CPI); (3) the Employment Cost Index Survey (ECI); (4) the National Longitudinal Study of the High School Class of 1972 (NLS-72); (5) High School and Beyond (HS-B); (6) The National Longitudinal Surveys of Labor Market Experience (NLS); (7) the Social Security Administration's Retirement History Survey (RHS); (8) The Social

Security Administration's Disability Program Work Incentive Experiments (WIE); (9) The National Medical Care Expenditure Survey (NMCES); (10) the National Medical Care Utilization and Expenditure Survey (NMCUES); (11) the Longitudinal Establishment Data File; and (12) the Statistics of Income Data Program (SOI). The surveys chosen for case study treatment were selected to represent a variety of sponsors, research questions and kinds of respondents. Each of the 12 case studies is described in the Appendix,,and they are frequently cited to illustrate important points throughout the text.

We hope that the chapters of the text and the case studies in the Appendix will convince readers of four points that emerged from the subcommittee's review of longitudinal surveys. First, longitudinal survey designs are appropriate, and even required, for certain kinds of research. These include, but are not limited to, such topics as gross change, the causes of change, or the role of attitudes in change. However, many longitudinal surveys have not made full use of their longitudinal design in the analysis.

Second, longitudinal survey design, operation, and analysis techniques are still evolving. There are a number of important design issues that are not yet explored or understood. An example is the optimal length of time between interviews, and the number of interviews to conduct to achieve research objectives. To some extent the variations in survey design

---

* A panel is a sample of persons selected to participate at a particular point in the longitudinal sequence. In a rotating panel

survey the sample units have a fixed duration. As they leave the sample, they are replaced by new units which are introduced at specific points in time.

reflect the wide and legitimate differences between the research goals that each survey was designed to accomplish. This does not explain, however, all the existing variation in methods . Decisions about sample design and attrition, about selecting the best respondent or analytical units, about the best estimation, imputation or weighting schemes, or about the impact of varying personal, mail or telephone interviews over the course of a longitudinal survey, have not always been consistent.

Third, the important question of the costs of longitudinal surveys compared to cross-sectional surveys has yet to be answered. There are conflicting reports about the relative costs of the two types of survey. Costs are usually cited as higher for longitudinal surveys, but the costs being reported are confined to data collection costs and processing costs. This does not compare the full range of survey costs including quality costs, costs of analysis, and other such elements which could, in the long run, change the picture of the relative costs.

The fourth and final point that emerged from the subcommittee's review was that the surest method for learning answers to design, operational, and analysis issues is to build an evaluation component into a longitudinal survey. By this means a record of comparative performance is created which benefits others. The case studies presented in this report, in particular, show how progress occurs when evaluation is built into survey operations, and how forethought and planning, far more than additional expense, are needed to increase our knowledge about longitudinal survey design.

This report is presented in 6 chapters. The first chapter is a review of the kind of research question for which a longitudinal approach is appropriate, illustrated with examples. The second and third chapters describe some of the problems encountered in planning and managing longitudinal surveys. Chapter four discusses problems related to sample design and analytical units in longitudinal surveys, and special problems of estimation and weighting. Chapter five describes and evaluates major approaches to the analysis of longitudinal surveys. The final chapter, number six, summarizes some issues the subcommittee members recognized as important, and outlines the need for building an evaluation component into prospective longitudinal surveys; both to answer questions about the quality of data derived from each survey and to answer questions about optimal design for future longitudinal surveys.

3

CHAPTER 1

THE GOALS OF LONGITUDINAL RESEARCH

There are at least five distinctive advantages to using a longitudinal survey rather than a cross-sectional survey some of these advantages are shared by rotating panel surveys.

1. A longitudinal sample reduces sampling variability in estimates of change. This is an advantage shared with rotating panel surveys such as CPS and NCS.

2. A matched longitudinal file provides a measure of individual gross change for each sample unit. This is an advantage shared to some extent by rotating panels, which can provide a measure of gross change, but not usually on an individual basis.

3. Longitudinal survey interviews usually have a shorter, bounded reference period that reduces recall bias in comparison to a retrospective interview with a long reference period. Rotating panels such as CPS and NCS also share this advantage. Longitudinal surveys with long intervals between interviews may lose this advantage.

4. Longitudinal data are collected in a time sequence that clarifies the direction as well as the magnitude of change among variables.

5. Longitudinal interviews reduce the respondent burden involved in creating a record that contains many variables. A single interview could not collect comparable detail without excessive respondent burden and fatigue. In addition, the quantity of data collected in a longitudinal survey is usually greater than that from several cross-sectional surveys because of the correlational structure of longitudinal data.

There are also some distinct disadvantages to longitudinal surveys. Some of these are:.

1. The analysis of longitudinal surveys is dependent on the assembly of the microrecord data. The full advantage of compiling a detailed longitudinal record with many variables may not be available until years after the start of data collection.

2. Beginning refusal rates may be comparable to those of cross-sectional surveys, but the attrition suffered over time may create serious biases in the analysis.

Principal Author: Catherine Hines

3. A longitudinal survey, including several data collections, is more costly than a single retrospective cross-sectional survey. A longitudinal survey may be less costly than a series of cross-sectional surveys. It is speculative whether a longitudinal survey is more costly than a rotating panel survey.

4. The estimates of gross change derived from longitudinal surveys tend to be inflated over time by simple response variance, The combined or net effect of such influences as simple response variance, response bias and time-in-sample bias effect on longitudinal estimates of gross change are still poorly measured.

5. Longitudinal surveys are often improperly analyzed, not taking into account longitudinal characteristics or

attrition.

For some research goals, the advantages clearly outweigh the disadvantages. For other research goals this may not be the case. Research goals that demand longitudinal surveys are described in this chapter.

A. Measuring Change

Both cross-sectional and longitudinal surveys can be used to measure change. The National monthly estimate of unemployment based on the CPS is always compared to the estimate for the previous month or the same month a year ago. Estimates of such things as crime victimizations, retail sales, housing starts, or health conditions are all compared to estimates from a previous time period. None of these data are currently based on longitudinal surveys.

Which measures of change need a longitudinal file structure? One example is the components of individual change. These are measures of gross change for the observational units between points in time.* Longitudinal data are frequently displayed in a time-referenced table, showing the characteristics, attitudes, or beliefs of the sample at time 1; cross-tabulated by the same characteristics, attitudes, or beliefs at time 2. Another example is the average change for an observational unit. As pointed out by Duncan and Kalton (1985), if data are available for several time points for each observational unit, then a measure of average change or trend can be estimated. Finally, a longitudinal design permits the measurement of stability or lack of stability for each observational unit.

Measures of gross change are of interest in several of the case studies described in this report. Respondents are followed through employment and unemployment (NLS), training and the labor force (NLS-72, HS&B;), into and out, of poverty (SIPP), or between health, treatment, and disability (NMCES, NMCUES, RHS, WIE). The focus is sometimes on movement across an arbitrary threshold (such as poverty, defined by household composition and income), and sometimes on a continuous measure.

The observation periods in a longitudinal survey are commonly called waves. A wave describes one complete cycle of interviewing, from sampling to data collection, regardless of its duration.

In independent (i.e., cross-sectional) samples, subpopulations with very different gross-change patterns are indistinguishable if the sum of the changes is similar. This has been important to studies of employment. The NLS, for example, can distinguish a hypothetical population where 15% of the people are never employed, from a population where at each interview a different 15 % respondents report unemployment. A cross-sectional survey could not make the same distinction, which is vital to the development of intervention policies. Another example can be cited from the field of social indicators research. A series of variables, measured longitudinally, can be used to construct models for estimation to examine change over time with great elegance. (See Land, 1971, 1975.)

Young adults in the years after full-time school are frequent longitudinal survey subjects (NLS Youth Cohorts, NLS-72, HS&B;)

because individuals in these years are known to pass between statuses (employment and unemployment, school and training programs, in and out-of the armed services, between households) rapidly and irregularly. Cross-sectional studies would miss all the individual reversals and repetitive change. To develop detailed models of the causes of change in these fluid populations, longitudinal measures are needed to capture the record of individual and gross change.

For example, cross-sectional studies of college enrollments have generally found relatively high stability over a number of years, whereas analysis of NLS-72 data identified frequent individual change occurring at a stable rate. A substantial percentage of the college students surveyed exhibited erratic enrollment patterns characterized by dropping out or transferring between 4-year and 2-year colleges. In light of these findings, student financial assistance (grants and loans) have changed.

Legislation has shifted aid to channel the funds directly to the students, who choose the college they wish to attend -- rather than channelling the funds to college officials, who decide how the funds are doled out to enrolled students.

Studying the relationship between attitudes and behavioral change poses particularly difficult problems in research design. The problems inherent in determining which variable in a pair changes first are present, and they are exacerbated by the problems encountered in surveys of subjective phenomena, such as attitudes. Using retrospective questions to ask respondents to reconstruct thoughts or feelings as they existed in the past has proved unreliable.

Prospective longitudinal surveys provide the most reliable data on change in knowledge or attitudes, because longitudinal measures are collected while the subjective states actually exist. This appears to reduce the bias frequently caused by suppression or distortion of respondent recall. In addition, unlike retrospective measures of attitudes, contemporary measures can sometimes be probed or even verified.

The longitudinal surveys of high school students (NLS-72 and HS&B;) demonstrate the method's power to collect data on changing subjective states, and to study causation. These surveys have measured attitudes and expectations about employment, and subsequent employment experiences and behavior. The data, which could not have been collected cross-sectionally, can be analyzed to understand the formation of attitudes, as well as to evaluate the effects that attitudes have on subsequent behavior.

When the research goal is to measure a component of individual change, longitudinal surveys have strong advantages. They are the only method available to collect data on a recent occurrence basis over a long period of time. Although a retrospective cross-sectional survey could be used to attempt the same thing, the recall bias may be a strong force against this decision. The bias from the attrition in a longitudinal survey has to be balanced against the bias or lack of information in a retrospective cross-sectional survey. The bias from attrition is usually preferred.

Price and wage changes are measured in longitudinal surveys (i.e., the CPI and ECI) because the longitudinal sample design holds other variables constant. The assumption can be made that whatever unknown sampling bias exists in later waves was also present in earlier waves, and can be dismissed as a possible source of the changes being measured.

B. Assembling Detailed Individual Records

Longitudinal surveys generally provide researchers with more detailed records for each individual than is practicable through a cross-sectional design. In a longitudinal design, an extremely detailed record can be accumulated for each subject without making any single observation period (i.e., interview or wave) excessively burdensome. By 1982, for example, records for the original respondents in the NLS contained up to 1,000 data items for each sample case. To create a record of comparable detail complexity would have required a one-time questionnaire of extraordinary length. In addition, responses referring to earlier time periods would have been reconstructed from memory, reducing their reliability. In many instances, researchers are looking for cause-and-effect relationships that are more likely to be accurate if the data are compiled on a current rather than retrospective basis.

C. Collecting Data That is Hard to Recall

Some surveys ask questions that respondents have difficulty in answering precisely or objectively after much time has passed. These include questions that call for the kind of detail that people seldom recall clearly (such as complete records of expenditures, or health treatments), and questions that refer to events that respondents tend to telescope, embellish or suppress in their memories after time has passed (such as crime victimization, health problems, or visits to the doctor).

Questions such as these have been used successfully in longitudinal surveys, in which the previous interview provides a clear marker to bound respondent recall, and which are constructed with short reference periods between interviews. For example, the Consumer Expenditure Survey, conducted as part of the CPI program, collects detailed records of household spending patterns through longitudinal interviews. (See Case Study no. 2 in the appendix.)

A longitudinal survey with relatively short reference periods is one of the best methods for producing aggregated data for a longer time period, such as a year. For example, the primary goal of the NMCES and,NMCUES programs was to develop estimates of medical expenditures for a calendar year. This was accomplished by obtaining medical expenditure data every 3 months and Compiling an annual total. A similar example is the new continuing Consumer Expenditure Survey, which covers all consumer expenditures. The SIPP program employs a similar design, using interviews at 4 month intervals to produce annual aggregates. The relatively short, bounded reference periods for these longitudinal surveys improve reporting by eliciting events closer to the time they occur. This increases the completeness of aggregated estimates and reduces error.

D. Modelling Studies and Pilot Programs

The detailed case histories built up in longitudinal surveys are important in analyzing the impact of alternative policies or intervention strategies. The complex individual case records accumulated in a longitudinal panel survey provide a microcosm in which the impact of changes can be simulated. Questions can be answered about the probable impact of changing a program's eligibility criteria, for example, or about the benefits which specified classes of respondents might anticipate under,various program changes. Intervention programs can be evaluated through longitudinal surveys to Study their effect on respondents with known characteristics. A sufficiently detailed record makes it possible to simulate alternative interventions, and predict a range of effects. (See Case Study 9 on the WIE, for example.)

In some cases longitudinal surveys, pilot intervention programs and Federal policy experiments evolved together in the 1960's. Several longitudinal surveys authorized as components of pilot or experimental intervention programs to measure program effects and ensure that decision-making information would be available when it was needed. Longitudinal data collection components were built into pilot income maintenance programs, for example, administered temporarily in cities in New Jersey, Indiana, Colorado and Washington State.

In conclusion, tho points about the periodicity of longitudinal research should be stressed. First, longitudinal data are never available immediately; any data that are based on the sequence of measures over time cannot be fully extracted until the final measures are collected. If information is needed at once, another research design has to be used which incorporates some alternative to a true longitudinal approach; such as retrospective measures, or the use of administrative records. Even if the quality of data from a longitudinal survey would be clearly superior, that would be irrelevant if the schedule outweighs these other considerations.

Second, longitudinal data can be used cross-sectionally to provide immediate data as long as the research focus is not specifically on changing measures over time. Each wave of a longitudinal survey can also be analyzed as a cross-sectional survey. Thus some data can always be made available immediately. Record data from non-going longitudinal surveys can be analyzed quickly from a cross-sectional perspective to serve certain analytical purposes without delay. It is also possible to add questions to the current waves of a longitudinal survey to meet immediate data needs, using an existing longitudinal sample and base-line demographic data for maximum efficiency. In these ways a longitudinal design adds analytical strengths without sacrificing the potential for cross-sectional research.

CHAPTER 2

MANAGING LONGITUDINAL SURVEYS

As described in the previous chapter, prospective longitudinal surveys have proved to be an important research approach, but certain limitations have also emerged that must be considered when these surveys are planned. The problems related to staff and management of longitudinal research differ in kind as well as degree from those encountered in cross-sectional research.

The core of the problem in managing a longitudinal survey is a conflict between the need for long-term and for short-term resources. Plans and funding must be stable over many years, but the need for staff rises and falls over the course of a longitudinal survey. Most organizations sponsoring longitudinal surveys have solved the dilemma through some combination of permanent and temporary staff. Fluctuations in resources are less pronounced in longitudinal surveys that employ non-going rotating panels (such as SIPP or, to some extent, the CPI) than they are in fixed panel surveys in which interviews are conducted at longer intervals (such as NLS, NLS-72, or HS&B;).

The major difficulty faced in planning and managing a longitudinal survey is in maintaining a core group dedicated to the project, and maintaining consensus between this group and senior agency staff. These groups tend to view long-term commitment of Staff and resources in different ways. The schedule, funding, and staff needs of a longitudinal survey are viewed differently by survey designers, by agency directors, and by those responsible for operations. It is a constant challenge to generate commitment to a long-term goal such as analysis of data, when senior staff with direct authority over the project often changes before the survey is completed.

A. The Need for Long-Range Planning

The need for long-range planning and organization for a longitudinal survey should be brought to the attention of senior staff very early with a planning document that outlines the workload, survey tasks, and anticipated products over time. The planning document should be prepared in conjunction with an analysis plan, and the design of the instruments and procedures will then follow once all groups are in agreement with the planning document.

Long range planning is vitally important to a longitudinal survey, because it promotes enduring support at a senior agency level, it widens the pool of sponsors and supporters; and it begins the process of documentation that ensure continuity of operations.

Principal Author: Lawrence Corder

A large-scale longitudinal Federal survey generally has at least nine principal management phases which may be briefly described as follows:

1. Budget Planning. Up to five years before data collection is to begin, a general plan must be conceived and provisions made to obtain continuing staff and funding

resources throughout the longitudinal project.

2. Development of Position Papers. These are draft planning documents which discuss options, costs, and yields associated with various sampling plans, data collection designs, or questionnaires. These ensure widespread and enduring support for the longitudinal research.

3. Procuring outside assistance. If a contract is to be awarded, requests for proposals must be prepared, cleared and advertised, and responses must be evaluated before a contract is signed. This is a common approach to levelling out resource needs.

4. Final Research Plans. This stage includes final OMB clearance, conduct of field tests, revisions as necessary, and detailed agreements with any other

cooperating agencies.

5. Data Collection. This refers to the full-scale field data collection. Longitudinal surveys (such as NLS) which have been extended beyond the original research period have repeated these 5 stages independently several times.

6 . File Preparation. Development of the system for data entry, data base design, processing, etc., may also require systems for optical scanning of questionnaires, machine/or manual edit steps, preparation of code books, the construction of composite variables, plans to preserve privacy in public data files, and numerous other activities. Each operation must be fully documented, to ensure comparability between waves.

7. Planning the Analysis. While the overall goals oft he analysis must be planned in the early stages, some details cannot be finalized until the data are available on computer files and code books are completed. Also, as policies shift, new analytical priorities must be met. In all cases, this process requires plans which may include in-house analyses and contracts for analyses. Contracts require a repetition of the procurement process described in phase 3.

8. Conduct of Analyses. These may go on for several years. Cross-sectional, analyses can be conducted as soon as one wave of interviews has taken place. Longitudinal analyses take place after some or all other waves are completed.

9. Publications. With in-house and professional peer reviews, these may continue for several years.

Each phase requires substantial time to complete, contains specific activities and results in the preparation of key documents. The final products of any longitudinal surveys are usually public-use data files and reports.* Ideally, these should be supplemented by rapid preparation of in-house documents as part of the policy-making process. Schedule milestones and due dates are part of any longitudinal survey, and the ultimate success of the project and even the usefulness of the analytical results may be judged against their timeliness.

It is not unusual for a longitudinal survey to consume a decade or more from inception to completion of the publication plan. The NMCES and NMCUES Studies, for example, both took 8 to 10 years to complete. While field operations and the period for analysis vary with each survey's objectives and resources, the successful pre-field period is probably very similar in each case. The planning period should be dedicated to achieving consensus internally, then to producing instruments and obtaining clearances and approvals (for contracts as well as for questionnaires). A typical schedule for completing pre-field activities alone (excluding budget planning) would frequently require 12 to 18 months.

Some of the most severe criticisms of longitudinal surveys have resulted from insufficient planning. It is not uncommon, for example, to omit thorough planning of the analysis. Then, at a production stage, it is discovered that people have different ideas on the tables and data to be produced and analyzed. It is also necessary to plan the linked files carefully so that the data needed for longitudinal analyses are readily available. Unfortunately, the planning of budgets and field work often takes precedence over the planning of processing and analysis, sometimes leading to delays, acrimony, and sometimes shifts in support.

B. Funding Longitudinal Research

The actual unit costs of doing longitudinal surveys may be no higher than for a series of cross-sectional surveys of comparable size and complexity (Wall & Williams:30). There is conflicting evidence on comparable costs, probably reflecting non-standard cost reporting on survey operations. Funds, however, must be committed over a number of fiscal years and budget plans are not easily altered. There is a trade-off to be made when errors are discovered or improvements can be implemented. Additional costs must be carefully considered, as well as the effect of changes in methodology on the longitudinal analysis. Errors, of course, should be corrected or, if too costly, an indication of their effects provided. Changes in methodology are different from changes necessitated by errors and must be thoroughly explored. Provision should be made to share information with analysts and data users on real change vs. methodologically-induced change. (The change to computer assisted telephone interviewing is one such change that needs careful exploration.) If errors or methodological changes result in higher costs, alternative methods of meeting those costs should be considered: higher funding, smaller sample size, more time between interviews, delayed processing, and so forth.

Surveys of business or industrial establishments are often an exception to this rule, to protect the identity of large firms that dominate certain samples.

Inter-agency cooperation can help meet long-term funding needs. The Health Care Financing Agency (HCFA) and the National Center for Health Statistics (NCHS) chose this approach in conducting NMCUES. Inter-agency agreements frequently involve the Census Bureau for data collection and analysis, but they may also be used between other agencies with related research goals. Inter-agency Cooperation in longitudinal surveys could take the form of joint sponsorship of a new longitudinal survey, or it could be in the form of using an existing longitudinal sample as a vehicle for research to save the cost of starting a new longitudinal survey.

The NLS-72 provides an example of a consortium approach: For the fifth follow-up interview in NLS-72, the National Science Foundation appended questions on math and science teachers, and the National Institute on Child Health and Human Development joined with the National Center for Education Statistics (NCES) to fund questions on child care and early childhood education issues. Longitudinal surveys are generally long term projects with significant start-up costs. If a survey can he constructed to serve more than one agency through an inter-agency agreement, start-up costs may be shared and several agencies will be bound to multiple-year funding commitments.

When agencies select outside contractors to conduct longitudinal research, competitive procurement is required. The decision to use a contractor to conduct a survey increases the time needed to start a project, because approval of contracting plans must be added to other planning tasks. One advantage of contracting out the survey work is that it gives an agency access to additional staff support in cases where the agency has no authority to add permanent staff.

Contracting for data collection by an outside agency may or may not be more expensive than employing a government organization for this purpose. In comparing costs, NCES found that the first NLS-72 follow-up, conducted by the Census Bureau, cost slightly more than the second follow-up, conducted by Research Triangle Institute (RTI), despite inflation. Other longitudinal surveys, including NMCES and NMCUES, have had just the opposite experience. The most cost-effective mode of operation appears to depend on the kind of survey, not on the agency conducting it.

The duration of longitudinal surveys often requires periodic recompetition once a competitive award has been made. As a result, agencies have found themselves switching contractors part way through the data collection phase of a longitudinal survey. The competitive award of each data collection wave can, however, help control overall survey costs, because it provides contractors with an incentive to hold down their costs.

The possibility of changing contractors over the life of a longitudinal survey requires a detailed documentation of methods that goes far beyond what is needed for any one-time survey. This level of documentation was not anticipated when the original contract to collect data for NLS-72 passed from the Educational Testing Service to RTI, and the change in contractors caused difficulties. Based on this experience, NCES now builds a sub-contract to the previous contractor into any subsequent data collection awards. As a result, a later transfer of the NLS-72 contract from RTI to NORC was accomplished without problems.

C. Staff Needs

Staffing requirements for a longitudinal survey typically vary substantially, both by number and by type of staff throughout the history of the project. Staffing is much more controlled in rotating sample surveys, whether they are longitudinal or cross-sectional. Funding and staff needs for a longitudinal survey are much greater during the data collection period than during any other phase. However, some of the types of people needed for data collection, such as interviewers, are not needed in later phases. Staff monitors for field work and data processing are in high demand at early stages as well as intermediate stages. Because of sporadic needs, the use of a core group of survey professionals in combination with temporary staff, or interagency agreements or outside contracts, can be the best method to ensure adequate staffing for the entire effort.

To distribute the costs of a contract more evenly over a longitudinal survey, NCES and NCHSR have used incrementally-funded contracts. During the longitudinal survey, separate contracts are awarded for each phase or wave. Each contract extends over two or more years. At any point, some survey tasks are being advertised for competition while others are being completed under contract. Looked at from the standpoint of each fiscal year, the total costs and level of effort remain more nearly constant. NCES has also found that giving agency survey analysts the responsibility for monitoring contract performance will help control variations in staffing patterns.

By employing temporary peripheral groups in addition to permanent staff groups, two problems are solved: Research staff needs are met without adding permanent personnel to an agency; and peak workload needs are met without jeopardizing tight survey schedules. Inter-agency agreements or contracts not only bind parties to a specified set of research goals, but they also permit the level of staff effort to rise and fall as needed.

D. Maintaining Core Staff

The duration of longitudinal research projects creates another management problem (which has been called a Methuselah effect by

Herbert Parnes). Each phase of a longitudinal study, such as planning, data collection, or analysis, is frequently carried out by different individuals, who may not even be part of the same organization. The relative inflexibility of a longitudinal study plan is an analytical necessity, but it could also prevent interim analysis or refinements in the design. For these reasons, it has been suggested that non-going longitudinal surveys may hold little interest for the calibre of professional staff that is needed for management or analysis (Wall & Williams: 35).

NCES, however, has successfully attracted talented analysts to manage the agency's longitudinal surveys. To some extent this may be because NCES ensures that the Agency's staff have challenging responsibilities for program analysis. Agencies which see only data collection as their primary mission may be more apt to encounter the staff problems recognized by Wall and Williams. in order to allow mid-course corrections and modifications of the survey plan, NCES uses a multi-phase sampling design (as in HS+B). This, too, contributes to the flexibility of the NCES longitudinal survey program.

E. Data Collection and Processing Schedules

Longitudinal surveys have become notorious for developing serious backlogs because data collection takes precedence over all other tasks. The schedule for observations is usually the least flexible aspect of the design, because each subject must have an identical record structure. As data collection continues, it creates an ever-growing backlog of other procedures, such as analysis.

Uncompleted tasks tend to accumulate, becoming increasingly difficult to finish. To prevent backlogs and delays, a longitudinal survey must be well-organized and planned so that analysis and data release keep pace with data collection.

Data collection schedules are not the only factor in backlogs. Another factor is data processing, including file linkage. Survey organizations that are more accustomed to doing cross-sectional surveys or other non-longitudinal surveys often have difficulty recognizing the special processing needs of longitudinal surveys. Databases need specification, key variables,need identification, and a policy on imputation needs to be thought through. Ideally, all this needs to be done when the survey questionnaire is designed, but this ideal is seldom, if ever, met.

F. Data Analysis

Data analysis is often looked on as the rewarding part of the job after the difficulties of data collection and data processing. Analytical interests often go beyond the agency conducting the study. Some agencies include analysis contracts in their contracting for services. Usually some analysis is done by agency personnel.

One possibility to counter some,of the delay caused by the time it takes to complete a longitudinal survey is to analyze each wave as if it were from a cross-sectional survey. This not only provides timely data, but raises questions to be answered at later stages, and generally whets the appetite for more data and more analysis. Recent data from non-going longitudinal programs can be analyzed relatively quickly to serve some analytical purposes without delay. It is also possible to add questions to the current data collections of a longitudinal survey to meet immediate data needs.

G. Release of Data

A principal goal of any longitudinal survey should be to produce public use data tapes and analytical reports rapidly, both for policy-makers and the interested public. If public use files are to be created, then procedures to

protect confidentiality must be worked out in advance, File structure and documentation need to be readily available. Variance estimation must be provided for those using the file. The permanent survey staff should maintain a role in the preparation of files and reports, so that their expertise and interest are not lost.

In conclusion, longitudinal surveys, sometimes taking 5 years or more to complete, inevitably encounter staff changes. Two management approaches can minimize the loss of institutional memory. First, it is vital that every survey activity be documented. Interview instructions, edit specifications, variable definitions, file layouts, sampling, weighting and imputation methodologies, all instruments and procedures should be recorded and readily available. This task is very labor-intensive and, unfortunately, apt to be slighted when staff time is short. Second, inter-agency agreements or contracts may clearly lay out both the procedures to be used and the final products. It is also wise to specify key contractor staff persons who cannot be replaced without sponsor approval. These actions are important to minimize the effect of staff changes and to prevent errors and delays.

CHAPTER 3

LONGITUDINAL SURVEY OPERATIONS

The principal differences between field and processing operations in one-time surveys and in longitudinal surveys are created by the use of time as a significant factor in research. Longitudinal surveys typically encounter changing conditions, and survey designers have developed and evaluated a variety of methods for controlling the problems that can be caused by change in the sample or changes in the design or administration of the survey.

A. Sample change over time

The composition of the sample may be expected to change across waves for a variety of reasons. Respondents may refuse to participate, they may die, they may move and cannot be found, or they may leave the sampling frame (e.g., by entering an institutional population or by moving abroad). The danger is that the sample becomes increasingly less representative of the target population as time passes. To minimize the effects of these problems, new observational units are routinely introduced into the samples of some continuing surveys as time passes.

1. Selection of new units into sample

For some longitudinal surveys, they are a number of concerns related to the length of time respondents are kept in sample. Respondent burden across several interviews may produce a decline in the quality of data gathered or may result in increasing refusal rates. Respondents may also leave the sampling frame, move and cannot be tracked, or die, thereby affecting the representativeness of the sample. for these reasons, it may be desirable to institute a rotating panel design, which regularly moves new respondents into the sample and retires other respondents after a fixed number of interviews or period of time.

The Survey of Income and Program Participation (SIPP), the National Crime Survey (NCS), the new Consumer Expenditure Survey (CE), and the Consumer Price Index (CPI) have all adopted rotating panels. SIPP introduces new respondents annually and retains them for 2-« years (7 or 8 interviews) before rotating them out; NCS introduces new respondents monthly and interviews them for 3-« years (7 interviews). The CE Survey introduces respondents monthly and interviews them five times on a quarterly basis, while the CPI introduces new respondents once every five years and interviews monthly or bimonthly.

Fienberg and Tanur (1983) note that rotating panel designs may create some problems of inference, according to conventional sample survey theory, in that random selections of respondents occur at different times for different respondents. The argue, however, that this is only important when date of selection is related to temporal changes in the phenomena the survey was designed to measure. The inferential difficulties which might result from a rotating panel design must be balanced against the reduction of attrition-related bias, which is the alternative.

2. Movers

Some respondents may be expected to move from originally sampled housing locations (or telephone numbers) during their time in sample. Depending on the purpose of the survey and procedures adopted to track movers, respondent mobility has varying implications for the representativeness of the sample over time. a number of factors may enter into decisions regarding whether, or how, to follow movers.

A crucial consideration is to determine the most important unit of observation for the survey. A longitudinal survey of persons may be designed to follow sample individuals or households, if the substantive goals of the survey would be served by retaining as many of the originally sampled respondents as possible. A number of surveys, such as SIPP and NLS, focus on individual and household economic data, which continue to be relevant to the purposes of the survey regardless of respondent mobility. Consequently, following movers is an appropriate means to maintain data quality over time for such surveys.

Following movers may create other problems, however. For instance, if there are ecological correlates for the phenomena of interest, such as crime or quality of housing, then following mobile respondents may result in deterioration of the geographic representativeness of the original sample, with a consequent potential for bias in some measures for later waves. A rotating panel design may minimize this problem, because newer respondents are more likely to reside in the originally sampled housing location.

Another reason for following movers is that respondents may move for reasons related to the substantive goals of the survey. This makes it important to know why they move. If this is the only reason for following movers, then collecting data for only one wave after a move may be enough. In NCS, for example, some respondents may move from a high-crime area to a safer neighborhood, and it would be important to determine the proportion of moves which were related to crime victimization can be measured, but not the future consequences of victimizations for such movers.

The SIPP is attempting to follow all individual movers. Because living arrangements vary according to economic circumstance --and affect eligibility for social welfare programs -- a change in residence can be related to changes in income and program participation. Thus, for SIPP it is crucial not to lose data on movers. The CPI, on the other hand, follows only those movers who provide services, such as doctors or lawyers, since their expertise is the item being purchased. When a commodity outlet changes location, this move is considered a unit "death" and the CPI record is terminated.

The actual procedures developed for following movers are likely to reflect the field procedures of the organization conducting the survey, the collection mode used, the distance involved, and the costs associated with tracking movers. If the organization conducting the survey uses decentralized collection procedures, a respondent moving from the jurisdiction of one regional office to another may be more difficult and more expensive to track. Also, the costs of following movers may be greater if a face-to-face collection mode is used, rather than a telephone design, where tracking procedures may

20

be limited to obtaining a new telephone number. Depending on the cost, administrative difficulty, and proportion of respondents who move far enough to create problems, it may not be desirable to follow all movers or to rely on standard collection modes. SIPP field procedures, for instance, indicate that personal interviews need not be administered if the respondent has moved beyond 100 miles from any sample PSU, and rules also differ for respondents younger than fifteen years of age. If survey procedures allow telephone interviews in lieu of face-to-face interviews, a phone contact may be a desirable alternative for movers who are difficult to reach.

The type of sample involved may also affect the ease with which movers may be located. For instance, it is usually easier to find a mover through neighbors or subsequent occupants of a sample housing unit if an area sample has been adopted rather than with a random digit dial sample. Asking respondents to notify the field office with pre-printed cards when they move can be a partial solution, but this option relies heavily on the respondent's cooperation.

3. Attrition

When projected across waves of a longitudinal survey, manageable levels of non-response in a cross-sectional survey can become significant sample attrition. The potential for attrition in a longitudinal survey sometimes limits sample definition. Tracing mobile respondents generally accounts for a large proportion of field problems as well as costs, and refusal rates are likely to grow over the life of the survey. Incomplete records and missing interviews create analytical complexities that are unparalleled in cross-sectional research. Attrition is most dangerous when it is correlated with the objectives of the survey. For example, there is evidence that sample attrition may be related to victim status in the NCS. To the extent that the sample loses victims at a faster rate than non-victims, estimates from later waves will be biased. Also, Fienberg and Tanur(p.17) note than in social experiments disproportionate loss of respondents for different treatments may be a problem, because treatments often vary in their attractiveness to participants.

Sample attrition between observation periods may create the illusion of change when means are compared between waves, without adjusting for non-response. In study focused on identifying change, there is a risk that changes are spurious, due to sample attrition. In addition, respondent participation that varies from panel to panel could produce the appearance of change even when aggregate non-response is stable. The estimates of central tendency (Cook & Alexander: 191). Mean test results from longitudinal panels of students taking ETS exams were compared to mean test results derived from a cross-sectional survey of the same population. The means were significantly different, which the analysts attributed to selective attrition in the longitudinal sample.

Effects of attrition in demographic surveys have been harder to predict. Attrition does not necessarily created unmanageable bias in a longitudinal survey: The NLS was still contacting 92 percent of living respondents 3 years after the original contact, and still contacting 80 percent of eligible respondents 12 years after the study began (U.S. Department of Commerce:321). In the ISDP panels of 1978 and 1979, attrition did not climb steadily over the five or six interviews administered to respondents. Instead, it leveled off and then declined slightly over all waves (Ycas:150). Nonetheless, a combination of attrition and varying participation from wave to wave can create serious

problems in creating complete records. In the 1979 ISDP panel, for instance, only two thirds of the original sample persons had complete interview records (Ycas:150).

Calculating the response rate in longitudinal surveys is itself difficult. The measures used in cross-sectional research are often not adequate for measuring non-response in complex records, as they do not reflect cumulative non-response across waves and do not take into account changes in the size of the eligible sample due to births, deaths, and the addition of new household members. The illustrate, non-response for entire housing units in the NCS is sometimes reported at 4 percent. However, when records for housing locations are linked to form a longitudinal file, it has been found that over half of the originally sampled housing units are missing at least one interview. This discrepancy is due to the fact that the former figure is a cross-sectional measure of unit non-response in a particular wave and does not account for the approximately 10% of sample housing units unoccupied at the time of interview (Fienberg & Tanur:14). This figure also dies not cumulate non-response over time. While the lower figure is an appropriate measure for many cross-sectional uses of NCS data, it clearly is inadequate for reflecting the completeness of linked housing unit records.

The methods that have been developed for tracing respondents in longitudinal surveys have been successful, but they have also proven to be expensive. The Census Bureau has estimated that the cost of contacting each wave of an ISDP research panel increase by 8 percent over the previous wave, due to the costs of following movers and interviewing additional households (Fienberg & Tanur:11-12, White & Huang). However, NCES also found that per-unit tracing costs for the High School and Beyond (HS&B;) Survey were approximately 20% less than the cost of base year sampling, which illustrates the economies which can be realized by mounting a longitudinal study, rather than separate cross-sectional studies. To control costs, as well as potential bias, each longitudinal survey must investigate the characteristics of respondents who move. Depending on empirical evidence about how atypical non-respondents are, a judgment can be made about the proper balance between the costs of tracing respondents and an acceptable level of non-response.

Sample definition offers another approach to limiting unscheduled attrition. The probability of becoming a non-respondent is not randomly distributed among the population. In longitudinal samples such factors as rural resident, interval since contact, and region of the U.S. affect the probability of maintaining contact (Artzrouni:21-24). Some longitudinal designs have therefore sought to minimize attrition by avoiding the respondent classes that are most susceptible to attrition.

Setting aside respondent classes to control attrition can conflict with attaining a sample that truly represents the reference population. However, a sample chosen without regard to eventual tracing difficulties may also gradually lose its representative power through attrition. Only empirical evidence can indicate the extent to which characteristics that predict attrition co-vary with the characteristics that the study is designed to investigate. A sampling design which sets aside respondent classes with potential attrition problems should be undertaken only after careful consideration of the relative magnitude of bias which could be introduced by such a strategy and other alternatives, such as imputation for missing data or performing analysis on the remaining sample cases of an initially representative sample.

In cohort or panel studies, which require measurement to begin and end at the same time for all respondents, implementation of a rotating panel design, which reduces the impact of attrition by replacing respondents over time, will clearly not serve the goals of the survey. One possible strategy for dealing with attrition in such studies is to impute.

missing data, based either on statistical models or on complete data from prior waves or from respondents with similar characteristics. Another possibility is to reweight the sample for each wave to reflect non-response for various demographic groups in the sample. (See Chapter 4.)

Duncan, Juster, and Morgan (1982) model such a procedure for the Panel Study of Income Dynamics (PSID), conducted by the Institute for Social Research (ISR) at the University of Michigan. They compare results for data gathered with persistent efforts to pursue respondents and for the data set which would have resulted if less intensive respondent contact strategies had been adopted. When the latter is reweighted to adjust for missing cases and compared with the first data set, there are minimal differences in outcome measures. While this procedure has promise for minimizing bias resulting from non-response across waves, it may also allow some relaxation in pursuing respondents, allowing cost reductions in survey administration. The authors do note, however, that reweighting entails some risk of covariation-related bias in multivariate estimates, especially for models that are not well specified, and that maintaining an adequate number of respondents in some key subsamples may remain a problem.

A reasonable precaution to minimize the deleterious effects of sample attrition is to minimize respondent burden, which has been variously described as the amount of time which an interview entails or as the complexity of the task required of respondents for successful completion of an interview. Under the Paperwork Reduction Act of 1980, each Federal statistical program is restricted to a limited number of hours available for data collection in a fiscal year, thereby encouraging reduction of the burden placed on respondents. In addition to the statutory reasons for limiting the length of Federally sponsored surveys, controlling respondent burden may also improve data quality for longitudinal surveys in a number of ways. An important aspect of this data quality enhancement is that, respondent participation may be encouraged by reducing interview tedium, thereby reducing refusal rates and enhancing the representativeness of the sample over time.

Respondent burden hours may be reduced by a careful evaluation of the utility of collecting information in every wave. The SIPP, for example, minimizes respondent burden by dividing the survey into a core questionnaire ad ministered at each interview, plus "topical modules" to collect data not required as regularly. Sometimes only a subsample of respondents should answer certain topics. Finally, lengthening and/or varying the intervals between waves should also be considered as a means for reducing respondent burden. The CPS, while not a longitudinal survey, adopts this strategy of varying tim e between interviews. Respondents are interviewed for four months in succession, not contacted for the following eight months, and then interviewed for a final four months.

4. Changes in Units of Observation

A slightly different sample of respondents participates in each wave of a longitudinal survey. Such changes in sample may result from scheduled introduction or retirement of sample units in a rotating panel design, from attrition, or from introducing new respondents when household composition changes. This variation causes difficulties related to defining the correct reference population, in weighting for item non-response, and in weighting respondents who enter and leave the sample. In addition, the changing sample of respondents and aggregate units creates unique difficulties in analyzing data above the person level A variety of approaches has been used to define units of analysis in longitudinal research, and each has specific problems and strengths. These are discussed in detail in Chapter 4.

It should be noted here, however, that all weighting adjustments should be planned simultaneously. The problem of adjusting for non-response is the converse of problems created by persons entering the sample, and the adjustments for entrants and non-coverage, once selected, can be accomplished in a single operation.

Split and merged households present particular problems for sample comparability across waves. Such recomposition of households creates obvious difficulties for longitudinal matching, which will be discussed below. However, changes in household membership also raise questions about how to treat new members of split households who were not members of the originally sampled household but who came into sample because of their associations with original sample persons. Rules developed by the ISDP offer one method which seems generally applicable to a number of surveys: New household members were added to the sample, but if they left the household, or if this household subsequently split, only those members who were selected for the original sample were followed. This procedure avoids excessive growth of the panel, thus minimizing artifactual changes in aggregate panel statistics, but still collects relevant household data which correspond to data from "stable" households.

Whether a change in a household constitutes the birth or death of the sample unit depends on the goals of the survey. If the survey samples households and does not follow movers, then a complete turnover in the household occupants would indicate the birth of a new unit. If housing locations are sampled, then such a turnover would not constitute a death as long as the hosing unit remains occupied. The death of a member of the household, or event he head, does not constitute death of the unit for a household-based sample, but a divorce or separation often will be defined as termination of the unit. If an individual respondent leaves the sample, the reason for the departure should be determined. If the respondent has died, then the individual record should be terminated. However, if the respondent leaves the sampling frame for other reasons (e.g., entering the military or moving abroad), it is possible that he or she may return during the life of the panel, and the record should be retained.

Often the death of a unit can be determined by observation. For instance, when a housing unit is vacant or destroyed and the sample is location-based, termination of the record may be indicated. However, in other cases respondents must be queried regarding the status of the unit. If the unit of measurement is the household, occupants of the sample location must be asked whether they lived at the current address when the previous interview took place to determine whether they should be considered part of the sample. (Rules for this decision will vary between surveys.) If only part of the household has moved since the previous visit, it may be necessary to determine the reason for the departure to ascertain whether the movers remain in the sampling frame. In designs which do follow movers and which allow the formation of new households during the life of the sample, permanent departure of individuals to form new households will indicate the need to establish new household records. (See Chapter 4 for a fuller discussion of these issues.)

B. Changes Related to Respondents' Time in Sample

Varying sample participation is not the only change over time which complicates inference from longitudinal data. A number of factors related to the time respondents remain in sample may produce changes in survey measures which are independent of any substantive changes in the phenomena under investigation. These factors include variation over time in the rules for interviewing particular respondents and changes in

respondents' approach to the interview based on increased experience with the survey instrument as the sample matures.

1. Response Variability Due to Changes in Respondent

The manner in which a survey is administered may vary from respondent to respondent. "Proxy" interviews may be administered, in which adult household members complete interviews on behalf of younger respondents, or in which available household members supply data for other individuals in the household. (In some cases such proxies are restricted to household members who are not present, but, in other instances, one household member will supply personal data for all individuals in the household.) Respondent rules are also frequently needed for collecting household information if there is more than one respondent per household. A number of possibilities exist for respondent rules. For example, one respondent in a household may be selected to provide household data, while personal data is requested from each respondent individually. Alternatively, all respondents may be asked for household data. In the latter case, inconsistencies might be reconciled in the field, for instance, when respondents report conflicting details regarding a household crime incident. A computer edit, or a postweighting algorithm might also adjust for differences in reporting, when household measures are simply the sum of individual measures.

Respondent rules can affect longitudinal data over tim e. For instance, during a longitudinal survey, younger respondents may become eligible to complete an interview without proxy, and may begin to report information of which previous proxies are unaware. There is also evidence that household-respondent status may affect the manner in which personal data are reported, particularly if the two types of information requested are related. Biderman, Cantor, and Reiss (1982), for example, find that respondents who report household data also report higher levels of personal crime victimization than do respondents who do not report household data. They also find that, if the household respondent changed between interviews, levels of personal victimization for the affected persons would also change. The authors hypothesize that the initial battery of household victimization items serves as a warm-up for personal items and aids recall for household respondents.

If the household respondent is allowed to change across waves, then two effects should be anticipated. First, the quality of personal data reported by a given respondent is likely to change over time, depending on whether he or she serves as the household respondent. Second, different household members will vary in their knowledge of the relevant data, so the quality of household data may also be expected to change over time and thereby bias transition estimates.

There are some obvious remedies for these problems. First, proxy interviews should be minimized, recognizing that obtaining certain information directly from younger respondents may be inappropriate or that there maybe no other way to collect data for some respondents.

Surveys vary in their reliance on data collected by proxy (eg., about 60% for NCS, 40% for SIPP), and such a policy is likely to produce an improvement in data quality proportionate to the fraction of data currently collected in this manner. Second, care should be ta ken in assigning responsibility for answering questions about the household over time, either by consistently assigning this responsibility to the same respondent or by requesting these data of all respondents. The latter procedure minimizes the effect of an unavoidable change in household respondent and makes any respondent effect consistent across all waves however, due to mandated ceilings on response burden for federally sponsored data collections, the additional precision realized may not justify the substantial number of redundant questions which are required. It should also be noted that the reconciliation procedures or post-weighting that would be required may make such a strategy very difficult to use.

2. Panel Bias

A number of factors associated with respondents' time in sample may produce changes in survey measures over time and thereby complicate explanation. The impact of these factors has been described as a history effect, secular effect, maturation effect, rotation group bias, time-in-sample bias, or Heisenberg effect. These factors include the reactivity of respondents to survey measures, changes in the performance of the respondent role, the "conditioning" effect of multiple administrations of the survey instrument, the aging of the panel, interaction between interviewers and respondents, interviewers' perceptions of their role, and the correlation between variables of interest and the probability of response. Changes in survey measures due to such effects present a danger for bias in longitudinal estimation. Consequently it is important to consider the influence of such factors when designing a longitudinal survey and to minimize the potential for such changes. This is a difficult task, because the reasons for the phenomenon are not clearly understood.

Ideally, the process of measurement should itself produce no change in the phenomenon under investigation. Research methodology in experimental psychology, for example, often involves disguising the purposes of research, so that the subject will produce the behavior under investigation with minimal "contamination" by the research procedure. In survey research, however, the respondent must not only understand the measures being collected but also must be led to appreciate the purposes and value of the research if response rates are to remain high. This is particularly important for longitudinal surveys, where retaining sample is a crucial goal Consequently the danger of reactivity between survey interviewing and the phenomena under investigation is a particular problem.

Researchers studying labor market experience, for example, have speculated that repeated interviews asking about job mobility might cause some of the mobility reported (Parnes:15). Questions about mobility may in fact cause subjects to consider the possibility and act upon it. National Crime Survey data also indicate that proportionately fewer crime incidents are reported in successive waves. This finding may stem from respondents' heightened awareness of vulnerability to crime, caused by participation in the NCS, which results in increased precautions taken against crime victimization. It has been suggested that respondents in a longitudinal sample might exhibit non-typical behavior Simply because repeated questioning regarding a topic may alter respondents' perceptions of the subject under investigation and change their behavior or attitudes accordingly.

For respondents no remain in sample, their responses can change over tim e solely as a function of longevity in the panel These temporal variations in response have implications for the quality of longitudinal data which are often unpredictable. In some cases, the quality of data may improve over time. Respondents may understand the respondent role better with repeated interviewing or pay greater attention on a day-today basis to the experiences being measured, with a consequent improvement in the richness or accuracy of the data gathered. Alternatively, if respondents or interviewers find the interview tedious or burdensome, they may become less enthusiastic about the

task over successive waves and avoid or give incomplete responses to survey items. One aspect of such a decline in data quality is the possibility that respondents may be "conditioned" by their participation over several waves to provide answers which produce artifactual changes over time. For instance, respondents may learn that a particular response will trigger a long battery of questions, which they may prefer to avoid in the future.

This is one alternative explanation for the decline in the rate of crime victimization reported in the NCS over successive waves. Respondents may learn that reporting a crime incident leads to an additional series of items for each incident reported, which results in a substantially longer interview. The Census Bureau's Current Population Survey (CPS), which is not strictly a longitudinal panel survey but which has many of the attributes of a longitudinal survey, exhibits a similar trend. Reporting unemployment triggers a battery of questions dealing with reasons for unemployment and activities directed towards looking for work. Reported unemployment invariably falls between the first and second waves of interviews in the CPS. This phenomenon in CPS could be related to several factors. One has to do with repeated interviewing and attrition. Williams and Mallows showed that, if the probability of response in a given save of interviewing was correlated with variables of interest, then, even with no change in the variables, a spurious change would occur.

The passage of time can also produce unintended change between observations because of gradual shifts in the meaning of questions and answers. Even when questionnaires are not changed, there may be evolution In the way respondents perceive or answer questions, which produces the appearance of movement (Parnes:14). This might be caused by events (including the survey itself), by maturation in the sample, or by non-response.

It is very difficult to determine whether a change across waves is real change or spurious change. Continuing validation research is necessary to identify panel bias in longitudinal data. Panel bias may be studied by comparing data collected in subsequent waves of a longitudinal survey to data collected in cross-sectional surveys (as in Cook & Alexander).

Although some conditioning or panel effects may be inevitable, several tactics can be used to minimize their impact. One option is to implement a rotating panel design to replace respondents after a predetermined number of interviews. This procedure affords two primary benefits. First, those respondents who have been in sample the longest are replaced with more "inexperienced" respondents. Second, the temporal overlap of old and new sample facilitates studies of time in sample effects. All respondents are administered the same instrument under the same conditions at the same time, which serves to test alternative hypotheses about panel effects.

Another possible means to attenuate or postpone the effects of panel bias is to minimize the respondent burden imposed by the interview. Careful construction of the instrument to minimize tedium and encourage respondent rapport should be central concerns in planning any survey but take on added importance in longitudinal data Collections, because of the need to sustain the active participation of respondents overepeated interviews. The overall length of the instrument may play a role in the respondents willingness to participate fully in successive contacts. However, design of the instrument to minimize tasks which the respondent is likely to find either tedious or particularly difficult is also an important consideration. Use of long follow-up batteries should also be minimized, to attenuate the effects of respondent conditioning.

C. Operations Change Over Time

Changes in the administration of a continuing survey are almost inevitable. Revisions to the instrument, redesign of the sample, introduction of new collection modes, and transfer of data collection responsibilities to another organization can all introduce changes in the data and compromise the validity of longitudinal comparisons. While a consistent time series may be difficult to maintain under such circumstances, means exist which allow the analyst to deal with the effects of such changes.

Eventually in most longitudinal research there is a pressure to change the survey measures in response to changing hypotheses. In addition, later findings frequently indicate a need for measures of new variables. Particularly when longitudinal research is exploratory and designed to identify significant correlates of change, researchers may be inclined to correct large a mounts of data to minimize future requirements for change in the questionnaire design. This aspect of longitudinal research may be costly, but it is an understandable precaution given the tendency for research hypotheses and/or policy-aims to change over time.

To accommodate changing methods, a survey may be run under old and new procedures simultaneously for a period of time, to allow comparisons between data collected before and after the change. Ideally, both old and new designs should be implemented at full sample, in effect twice the usual sample size, but budget constraints will often make this impractical The CPS has adopted this double-sample strategy to phase in new samples based on the 1980 Census. The CPI also used both old and new sample designs simultaneously for a six- month period in 1978, when the survey was revised.

Another strategy to consider when a questionnaire item is rewritten or a derived variable in a file is altered is to make changes in such a way that analysts may record the revised variable to correspond to the original variable (and vice versa), or to retain old questionnaire items in the revised instrument for some time. NCES adopted the latter strategy for the HS&B; survey when it adopted an "event history" approach to gathering employment and education data. In addition to the new items, the previous "Point in time" activity item was continued, allowing calibration of new items to the old and providing a degree of comparability between versions.

To reduce field costs, many sponsor agencies have approved designs which permit data collection by telephone after the first visit. NMCES and MNCUES, for example, used phone contacts for follow-up interviews. The available evidence suggests that such changes in mode may not produce uncontrollable fluctuations in the measures obtained. Benus (1975) notes that data collected by telephone and by personal visit for the Panel Survey of Income Dynamics (PSID) are quite similar. Groves and Kahn (1979) found overall that univariate distributions and bivariate relationships were not significantly different for 200 questions ad ministered by telephone and in person. However, they note that telephone interviews elicited more rounded financial figures, less detailed responses to open-ended questions and narrower distributions on some attitude items. They also indicate that respondents tend to perceive telephone interviews as longer than personal interviews of the same length. Findings that telephone respondents tend to give more "don't know" answers to filter questions triggering other questions may be related to this difference in perception of length. Telephone respondents may be more eager to bring the interview to a close. Consequently minimizing respondent burden seem s particularly crucial for interviews conducted by telephone.

While the research literature on the effects of interviewing mode on survey response is generally encouraging, there are enough examples of differences in respondent behavior to indicate that a mixed mode design should not be implemented without adequate pretesting and analysis of the effects. One danger is that a particular questionnaire design or questions about a certain subject area might trigger mode-related differences in respondent behavior. To facilitate measurement of such mode-related response variability, it is desirable to design shifts in mode of data collection so that the changes across waves are systematic, making the effects measurable. It is also important in surveys which do not require interviews with all household members to ensure that interviews are obtained from the same household members when the interviewing mode varies across waves, as respondent availability may vary by mode.

In conclusion, prospective longitudinal surveys require administrative and operational features that are different in kind as well as degree from those in cross sectional research. The long-term analytical goals of the survey must be considered in planning every aspect of sample definition and weighting. Provisions should be made for validation studies to evaluate such factors as attrition and panel bias. Finally, changes in format, operations and staff must be anticipated and managed in ways that ensure the comparability of measures from wave to wave.

In practice it is worth noting that there are only a limited number of organizations which handle nearly all large-scale longitudinal surveys. Due to their experience, these organizations have a high level of expertise, and the continuity of experience contributes to successful planning and implementation. However, the concentration of longitudinal research in such a small number of organizations increases the impact that any errors, such as limitations in the sampling frames most commonly used, would have on the representativeness of longitudinal research.

D. Processing

While the measures collected in longitudinal research may be similar, to those collected in cross-sectional studies, there are special problems in controlling and interpreting them. The sheer size of the data files created in national longitudinal surveys creates special problems in processing and analysis. The massive files can be difficult, expensive, and slow to process, which has often limited their use to organizations with the staff, equipment, and often complex software capable of handling complex data sets. As a result, data analysis has typically lagged behind the accumulation of data (Kalachek:17). Fortunately, this situation is changing with the advent of public use files for multivariate analysis and with the dissemination of m ore user-friendly "statistical data base" packages to facilitate data management and analysis.

In processing data from longitudinal surveys, difficulties are encountered related to cross-wave case matching, cross-wave data revisions, and preparation of data files for analysis. Often there is no single "best" procedure for processing, because ease of processing and analytical requirements are not always compatible goals.

Errors in individual record files can cause multiple problems. Often items which should remain consistent across waves (e.g., race and sex) or which should change only in predictable ways (like age and marital status) will exhibit changes due to respondent confusion, transcription error by interviewers, or keypunching errors by processing staff. Detecting these errors is important, not only because such items often define key demographic variables for analysis, but because such items are frequently needed to match cases. Errors are also inevitably introduced when imputations are made for missing data.

Several procedures are possible to minimize errors. For SIPP, the field office staff immediately checks completed interviews to reconcile discrepancies, avoiding more costly correction of data after they have been keyed. Another possible procedure is to build computer edits into the processing system to detect inconsistencies between current and prior interviews. NLS-72 and HS&B; use machine edits to identify and resolve inconsistencies for about thirty critical items. Another option, utilized by CPI, is to create a machine-generated control card, which avoids errors in transcription and which provides interviewers with prior-wave data necessary to reconcile discrepancies in the field. This latter procedure, however, can also lead to reduced reporting of actual change.

1. Cross-Wave Matching

In order to link data across waves, variables must be created to match records at the desired unit of analysis. A number of data management issues must be addressed, including the consistency of linking variables across waves, providing for longitudinal matching at multiple levels of analysis, and rules for matching merged and split households.

If longitudinal records are not matched correctly between waves, the effects can be similar to sample attrition or nonresponse. The records of one or more observations will be missing from a respondent's longitudinal file, giving the appearance of missing interviews. One possible consequence of matching errors is error in analysis, either because incomplete records are deleted, or because missing data are imputed. If records are linked incorrectly, longitudinal data are also likely to produce flawed results by showing false changes in status. Even cross-sectional analyses may be in error, if control card information or data from previous interviews are carried over onto the improperly matched record by the processing system.

A number of procedures are possible for linking units accurately from wave to wave, including matching of household and individual line numbers, or matching independent person and/or household identification numbers. Economy in the number of variables used for a match is generally a virtue, because the opportunity for mismatches due to transcription or coding errors increases with the number of variables used. So does the likelihood of missing data, which often results in the computer assigning a missing data code, which hampers matching. Limited redundancy in linking variables can, however, provide some protection against false matches, in that such cases are more likely to be flagged in the matching process.

Validation procedures to detect longitudinal mismatches should be incorporated into the processing system and can often rely on demographic variables which either should not change over time (e.g., race, sex, or date of birth) or which can be expected to change in predictable fashion (e.g., marital status or age). Such methods are particularly useful when person-level matching is performed using the assigned line number of respondents within household. It is also useful to imbed check digits in key linkage numbers, to detect miskeying. In addition to careful design of validation variables, immediate error checking by the field office of items important for matching and validation is likely to reduce the number of mismatches significantly.

Often, person records are linked across waves by matching on household ID and on the line number of an individual within the household record. This is usually cumbersome, and it makes linking individual data across waves extremely difficult if an individual moves out of the sampled household, if the household dissolves, or if the household merges with another household, all of which render the previously assigned household ID obsolete. Consequently, for surveys which are intended to follow individuals, regardless of the duration of their association with a sampled household or household location, assignment of an independent person ID is highly desirable. This is not to argue that ID is at other levels of observation are not useful, as longitudinal analysis at household, person, or event level is often needed. The important consideration is that linking variables be designed so that changes in sample composition do not prevent record matches.

SIPP has implemented an ID which, while complex, illustrates the sort of linkage which is often desirable. (Cf Jean & McArthur, 1984). The ID consists of:

PSU number - 3 digits

Segment number - 4 digits

Serial number - 2 digits

Address ID - 2 digits

Entry address ID - 2 digits

Person number - 2 digits

Household ID consists of address ID, PSU, segment, and serial numbers. The latter three numbers are fixed once assigned. The entry address ID also does not change. The first digit of the address ID indicates the wave at which the household was interviewed at that address. The second digit sequentially numbers, by address, households resulting from a split into two or more households by original sample persons. The first digit of the person number indicates the wave at which the respondent entered the sample, and the second two digits sequentially number persons within the household. This ID also remains fixed.

Linking households or individuals with the SIPP system is fairly straightforward. Households whose composition does not change require the household ID, and individuals require the household ID and person number to provide a match. The inclusion of a fixed entry address ID also facilitates matching records for individuals or households who move, and for split households. Combining the person number and the entry address ID provides a person number which remains constant regardless of changes in address and household composition. This provides a link to data collected for an individual across all waves, allows a match to the initial household, and permits the analyst to filter data for only the original survey respondents, if desired. This system remains adequate for multiple movers or for households which split a number of times.

In 1979 two waves of interviews from an ISDP panel were merged into a single longitudinal file using personal identification variables. Mismatching between records proved to be a significant problem, and there was evidence that additional matching errors were undetected (Kalton & Lepkowski:26). A second file was created using ID numbers rather than personal characteristics. This file had significantly fewer discrepancies during edit checks for such items as sex and age, indicating that fewer matching errors occurred with the use of the ID number for linking.

Sometimes the potential of longitudinal data has not been exploited because of the complexities involved in updating data with information collected in subsequent waves. For instance, a respondent may report a crime victimization or a health problem, but information on insurance coverage will remain incomplete, because the claim had not been settled at the time of the interview. It is frequently desirable to revise or add data during a later interview and to create an automated control system which would allow revision of the original record. One possibility is to provide a check item on the instrument for information which is frequently incomplete. The control system could then flag incomplete data during processing and direct the interviewer to follow up on this question in a later wave. Similar procedures were used in N M C E S and N M C U E S, which allowed validation of data collected on health care payments and insurance coverage during later interviews.

Revising files obviously creates some complications, and there are trade-offs between ease of processing and ease of analyzing the revised records. One of the simplest procedures for processing is to reserve a field for follow-up data in the interview along with an incident or event ID which allows a match to the original record. This procedure unfortunately would make the analyst's task considerably more difficult, in that several files would have to be scanned to locate all updated material. The required matching and file restructuring routines would also be rather cumbersome and expensive to run, unless the data were released in a form compatible with a statistical data base which performed the matching. These complexities create potential for data management errors, particularly for inexperienced users accessing public use files.

The alternative is to correct the original records based on followup data and to release the updated files. A disadvantage of this procedure is that several versions of the same, file would be in circulation.* Nonetheless this procedure appears to have greater potential for facilitating straightforward analysis and management of the data, particularly if early versions of a file are labeled as "preliminary."

2. Data Structures to Facilitate Analysis

A number of strategies may be used to create longitudinal data files. One is to create, a separate fixed length record for each case at the smallest unit of analysis, with separate fields devoted to repeated measures of the same variable. Often this is not feasible, because this procedure entails a thorough revision of the file every time a new wave is completed. It is often preferable to produce a separate file for each completed wave or even more frequently if data collection extends over a lengthy period and to include in the files a number of linking variables which remain constant for each case across waves. Other than the size of the files produced, the main difference between these two approaches then is in the processing system adopted: The former produces Integrated longitudinal files, while the latter produces files resembling crow-sectional data sets which allow the analyst to link the records later.

Producing a file which uses the smallest unit of observation as the basis for a record is often not the most efficient structure for a data set. A number of surveys

---

*This is not as serious a problem for longitudinal files, the latest version of which can more easily be identified, as it is for cross-sectional files created from a particular wave.

collect data on households, individuals within households, and discrete events experienced by the household in aggregate or by individual members. Given the implicit "nesting" of such data, creating a file based on the smallest unit will result in much redundant information for higher level units. The number of events recorded and the number of household members may also be expected to vary between households, and variable length records will result, necessitating extensive "padding" to create a rectangular file.

A more efficient strategy in such cases is to produce hierarchical files with the data pertaining to each level of observation appearing in separate records and with variables appearing in more than one type of record to allow for linkage across levels. A number of software packages such as SAS and OSIRIS now exist which can process and analyze such files. In addition, a number of "statistical data base" packages are available, such as SIR, Canada's RAPID, and Mathematical Policy Research's R A MIS, which provide sophisticated capabilities for matching across waves and levels, and which thereby simplify the analyst's data management tasks in working with longitudinal files.

Decisions regarding the optimum structure for a longitudinal file also need to take into account the expected size of files. Limits on the number of records many soft ware packages can process may be exceeded by the size of large federal data collections. Consequently, file structure options for facilitating analysis of longitudinal data may be constrained. Sponsors may find it necessary either to forego compatibility with some otherwise useful software packages or to release subsets of their data to provide compatibility with a wider range of software packages.

3. Confidentiality

Processing operations and data structures for analysis cannot be designed solely to reduce costs, complexity, or bias. They must also protect respondent privacy as far as possible. This is sometimes not compatible with maximum efficiency. Procedures for protecting confidentiality of paper records and of tape records must be thought through carefully.

The problem of maintaining respondent confidentiality is more difficult in longitudinal surveys than in cross-sectional surveys. In cross-sectional research, the confidentiality of a response can be protected by stripping responses of identifiers at an early stage in processing. In longitudinal surveys, response records must be linked to personal identifiers, sometimes for decades, until data collection and analysis are complete. Longitudinal records commonly contain multiple identifiers in order to facilitate tracing and to ensure that records can be matched after each wave, regardless of missing data. Name, address and Social Security number are often augmented with the name and address of family, neighbors, or friends who are to be contacted in tracing respondents who have moved. The large number of identifiers, plus their dispersion across records and across time, makes protecting confidentiality in a longitudinal survey far more difficult than in cross-sectional research. However, most research organizations have learned over the years how to protect paper records.

An illustration of one solution to problem is that adopted by N C ES for the NLS-72 and HS & B: Identifiers are stripped from the tape prepared by the contractor before it is turned over to the sponsor agency. These data are maintained by the contractor but may only be used with the explicit approval of the sponsor. The procedure provides a complicated, layered procedure which inhibits any unauthorized access by sponsor, contractor, or public users and provides protection similar to that of a cross-

33

sectional study.

This example illustrates a number of the basic safeguards which should be integrated into any longitudinal data collection effort. First, identifiers should be used only to maintain the quality of the data, e.g., for tracing respondents or for matching purposes. Second, only staff performing these functions should be allowed access. Hardcopy media containing identifiable data should be stored in a secured area to limit access. Electronic files should be similarly secured and, when in use, access should be restricted by the operating system to authorized processing personnel only. Third, all privacy- relevant data should be stripped from public use tapes before release. Ideally, the collection agency should separate identifiers during processing and store them on a file separate from the substantive data. Finally, when data Section is complete, all copies of identifiers should be destroyed. Even when such measures are taken, agencies and research organizations must consider the possibility of confidentiality breaks. The quantity of information available about respondents creates the possibility that a series of rare responses can identify respondents. Current research in confidentiality is addressing this problem and should provide useful guidelines for enhanced security measures in the near future.

CHAPTER 4

SAMPLE DESIGN AND ESTIMATION

There are many issues in the design and estimation strategies for longitudinal surveys that are identical to those for cross-sectional surveys. Some issues, however, such as weighting and compensating for nonresponse become more complicated with a longitudinal survey. Usually the complications arise because of the changing nature of the population, as discussed in Chapter 3. In this chapter, we discuss some of the major design and estimation problems, many of which need more research.

A. Defining a Longitudinal Universe

Defining the initial study universe for a longitudinal survey is no more complicated than defining the universe for a cross-sectional study, The initial universe is fixed at a specific point in time and is explicitly d fined. Sample units can be selected and the only difficulties are related to the sampling frame itself. Time, however, gradually complicates the problem of defining a longitudinal universe.

The study universe usually does not remain constant over the period of the longitudinal survey, as was discussed earlier., The universe of individuals, households, families, or establishments changes over time. If a universe changes slowly along the critical dimensions of the survey, the problem of a longitudinal universe definition may be ignored. However, if changes in the universe over time are not trivial, a static universe definition may not be sufficient. The choice of definition for the longitudinal universe will have a direct effect on data collection and analysis.

Judkins et al (1984) describe three methods for defining a longitudinal universe. These ideas are generalizable to any longitudinal study of persons or other units. One method for defining a longitudinal universe is to select a specific time during the course of the study as the point that defines the universe. If the universe is defined at the time of sample selection, it is called a cohort study. Units in the sample are defined at the time of the first interview. At later waves of interviewing, data need be collected only from these units. All inferences and estimates refer only to the universe in existence at the time of the first interview. For example, for the CPI commodities and service sector, the universe is a set of cohort samples with attrition due to deaths. Births are introduced only when an entire cohort is replaced with a new sample.

Principal Authors: Daniel Kasprzyk and Lawrence R. Ernst

The longitudinal universe may also be defined at a time other than the time of sample selection. Under both scenarios, statistical, operational and methodological problems may arise because the sample was selected at one point in time and the analyses of the study universe reflect a different point in time. It is possible that elements of the study universe at the time of sample selection are no longer part of the longitudinal universe; it is also probable that elements of the longitudinal universe which exist at the time of definition were not in existence at the time the sample was drawn. This creates an operational problem -- whether to collect data from these "entrants" to the longitudinal universe -- and it creates a statistical issue, the development of estimation methods for this universe. For example, in the SIPP universe (the non-institutional population, and members of the military not living in barracks) individuals may leave the universe by moving outside the United States, to an institution, to military barracks, or by dying. At any time during the study period persons may enter the SIPP universe by returning from overseas, institutions, or military barracks, or through birth.

A second method of defining a longitudinal universe extends the first method by looking at more than one time point. Several time points are selected, each one defining a universe at that time. Then the entire set of units -defined by these different cross-sectional universes is included in the longitudinal universe. Thus, if a person entered a sample household by being born or returning from overseas sometime after the initial interview, that person would be included in the longitudinal universe. People can be added to the universe, and anyone who is in the universe for any of the time periods should be included in the estimation.

For analysis of aggregations of persons, such as households and families, some identification of aggregations at each time point is necessary. Since these aggregations can and do change over time, conceptual, operational and statistical difficulties occur. See, for further discussion of this subject, the section on units of analysis in this chapter. This approach, however laden with difficulties, is the approach which best captures the dynamics of the longitudinal universe.

The third method for defining the longitudinal universe is also an extension of the first method, but instead of including all units that enter, leave or stay, this approach includes only those that are common to all the selected time periods. In this approach, one includes in the definition of the longitudinal universe only those elements which were members of all cross-sectional universes. This definition leads to a static universe containing only those elements which do not enter and exit the universe. For example, for households, families, and establishments the universe contains only those units in existence throughout the entire survey period.

As discussed above, defining the longitudinal universe can be a problem when it contains units which enter and leave the cross-sectional universe. When the units are establishments or a group of individuals, some decision concerning "rules of continuity" is necessary. The next section briefly reviews models for longitudinal household (family) units of analysis.

Units of Analysis

Aggregations of persons, such as households and families, present difficult conceptual and practical problems in longitudinal surveys. Over time individuals enter and leave households, and set up new households. It is no longer obvious how a household or family should be defined when time becomes an integral part of the definition. McMillen and Herriot (1985) attempt to reduce the possible definitions to a reasonable number, in order to conduct an empirical evaluation of alternative concepts. They also provide a brief review of the historical basis for a longitudinal definition of households. Much of the discussion below is based on the McMillen and Herriot (1985) paper and one by Kasprzyk and Kalton (1983).

Three models have been used to describe household and/or families over time:

1) a static model; 2) an attribute model; and 3) a dynamic _model. The static model of households (or families) classifies households at one point in time, and reflects a cross-sectional perspective. Households and their members are defined at one point and individual characteristics are aggregated over the survey

period to provide summary statistics for aggregated analysis units. A critical, but false, assumption has to be made that the household composition remains fixed during the survey period. This definition is not truly longitudinal, because it ignores any changes that each unit may undergo. In this approach weighting the so-called longitudinal sample corresponds to weighting the cross-sectional sample. Note, however, that for CPI or any Laspeyres type index the assumption of fixed composition is what is desired, since the change in composition of sales is being held constant so that price change is the only thing measured.

The second model for defining households or families over time is the attribute model. In this model, the individual is the unit of analysis, and household and family characteristics are treated as individual attributes. As a result, the problem of changing units over time is avoided. Results under this approach are expressed as "X% of persons live in households with attribute "Y", rather than "X% of households have attribute Y." Household characteristics are, therefore, attributes of the individual. The attribute model has been used extensively by the Survey Research Center of the University of Michigan for the analysis of data from the University of Michigan's Panel Study of Income Dynamics.

Dynamic models, the third type, represent the most difficult conceptual and operational problems. In these models, households (or other groups of individuals) are defined over time, not at one point in time, by a set of rules. These rules, often referred to as continuity rules, identify the initiation, continuation, and termination of the analytic unit. Three examples of continuity rules which have been proposed as dynamic definitions of households are presented in McMillen and Herriott (1985). It is not obvious that one set of rules is better than others; in fact, one concept may be more useful for certain kinds of analyses, but not for others. Little empirical work using alternative dynamic concepts has been published, although Citro (1985) has recent begun an investigation using data from the SIPP development program it remains to be seen whether the dynamic concepts can be properly interpreted and employed to provide useful results for policy application.

C. Sample Design

For a longitudinal study with a static population, that is, one in which there are no additions over time, the need for longitudinal estimates presents no special difficulties in sample selection. It is only necessary to choose a single sample at the selected point in time, as if a one-time survey were being conducted, and then follow the sample units initially chosen. For such a study there is, in general, no ambiguity about the analytic units, and no additions are permitted to the population. The longitudinal studies of the National Center for Education Statistics (NCES) are examples of this approach.

The populations for all the other longitudinal surveys described in this report are dynamic in nature. For these surveys initial sample, selection presents no particular difficulties. It is only necessary that each unit in the population at the time the initial sample is chosen have a known probability of selection. Complications arise, however, because of the additions to the universe, and the care that must be taken in order to follow the sample units of analysis over time.

Ideally, provision should be made at the design stage to give additions to the universe a chance of entering the sample, or, failing that, to make adjustments for their absence at the estimation stage. For SIPP, Employment Cost Index (ECI) and items in the CPI for which the Point of Purchase Survey (POPS) is the source, the problem of new units is partially alleviated by employing a rotating panel design. Thus, all additions to the universe will eventually be given a chance of selection, with the length of time between panels as the maximum lag. For the ECI and the CPI, because of the difficulty of identifying births quickly, this is the only provision made for additions at either the design or estimation stage. In general, additions to the universe in these surveys have no chance of affecting the estimates until the selection of the next sample or panel. This again is consistent with the Lespayres concept of a fixed set of items and outlets for Measuring price Change only.

In contrast to the ECI and the CPI, the designs of NMCES, NMCUES and SIPP give individuals, families and households that are additions a chance of selection as soon as they enter the universe. At each round of interviewing in these surveys not only is the initial sample interviewed, but so are all individuals currently residing in a household with the original sample people. Individuals joining the universe and moving into a household containing at least one person who was in the universe when the initial sample (or most recent sample) was chosen have a chance of entering the sample. So does any family or household joining the universe that contains at least one individual who was in the universe when the initial sample was chosen. Other individuals, families and households that join the universe have no chance of selection. To cite another example, the CPI rent survey samples building permits in order to identify new units quickly.

Care must be taken in the design of longitudinal surveys to assure that the analytic units used in the estimation process for a specific time interval are followed throughout that time interval. In general, this is not a serious problem with surveys such as the ECI and CPI, since the definitions of analytic

units for these surveys generally include a fixed location such as an item at a specific outlet. Furthermore, in cohort studies such as the High School Class of 1972 which only makes estimates for individuals selected in the initial sample, there are no difficulties other than the operational problems associated with following people. However, for NMCES, NMCUES, and SIPP there are difficulties associated with following certain sample analytic units.

A key reason for these difficulties is that a household or family may continue to exist under most longitudinal definitions even though it no longer contains any individuals who were initially in the sample. Under the procedures established for each of these surveys, the household or family will no longer be followed. Ernst, Hubble, and Judkins (1984) discuss this problem in detail. Any individuals who are additions to the universe and who are to be used in the estimation process should also be followed. Provisions were made to do this in T#ICES and NMCUES but not in SIPP. Tn fact, it has not been decided whether additions will be used at all in SIPP for longitudinal person estimation. Judkins et al (1984) discuss this question.

D. Weighting

There may be several stages of weighting a sample. One is to reflect the original universe; another is to adjust for nonresponse; a third may be to adjust for sample coverage. Longitudinal surveys have the usual weighting problems of cross-sectional surveys and then at least one additional problem. That is to provide a longitudinal weight to be used during analysis. In this section, we discuss the simple unbiased weighting and adjustment to independent estimates. Nonresponse, since it can be handled either by weighting or imputation, is deferred to the next section.

I. Unbiased Weights

Typically, the unbiased or base weight for a sample unit is the reciprocal of its probability of selection. In longitudinal surveys, this has generally been the weight assigned to sample units which were in the universe at the time the sample was selected.

The development of base weights becomes more complicated-in surveys such as NTICES, WCUES, and SIPP which incorporate additions to the universe in the estimation process, since it is often not practical to compute selection probabilities for such analytic units. For example, NMCES and NMCUES families which are additions to the universe will generally be used in the estimation process if, and only if, at least one member of the new family had been a member of a sample family during the first round of interviews. It would be extremely difficult to determine the first round families for all the members in the new family, and then compute the probability that at least one of the first round families could have been selected. Fortunately, it is not necessary to know the probability of selection in order to obtain base weights which yield unbiased estimators. See Ernst, Hubble and Judkins (1984) for a description of this methodology.

Several longitudinal weighting procedures will now be described. Since most of them will be defined in terms of cross-sectional weights, it is useful to define what is meant by the cross-sectional weight. The first round cross sectional weight for a sample household is taken here to be the reciprocal of the probability of selection. For all nonsample households in the universe this weight is zero. For any time period after the first interview it is defined to be the mean of the first round cross-sectional household weights for all persons in the household who were in the universe during the first interview. This type of weighting procedure is currently used in SIPP to produce cross-sectional household and family estimates.

There appear to be only two precedents for the weighting of longitudinal households and families -- NMCES and NMCUES. For these surveys each family was assigned its cross-sectional weight at the date the family was first formed (See Whitmore, Cox, and Folsom (1982)). The only other survey where serious consideration is being given to the longitudinal household estimation issue is SIPP. Five alternative methods for obtaining unbiased longitudinal weights are discussed in Ernst, Hubble, and Judkins (1984):

1. The NMCES/NMCUES procedure, assigning each longitudinal household (family) its cross-sectional weight at the date the household (family) was first formed.

2. For any time interval, assigning each longitudinal

household (family) its cross-sectional weight at the beginning of the time interval.

3. For any time interval, assigning each longitudinal household weight the average of the first round weights for all persons who remain members of the household throughout the time interval. If there are no such people, the longitudinal household weight is zero. This procedure generally has a slight bias.

4. For any time interval, assigning each longitudinal household the average of its monthly cross-sectional weights.

5. If a longitudinal household is defined as an attribute of a specific individual, such as the householder or principal person, then assigning the longitudinal

household the first round weight for that specific individual.

The procedures listed apply to the restricted universe of all households in existence throughout the time interval of interest. Some modifications are necessary to apply these procedures to the unrestricted universe of all households in existence for a portion of the time interval of interest. There are advantages and disadvantages to each procedure. They differ, for example, in their need for data from longitudinal households which no longer contain any first round sample persons, or their need to ask retrospective questions in order to determine the appropriate weights.

Finally, we briefly discuss longitudinal person estimation. NMCES and NMCES employ longitudinal person estimation that incorporates additions to the universe. Each additional person is associated with a first round family

40

and then assigned the first round weight of that family. For SIPP, it has not been decided whether individuals who are additions to the universe will be used in the person estimation process or, if so, how they would be weighted. One procedure being considered is to assign to persons who join the universe the cross-sectional weight of the household that they are a member of at the time they join the universe.

2. Adjustments to Independent Estimates

As a final step in the weighting process for several longitudinal demographic surveys, the population is partitioned into demographic groups and individual weights are adjusted so that the sample estimates of the demographic subpopulations agree with independently derived estimates. In general, this estimation step reduces sampling variability and biases resulting from undercoverage.

In the National Longitudinal Surveys (NLS) this adjustment was done for age-race-sex groups for the time of initial sample selection. The adjusted estimates of totals for each group were made to agree with independently derived Bureau of the Census estimates. The Census estimates are obtained by carrying forward the most recent census data to take account of subsequent aging of the population, mortality and migration between the United States and other countries. Since the same sample cases are followed throughout the life of the survey, no subsequent adjustments to independent estimates were made with the following exception: an annual adjustment was made for the cohort of young men (ages 14-29 in 1966) to maintain agreement with the independent estimates. This adjustment corrects population underestimates for men who were not represented in the original sample because they were in the Armed Forces at the time the sample was selected and who subsequently returned to the civilian population.

For annual data files from NMCES and NMCUES, family weights were adjusted so that the estimated number of families existing as of March 15 of the interview year agreed with counts from the March Current Population Survey. For each demographic group the adjustment factor used for sample families in existence on March 15 was also applied to families that did not exist on this date. This was done with the assumption that the rate of undercoverage and nonresponse was the sane for all families in a demographic group, irrespective of whether or not the families existed on March 15. Details of this procedure are given in Whitmore, Cox and Folsom (1982).

For person estimation in the NMCES' and NMCUES' annual data files, the adjusted family weights for each sample individual's first round family were further adjusted separately for each individual to produce agreement with independently derived age-race-sex estimates. The adjustment factor applied to each sample individual in a group was such that the average of the adjusted-sample estimates of numbers of individuals in each group at four times during the year agreed with the average of the independent estimates at the same four times. Details are provided by Jones (1982).

No decision has been made yet on how longitudinal weights for

SIPP will be adjusted to agree with independent estimates. One possibility is to use procedures similar to the NMCES and NMCUES procedures. A potential drawback to that approach is that survey estimates will agree with the independent

estimates at only one point in time. If agreement is required at other points in a time interval, then adjustment procedures could be modified so that the adjustment factor is not the same for each sample unit of analysis within a demographic group, but instead is also a function of the starting and ending date of that sample unit. This modified approach to adjustment has several disadvantages, such as possibly requiring some weighting factors to be very large.

E. Nonresponse In A Panel Survey

Nonresponse in longitudinal surveys can be treated from either the cross-sectional or longitudinal perspective. References concerning the treatment of nonresponse in panel surveys are in Kalton, Kasprzyk and Santos (1980), Kalton, Lepkowski and Santos (1981), Kalton and Lepkowski (1983), Marini, Olsen and Rubin (1980), David, Little, and McMillen (1983), Little (1984, 1985). Assuming the data, requirements for the survey mandate a longitudinal analysis, then the longitudinal perspective is clearly the more desirable, since it reflects the survey design.

If nonresponse in a longitudinal survey is treated from a cross-sectional perspective, each wave is treated as a separate survey. This has practical advantages in that the release of wave data may occur more quickly than if the separate waves were first linked, and linkage problems resolved. A disadvantage is that records with imputed data will be inconsistent from wave to wave because data processing and estimation procedures are implemented independently from one time to the next. Despite the inconsistencies at the micro-record level, changes in aggregates from the wave to another can be investigated. From a longitudinal perspective, nonresponse in a longitudinal survey is viewed not as nonresponse in a set of unrelated observations but as nonresponse in a set of variables with some logical dependency between two or more points in time. For example, in the CPI missing prices at time t are imputed based on prices obtained at time t-1, and on current average price movement for the item. This view adds considerable information to the data set for the treatment of nonresponse. However, it raises issues concerning the treatment of nonresponse which have not been addressed from the cross-sectional perspective.

Longitudinal surveys can be treated as cross-sectional to generate point-in-time estimates. Because of the repeated interviews, however, indicator variables can measure status over time, thus providing better information on patterns of behavior, transitions from one state to another, and the length of time in a particular status. The importance of obtaining this kind of information justifies linking the waves as quickly as possible and treating nonresponse from a longitudinal perspective.

The treatment of nonresponse in longitudinal surveys is in many ways no different then in cross-sectional surveys. The above discussion attempts to provide some indication of the similarities and differences in the two approaches. The time dimension adds a level of complexity for all decisions related to the treatment of nonresponse. First is the problem of longitudinal data base construction; efforts need to be made to construct longitudinal files which allow analysts to use the panel aspect of the survey. This includes, at a minimum, ensuring that sample units in one wave are linked to sample units in other waves and that critical data items remain consistent from

one interview to the next. Second is the problem of selecting imputation or weighting to handle nonresponse on one or more waves. Third is the problem of timing for release of data. Cross-sectional imputation offers the practical convenience of releasing data as soon as each wave's data are available. However, not all data useful for good imputation are available this way. Imputed values are likely to be better when a combined data set is used.

Fourth, in spite of the fact that longitudinal imputation is frequently more effective than cross-sectional imputation, a back-up system is necessary to handle cases where values needed for longitudinal imputation are missing.*

1. Types of Nonresponse

Three types of nonresponse occur in surveys: noncoverage, unit nonresponse, and item nonresponse. Noncoverage is the failure to include some units of the survey population in the sampling frame, which means they h ave no chance of appearing in the sample. This may occur, for example, because of incomplete listings at the final stage of selection. Unit nonresponse occurs when no information is collected from the designated sample unit. It can occur because of a refusal, because of a failure to contact the unit (no one at home), or because the unit is unable to cooperate (language difficulties).

Item nonresponse occurs when a unit participates in a survey, but does not provide answers to all the questions. It may occur because:

1. the respondent does not know the answer to the questions;

2. the respondent refuses to answer the questions;

3. the interviewer fails to ask or record the answer to the question;

4. the response is rejected during an edit check (e.g because it is inconsistent with another response.

The distinction between noncoverage and total and item nonresponse is important because it affects the type of compensation procedure adopted. With noncoverage, the survey can provide no information other than that

---

* The following sections describe imputation and reweighting to handle item and unit nonresponse in connection with improving finite population estimates. Imputation and reweighting strategies are not used, however, when estimating mathematical models of an underlying random mechanism or process. Since such analyses focus on estimation of model parameters, neither assigning values to individual cases nor adjusting to independent estimates is appropriate. Instead, methods of model estimation are used to account for the missing data under the assumption that the same model applies to all sample cases, even though some cases provide more complete histories than others. Model estimation by the method of

maximum likelihood is the most common approach (Tuma and Hannan (1984), chapter 5). The contribution of each sample case to the likelihood function is derived; and if the observations are statistically independent, then the likelihood function is, in most cases, the product of the individual contributions.

available on the sample frame. Compensating for noncoverage is usually carried out by using sources external to the survey to produce some form of weighting adjustment, as described in the last session.

Noncoverage in a longitudinal survey can be problematic depending on the population which is to be measured. If the population is approximately static, (that is, the amount of change in the population over the life of the panel is not substantial), then the treatment of noncoverage from the longitudinal perspective is not any different than from the cross-sectional perspective. To be precise, however, changes in the survey population should be reflected in later waves of the panel. Often this does not occur because of operational reasons or because such a small proportion of the population is involved.

For example, in SIPP the person population does not change greatly over the life of a panel. The principal changes are children who reach adulthood during the life of the panel, deaths, immigrants, emigrants, and persons returning from military barracks and institutions. The survey design captures information about new adults, deaths, and emigrants; however the design does not cover new entrants to the population who live in households which do not include adults eligible for initial sample selection, such as households in which all members are from the following sectors:

1. U.S. citizens returning from abroad;

2. immigrants who move into the U.S. after the first wave of interviewing; and

3. persons who return from military barracks or institutions.

The different approaches suggested for treating total and item nonresponse illustrate a concern for the kind and amount of data available for use in compensation procedures. Total nonresponse is typically treated by some form of weighting adjustment, using data available from the sample frame in addition to observations obtained by the interviewer. With item nonresponse, the responses to other survey questions may provide information. To use other responses effectively, item nonresponse is usually treated with some form of imputation (that is, by assigning values for missing responses based on responses from respondents with similar characteristics) rather than with weighting procedures.

From the longitudinal perspective, the issue of unit and item nonresponse is not very well defined. From this perspective, a unit's record consists of all information collected on the unit over the life of the panel. This suggests, however, that data missing for one or more waves of a panel can, in fact, be treated as item nonresponse. Nonresponse on one or more waves of the panel may logically,be treated as item nonresponse for all variables that should have been recorded for that wave(s). The distinction between unit and item nonresponse is not obvious, and, often, in the interest of simplicity, a judgment must be made identifying the appropriate level of response necessary to treat a case for item nonresponse rather than unit nonresponse. Ultimately, these issues are best resolved after empirical research on the nature, extent, and patterns of the missing information. This, along with knowledge of the uses of the data, will help determine a strategy for handling nonresponse in a panel survey.

2. Total Nonresponse

Total nonresponse in a cross-sectional survey means that no one at the household responded for one reason or another. It is often called unit nonresponse in cross-sectional surveys. It is generally handled by weighting adjustments, using data available on the sample frame such as region, city, block, type of area; or available from interviewer observation, such as race of householder. Usually the data available for weighting adjustment is quite limited.

In a longitudinal survey the concept of total nonresponse can take on a different meaning, including units which provided information for some, but not all, of the waves of the panel. Thus, viewing the entire longitudinal record as complete response, and responses at one or more waves as partial responses, the definition of total nonresponse can be reconstructed to include units which participate in the survey some part of the time. These units, despite having provided more data than "true" total nonrespondents, can be treated as total nonrespondents. In NMCUES, for example, total nonresponse is defined to include units (individuals), responding in fewer than one-third of the waves they were eligible for interview. (See Cox and Bonham, 1983,, and Cox and Cohen, 1985).

3. Unit Nonresponse

For the purpose of this discussion, unit nonresponse will refer to individual or person nonresponse to one or more interviews in a longitudinal survey. The length of a longitudinal survey increases a) the amount of data available for nonresponse adjustments and b) the complexity of nonresponse compensation procedures. Each individual's microdata record does not consist of unrelated, independent observations taken at different points in time, even though the data may be collected in that manner. Many variables reflect the same measure at different points in time. The status of a variable, such as income, at one point is frequently related to its status at a previous point. In a cross-sectional survey only two response categories exist, response and no response. In a longitudinal survey of n-waves there exist 2n possible patterns of response. For example, in a 3 wave study there are eight possible response patterns illustrated as follows (where NR refers to nonresponse and R refers to response):

1. R R R
2. R R NR
3. R NR R
4. NR R R
5. R NR NR
6. NR NR R
7. NR R NR
8. NR NR NR

Response patterns are usually classified as forming a "nested" pattern of nonresponses (i.e., variables from early waves of the survey are observed more often than variables from later waves), or as "non-nested". Attrition is a form of nested nonresponse, and estimators for dealing with nested nonresponse have been discussed in the incomplete data literature. (See Anderson (1957).: Rubin (1974), or Marini, Olsen and Rubin (1980).)

The three wave study example illustrates the kind of difficulty which can occur when one or more waves of data are missing. Case 1 is an example of total response -- an interview is obtained in each wave of the panel. Cases 2 and 5 illustrate attrition and nested nonresponse. Cases 3 and 4 illustrate non-nested patterns of response (two out of three interviews obtained) and cases 6 and 7 illustrate different non-nested patterns of nonresponse with only one of three interviews obtained. Case 8 is an example of total nonresponse. The difficult decisions about nonresponse which must be made for a three wave study are indicative of problems with surveys of more than three waves.

One way of treating unit nonresponse in a panel survey is to define the level of response necessary for a unit to be considered a "responding" unit. All units which exceed this response level would be treated as if they were present in all waves of the panel and their missing interview data regarded as a form of item nonresponse; units with a response level less than the standard would be treated like total nonresponse.

Underlying these alternative strategies for handling wave nonresponse is the issue of whether it is better to use imputation or weighting to adjust for wave nonresponse. The weighting procedure simultaneously compensates for all data items of a nonrespondent, but reduces the sample size available for analysis. Weighting adjustment procedures also typically incorporate many fewer control variables than an imputation procedure, although David and Little (1983) suggest a model based approach which increases the number of variables used in the adjustment.

Imputation, whether it be cross-sectional or longitudinal, fabricates data. The uninitiated user may not understand this and may attribute greater precision to the estimates than is warranted. Imputation techniques by their nature may fail to retain a covariance structure of the data. However, by identifying critical data items in advance, an imputation procedure can be developed to control key covariances. In practice, a two fold strategy of using both weighting and imputation procedures may often be the best solution (David and Little (1983)). A more detailed discussion of the weighting versus imputation issue for wave nonresponse can be found in Kalton (1985) and in Kalton, Lepkowski and Lin (1985).

4. Item Nonresponse

In the previous discussion it was noted that one way of treating unit nonresponse was to consider it a "form of item nonresponse" in a longitudinal record and use imputation techniques. That is, in a longitudinal survey, unit nonresponse can be treated conceptually as item nonresponse. Item nonresponse, because it typically refers to missing data item(s) in an otherwise completed interview, provides a good illustration of the fact that there is nothing theoretically special about longitudinal imputation. As Kalton, Lepkowski, and Santos (1981) have stated, longitudinal imputation for item nonresponse is simply imputation for item nonresponse using auxiliary data from a larger data base, including using longitudinal data elements as well as cross-sectional ones. The principal distinction is the availability of data which are highly correlated with the missing data, usually the same variable measured at different points in time. For example, the imputation in CPI is done from this perspective.

Theoretically, a decision concerning cross-sectional versus longitudinal imputation in a longitudinal survey is obvious. The longitudinal approach can certainly do no worse than the cross-section approach. The longitudinal approach can use any of the variables measured on a wave, but in addition it can use variables from other waves. As Kalton and Lepkowski (1983) point out,,if response on an item is highly correlated over time, then the value from a previous interview will be a good predictor of the missing value at the current interview.

Two exceptions to this statement should be noted: (1) the predictor variable must be reported at more than one point in time; and (2) the variables used in a cross-sectional imputation system are known to be poor predictors of the missing value and thus would likely be poor predictors in a longitudinal system. The two limitations are important because they suggest that empirical analysis of cross-wave data is necessary before developing a cross-wave imputation system. They also point out that in addition to an imputation system using two or more waves of data a fallback cross-sectional method is often needed to compensate for items which are missing in every wave of the panel.

Using cross-wave measures as auxiliary variables in an imputation scheme has special significance when individual changes will be analyzed. Obviously, if imputed values are assigned without conditioning on the previous wave's value, measures of change are very likely to be distorted. In this case, modeling state-to-state transitions becomes extremely important in developing an imputation system.

Some methods for longitudinal imputation Are discussed by Kalton and Lepkowski (1983). These methods make use of the stability a variable may have between successive waves of a panel, and they include:

1. direct substitution

2. cross-wave hot deck imputation

3. cross-wave hot deck imputations of change

4. deterministic imputations of change

A simulation to compare results using these 4 approaches is also described in the same source.

CHAPTER 5

LONGITUDINAL DATA ANALYSIS

INTRODUCTION

In the past, much longitudinal analysis has been done cross-sectionally, with each wave of a survey analyzed independently. The linked records were often difficult to use and discouraging to analysts. With improved data bases and the use of statistical techniques to analyzes transitions, trends, and change, longitudinal surveys are now showing their distinct analytical advantages.

A.   Determinants of Longitudinal Analysis Methods

Longitudinal analyses study,the change in some unit -- a person, a family, a business and so on -- over time. The focus is not on a description of the current status of the unit. Rather, interest is usually directed at the underlying process that determines any observed change.

The methods employed in the analysis of longitudinal surveys depend on four factors: (1) the nature of the process being studied, (2) the type of variables being measured, (3) the analytic objectives, and (4) the method of data collection. These factors taken together determine the kind of mathematical models of the process that are appropriate and estimable.

1. The Nature of the Process Being Studied

Many processes can be represented as the flow of a unit between some set of categories (states), such as the change in a person's employment status from employed to unemployed. Such a representation requires an enumeration of the possible categories and a probabilistic description of how movement takes place from one category to another. The flow of the process may be discrete or continuous in time. In a discrete time process, change of state occurs only at a fixed set of points. For example, eligibility for many government benefit programs is a discrete time process. Social Security Administration old age and disability programs, AFDC and many State welfare programs all pay monthly benefits. Eligibility for benefits changes only at discrete points in time, spaced one month apart. Other processes, such as change in health status, changes in price level, death, change in attitudes, or employment, can change state at any point in time and are therefore continuous in time.

The process under study may be time stationary or time nonstationary. A process is time stationary if its probabilistic structure and its governing parameters are not themselves changing over time. Processes which are not stationary in time are the most common. The payment of benefits under government programs often undergoes structural changes as the result of legislative and administrative actions. Morbidity is

Principal Author: Barry V. Bye

continuously affected by advances in medical science, and individual labor force decisions are in part determined by changes in the national economy.

2. The Type of Variables

A process may be described by variables which are discrete or continuous; and variables may be either observable or unobservable. Labor force status -- employed, not employed, out of the labor force -- is a discrete observable variable with three mutually exclusive and exhaustive states. Variables such as well being and satisfaction, on the other hand, are often taken to be continuous variables that are usually measured only imperfectly by a set of indicator variables.

3. The Analytic Objectives

The analysis of longitudinal data may have several objectives. Descriptive analyses are concerned with the regularities of the process under study. Such analyses often use cross tabulations at two or more points to show gross and net change of the units. There are other descriptive statistics: the number of times that a certain state has been entered since the last measurement, the average length of time spent in a given state, the distribution of probabilities for the next transition and the derivation of calendar period estimates not based on retrospective reports. Hypotheses tests often deal with differences in these statistics among several subpopulations.

Researchers interested in causal analyses tend to focus on the underlying structure which governs the process. Mathematical models of the transition from one state to the next become prominent in causal analyses, and the estimation of the parameters becomes the, primary statistical goal. The signs and statistical significance of the estimated parameters are usually interpreted in the context of some higher level generalization or theory.

Sometimes longitudinal analyses are designed to project a process into the future. Projection is of primary concern in evaluating changes in government programs or the results of field experiments, particularly When the full effects of the changes have not yet been realized. Projection usually requires a mathematical model of the process. The parameters are then estimated from longitudinal data.

4. The Method of Data Collection

Two major strategies are used in gathering longitudinal data. In the first approach a complete history of the process is obtained. This approach is the event history method. Measures include the sequences of states occupied by the individual units, and the times when changes in state occur. The second approach is the multi-wave method. Tn this approach the current status of the units is obtained at two or more points in time, but information is often lost on the duration and sequence of events, and on the possibility of multiple changes between measurements. Information on the duration of events may not even be collected in the multi-wave method. For example, at the initial interview, there may be no data concerning the initial status of the process. At the final interview, there are no data concerning the next state of the process.

To summarize, the appropriate data collection strategy for a longitudinal survey is chosen by assessing the nature of the process, the variables, and the research objectives. For example, structural analyses of discrete, observable processes will require event histories (see Tuma & Hannan, 1984). On the other hand unobservable variables such as attitudes can only be measured in a multi-wave panel context, because the best one can do is measure the current status at any fixed point. Multi-wave methods have been used in most large scale surveys even when the focus is on observable processes. The resulting logs of information often severely restricts the analyst's ability to recover the underlying parameters and to discriminate between competing mathematical models. (See Coleman, 1981, and Singer & Spilerman, 1976).

B. Analysis Strategies for Longitudinal Data

Many of the approaches that are used for the analysis of cross sectional data are applied to longitudinal data as well (see Dunteman and Peng, 1977). There are two ways to use longitudinal data in these analyses. In some cases, variables are measured repeatedly over time. In other cases, longitudinal data are used to establish the temporal sequence of a set of variables. Establishing the correct temporal sequence of a set of variables is important for assessing causal linkages within the set.

Categorical data are collected in longitudinal surveys as well as cross sectional surveys. These data can be arrayed in cross tabulations showing the relationship between antecedent and outcome variables. When the status of a particular variable is measured at more than one point in time, cross tabulations can be constructed that describe the change in status of the sample units over time. When longitudinal data are placed in cross tabular form, the statistical techniques used to analyze cross-sectional data may be applied. These contingency table analysis techniques include the general testing of hypotheses about the structure of the table (Landis & Koch), the use of log-linear modeling (Bishop et al, 1975, Dunteman & Peng, 1977, and Hauser, 1978), and the development of certain classes of latent structure models (Clogg, 1979).

In longitudinal studies where the outcome variable is continuous, a number of cross-sectional analysis models have been applied. These models fall within the realm of regression analysis and the analyses of variance and covariance. One of these methods, path analysis (see Blaylock, 1970), involves estimating a sequence of regression equations where all endogenous variables, ordered in time, are regressed upon all preceding variables. Path analysis methods have been extended by J™reskog and Sarbom (1979) to cases where the outcome and predictor variables are in principle unobservable (latent) and can only be measured imperfectly by a set of indicator variables. When such variables are measured at several points, J™reskog's methods can be used to determine whether the nature of the construct is changing over time and which predictor variables account for the changes.

While cross-sectional analysis is often adequate for describing changes in status and identifying determinants, these methods are usually unsuitable for the analysis of the underlying process that generated the data. Social processes are often better represented by discrete-state,

continuous-time stochastic models. The first step in constructing this kind of model is to specify rates of transition between states. A number of researchers (see-Coleman, 1981, Ginsberg, 1972a and 1072b, and Tuma, 1976) have shown that regression analysis -- usually specified in terms of linear or logistic equations with the outcome as the dependent variable -- can supply information about the rates of transition only for a severely limited class of models. In those cases where regression is useful, the process must have run a sufficiently long time that the observed proportions in the outcome categories are not themselves changing over time. Even when cross tabulations show status change between two (or more) points, model identification can be problematic. The data are often equally compatible with more than one model.

Because of the problems encountered when applying cross-sectional analysis methods to longitudinal data, current analysis strategies focus directly on the rates of transition from one state to the next. In the biological sciences these investigations fall under the rubric of survival analysis (see Elandt-Johnson & Johnson, 1980). In the social sciences, general theories of stochastic processes are applied (see Bartholemew, 1973 and Tuma & Hannan, 1984). While these new methods permit a richness of analysis not possible with cross-sectional methods, they can have a significant impact on sample design and data collection issues.

Many of the techniques require event history data rather than multiwave panel data. In those cases where only longitudinal data are obtainable, observations at unequally spaced survey dates are often required. Many of the new approaches utilize non-parametric methods or rely on maximum likelihood techniques for the estimation of model parameters. Applying these techniques properly to the complex sample designs found in longitudinal surveys remains a largely unexplored area in statistical research.

C. Examples of Longitudinal Analysis

Because there is such a wide variety of methods, the flavor of longitudinal analysis is best captured through examples. Two Social Security Administration projects will be discussed; the first is the Social Security Administration Retirement History Study (RHS). In this project some examples of the more familiar cross-sectional approaches are presented. The second is the Social Security Disability Program Work Incentive Experiments (WIE) which provide examples of some current analytic strategies.

1. Social Security Administration Retirement History Study

The Social Security Administration's Retirement History Study (RHS) is a multiwave survey designed to address a number of policy questions relating to the causes and consequences of retirement. Among these questions are: Why do individuals retire before age 65? How well does income in retirement replace preretirement earnings? What happens to the standard of living after retirement? The original sample of 12,549 persons was a multi-stage area probability sample selected from members of households in 19 retired rotation groups from the Current Population Survey. The sample was nationally representative of persons 58 through 63 years old in 1969. Initial interviews were conducted in the spring of

1969 and then in alternate years through 1979. Data collected during this period provide

detailed information on work history, sources of income, expenditures, health, and attitudes toward and expectations for retirement. Results from the RHS have been reported in a number of Social Security Administration research reports (listed in SSA publication #73-11700). The data have also been analyzed by researchers outside the government via public use tapes.

An interesting variety of cross-sectional analytic methods suitable for multi-wave data have been used with the RHS data. One example is a two-wave descriptive analysis of the change in income between 1968 and 1972 using simple turnover tables (Fox, 1976).

The second example is a three-wave structural equation model of income satisfaction (Campbell and Mutran, 1982).

a. Analysis of income change

Fox examined income level and change between 1968 and 1972 by constructing simple turnover tables. One of these tables (table 1 on page 59 and 60) classified respondents or couples by their income position in 1968 and 1972. The table shows the marginal distributions each year and the joint probability of change separately for married couples, unmarried men, and unmarried women, crossed by work status in 1968 and 1972. The table indicates some increase in income over time for persons either employed or not employed in both years, and, as expected, a substantial decrease in income for persons employed in 1968 but not employed in 1972. Among this latter group, Fox (1976) noted that income loss for unmarried men appeared greater than for unmarried women.

Fox's findings are examples of general questions that can be answered by the analysis of turnover tables.

1. Are income changes between the two points different for different subpopulations?

2. Are there differences in marginal income distributions between sub-populations at a given time?

A number of authors (Bishop et al, 1975, Hauser, 1978, Landis & Koch, n.d., and Singer, 1983) have shown that hypotheses involving marginal distributions and attribute-by-time interactions can be specified and tested using existing methods for the analysis of categorical data. For example, testing whether income changes vary by subpopulation is the same as testing for a 3 (or higher) way interaction between income level at time one, income level at time two, and subpopulation characteristics. The weighted least squares approach (Landis et al., 1976) would be an appropriate methodological approach for testing this kind of hypothesis, especially for complex sample designs. Given a consistent estimate of the sampling covariance matrix for the table cells, appropriate test statistics for a wide variety of hypotheses can be computed.

Fox's analysis also illustrates two additional methodological issues. We are informed in the technical note to his report that only 63 percent of the sample respondents had usable income data in both 1968 and 1972 due to the 'very conservative editing" of income response. In both years, respondents had to give usable answers to about 20 different income components (twice that, if married). An inadequate response to any one of these components was enough to cause a nonresponse for the entire set. Three questions immediately arise. What is the effect of response error for individual income items on the analysis of the turnover tables? Would imputing missing income items affect the analysis? How did analyzing only the partial data set affect the analysis?

Response errors are likely to result in-an overestimate of change in income class, because some of the observed change is due to reporting error rather than to real change over time. Generally, in order to separate real change from classification error, an observation at a third point is required. This third observation could be a reinterview, taken soon after one of the regular waves, designed to measure reporting error directly. However, under certain modeling assumptions, three widely spaced observations can also provide estimates of real change and classification error (see Bye & Schecter 1980 and 1983). A second problem resulting from classification error arises when attempting to measure differences among various subpopulations. There may not be real change at all; the analyses may simply reflect differences in the propensity for response error among the subpopulations, leading to incorrect interpretations.

The effect of imputation on the analysis of turnover tables will depend on the specific imputation scheme. If, for an individual, responses from other waves are used to impute missing values for a particular wave, real change may be understated. If, on the other hand, the amputations are carried out separately for each wave, real change will most likely be overstated. Particular care must also be given to substantive interpretations, when the same attributes are used both for imputation and for substantive analysis.

Analyzing partial data sets requires an assumption that the nonresponents are like the respondents. Usually no studies have been carried out to support that. To the extent that nonrespondents are different, as they frequently are in health and income studies, the data set is biased and the interpretation is inadequate.

b. Stability of income satisfaction

Campbell and Mutran (1982) present an analysis of the stability of income satisfaction over time using data from three waves of the RHS -- 1969, 1971 and 1973. They assume that income satisfaction is an unobserved continuous variable measured imperfectly by two indicator variables. The two indicators are "satisfaction with the way one is living" (SAT), and "ability to get along on income" (GET). Figure C (page 61) presents a path diagram for one of the models estimated by Campbell & Mutran (1982). (The estimated covariance matrix of the observed variables is shown in Table 2, page 62.)

Campbell and Mutran posit that income satisfaction is in turn a function of health status, (an unobserved variable with three indicators), of actual income level in 1969, and of the number of times in the hospital in 1970. The authors note that this path model is significantly underspecified but provides an interesting example of the use of LISREL methodology (J™reskog & S™rbom (1978) and (1979)).

LISREL unites factor analysis and structural equation modeling for a wide variety of recursive and nonrecursive models with and without measurement errors. (see J™reskog & S™rbom, 1976). The LISREL approach assumes that both measurement and structured equations are linear in the unknown parameters and that all variables are normally distributed.

2. Social Security Administration Disability Program Work Incentive Experiments

Under the provisions of the Disability Insurance Amendments of 1980, the Secretary of Health and Human Services was directed to develop and carry out experiments and demonstration projects designed to encourage disabled beneficiaries to return to work and leave the benefit rolls. The primary objective of the experiments is to save trust fund monies. The bill itself contains several examples of the kind of change in the post entitlement program that Congress had-in mind. These include changing entitlement provisions for Medicare benefits, lengthening the trial work period, and modifying treatment of post entitlement earnings, such as the application of a benefit offset based on earnings.

Congress imposed important constraints on the experiments: they must be of sufficient scope and size that results are generalizable to the future operation of the disability program, and no beneficiary may be disadvantaged by the experiments as compared to the existing law.

Eight treatment groups and a control group have been proposed (see (SSA, 1982, for details). Each treatment group represents an alternative to the current post entitlement program representing either some change in the law or administrative practice (or both).

A two stage stratified cluster sample of 31,000 newly awarded beneficiaries was planned for the experiments. The sample would be representative of all beneficiaries under age 60 at the time of award. The sample beneficiaries would be assigned at random to one of the nine experimental groups in such a way that the full experimental design is replicated in each geographic cluster. The total sample size in each treatment group would be 3,000, and there are to be 7,000 in the control group.

Under the current disability program, a beneficiary who returns to work despite continuing severe impairment is granted a 24 month period in which to make a work attempt while remaining on the benefit rolls (the first 12 months with full benefits, the second 12 months with benefits in suspense.) workers are expected to need 1 or 2 years to return to work and 2 or 3 years to complete the trial work period and be terminated from the rolls. Thus an observation period of 4 to 5 years is required to track beneficiaries through the shortest of the post-entitlement outcomes. Observed short-run labor force response will provide some information about the effects of the treatments, but trust fund savings will be significant only if employment is sustained in some groups. Thus, sustained work is the key labor force parameter in the evaluation of the work incentive experiments.

At the same time, the analysis of short run labor force outcomes, commencing about 2 years after the experiments begin, is a necessary first step in gauging trust fund effects. The data available for the short run analysis will consist of a voluntary baseline questionnaire (face-to-face interview) covering socioeconomic and demographic background items plus a series of mandatory quarterly reports (mail with telephone followup) showing the beginning and end of work attempts and monthly earnings for each month of the quarter. The response to the quarterly reports is mandatory because work reports and monthly earnings are required for administrative purposes.

a. Short run longitudinal analysis of return to work

The first step in the analysis of return to work will compare the proportion of beneficiaries who have made a work attempt among treatment and control groups. However, short run differences could be misleading if the full effect of the treatment has not been realized. Consider the hypothetical outcome in figures A and B below.

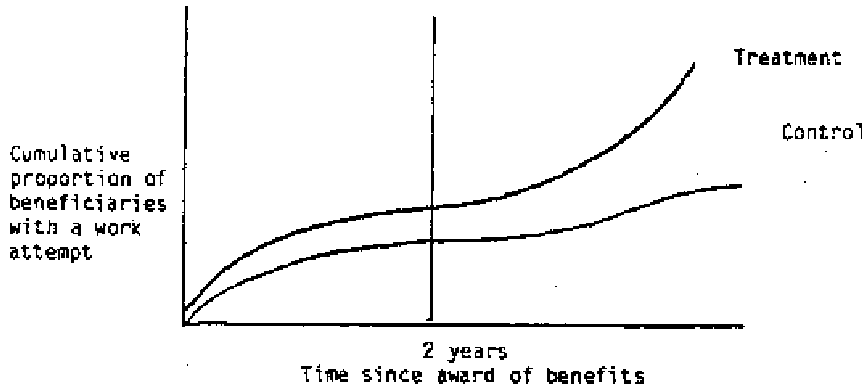


Figure A

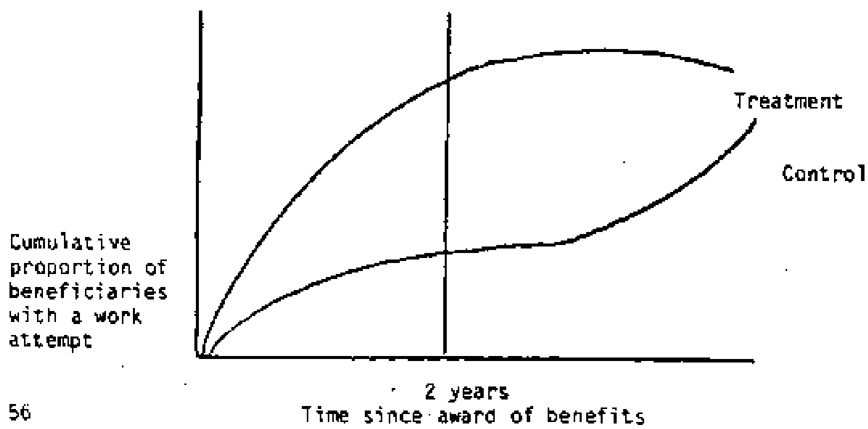


Figure B

56

In Figure A the difference between treatment and control is small for the first two years, but becomes large afterwards. In Figure B, short run difference appears large at first but then becomes smaller. Clearly, change over time in the proportion of beneficiaries who return to work is most important in determining the experimental effect. The rate of change of this proportion over time for beneficiaries who have not yet returned to work is called the hazard rate function (or hazard function). A short run evaluation of return to work will focus on differences in rates of return to work among treatment and control groups.

Using individual observations of the time of return to work, the first analysis of return to work will be to estimate and graph the cumulative hazards of return to work for treatment and control groups and test the difference between the hazards.

If there are differences between treatment and control groups, the graphical displays of the cumulative hazards should provide a useful guide. These can then be used to project long run differences in the probability of return to work among the experimental groups. Introduction of covariates from the baseline questionnaire might also improve the accuracy of these predictions (see Hennessey, 1982).

b. Structural Models of Duration -- Testing a Sociological Theory

It has been suggested that the longer a beneficiary remains on the disability rolls, the less likely he or she is to return to work. The reason given is that the beneficiary makes the necessary social and psychological adjustments to continue in the role of a disabled person. The fact that population rates of return to work for disabled beneficiaries decline over time is often taken as evidence supporting this theory. However, one can show that population heterogeneity can account for an apparent decline in population transition rates over time, even if the individual rates are constant or increasing. (See Heckman & Singer, 1982, for example.) Therefore any assessment of the apparent negative duration dependence must account for population heterogeneity.

One way to examine this issue is to specify and estimate a structural model for the hazard function for return to work. The parameters are usually estimated by maximum likelihood methods, incorporating the likelihoods for sample cases moving from nonwork to work at time t, and for sample cases which haven't yet moved by time t (which, in this case, is the end of the observation period).

c. Estimating long run trust fund effects

The Disability Amendments mandate that the primary evaluation of the experiments be in terms of trust fund effects. In general, the cost to the trust funds of an individual beneficiary is the sum of the expected costs to the Disability and Medicare funds between initial entitlement and the termination of benefits or the attainment Of age 65. The cost to the disability trust fund can be further broken down into the sum of the cash benefit payments plus the cost of vocational rehabilitation (if applicable)

minus the payback of FICA contributions (if the beneficiary returns to work) during this period. The estimation of long run effects requires the projection over time of the probability of receiving cash benefits for disability, the expected amount of those benefits, the probability of working, and the expected earnings level.

An analysis plan for the WIE is being developed which is based on a continuous-time stochastic model. The state space for the process admits four possibilities:

E.1 : Recovered

E.2 : Deceased

E.3 : Nonworking Beneficiary

E.4 : Working Beneficiary

At the time benefits are awarded the beneficiary is assumed to be in state E.3. The beneficiary can switch between states E.3 and

E.4 until he or she reaches state E.1 or E.2 (which are taken to be absorbing states) or reaches age 65 (and is automatically converted to the old age program.)

A semi-Markov model is proposed to link the various work and non-work episodes over time. This model assumes that each work and non-work period is independent of prior work history (but might depend on age and other exogenous factors which can be incorporated into the hazard functions.) Although it is unlikely that this sort of independence does in fact exist, the short observation period effectively precludes the ability to detect the real dependencies.

In conclusion, once the hazard functions are estimated separately for each experimental group, future work and benefit status histories will be simulated. These histories together with estimates of earnings and benefit levels will allow the estimation of long run trust fund costs for each experimental group.

Using four years of administrative data, Hennessey (1982) found that semi-Markov models of work and benefit status for male beneficiaries can accurately predict the histories three years hence. His,results provide encouragement for this overall analysis strategy.

Table 1.

Total money income, 1972, by total money income, 1968 (in current dollars): Percentage distribution of respondents, by employment status in 1968 and 1972, marital status, and sex [1]

Total money income, 1972, and employment status, 1968 and 1972	Total	Less than $2,500	$2,500–4,999	$5,000–7,499	$7,500–9,999	$10,000–14,999	$15,000–19,999	$20,000 or more
	Married men and spouses							
Employed in 1968 and 1972								
Total number	[illegible]							
Total percent [1]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Less than $2,500	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2,500–4,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5,000–7,499	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7,500–9,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10,000–14,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
15,000–19,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
20,000 or more	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Employed in 1968 but not in 1972								
Total number	[illegible]							
Total percent [1]	100	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Less than $2,500	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2,500–4,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5,000–7,499	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7,500–9,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10,000–14,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
15,000–19,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
20,000 or more	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Not employed in 1968 or 1972								
Total number	[illegible]							
Total percent [1]	100	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Less than $2,500	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2,500–4,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5,000–7,499	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7,500–9,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10,000–14,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
15,000–19,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
20,000 or more	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	Nonmarried men							
Employed in 1968 and 1972								
Total number	[illegible]							
Total percent [1]	100	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Less than $2,500	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2,500–4,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5,000–7,499	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7,500–9,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10,000–14,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
15,000–19,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
20,000 or more	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Employed in 1968 but not in 1972								
Total number	[illegible]							
Total percent [1]	100	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Less than $2,500	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2,500–4,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5,000–7,499	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7,500–9,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10,000–14,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
15,000–19,999	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
20,000 or more	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

From:

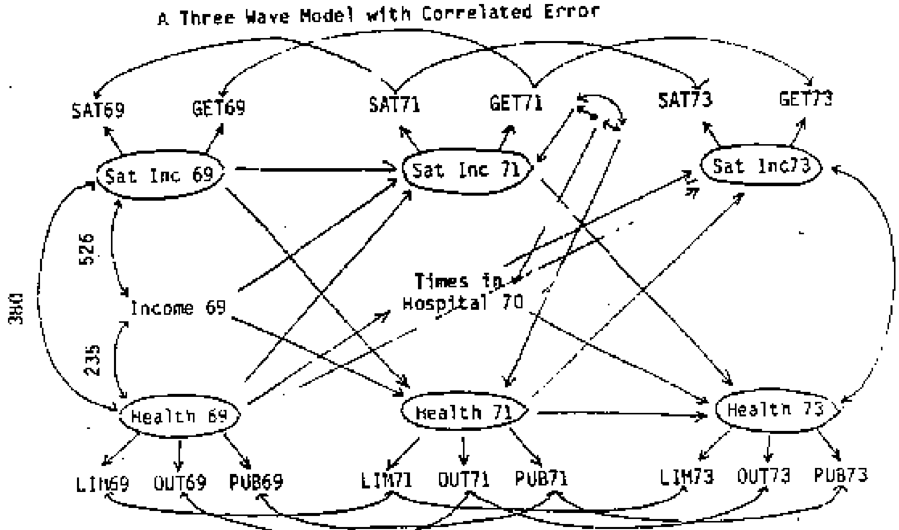
Fox, Alan
1975 "Work status & income change, 1968-1972: Retirement History Study Preview" in Social Security Bulletin, DHEW Publication No. (SSA) 76-11700.

Total money income, 1972, and employment status, 1968 and 1972	Total money income, 1968							
	Total	Less than $2,500	$2,500–4,999	$5,000–7,499	$7,500–9,999	$10,000–14,999	$15,000–19,999	$20,000 or more
	Nonmarried men—Continued							
Not employed in 1968 or 1972								
Total number	107							
Total percent [1]	100	72	19	6	3	0	1	0
Less than $2,500	53	50	2	0	1	0	0	0
2,500–4,999	30	20	10	1	0	0	0	0
5,000–7,499	6	1	7	6	0	0	0	0
7,500–9,999	4	0	0	4	0	0	0	0
10,000–14,999	4	1	0	1	2	0	0	0
15,000–19,999	1	0	0	0	0	0	1	0
20,000 or more	0	0	0	0	0	0	0	0
	Nonmarried women							
Employed in 1968 and 1972								
Total number	514							
Total percent [1]	100	22	38	24	9	6	0	1
Less than $2,500	16	11	4	0	0	0	0	0
2,500–4,999	31	10	19	2	0	0	0	0
5,000–7,499	25	1	13	10	1	0	0	0
7,500–9,999	14	0	2	10	2	0	0	0
10,000–14,999	11	0	1	2	5	3	0	0
15,000–19,999	2	0	0	0	0	1	0	0
20,000 or more	1	0	0	0	0	0	0	0
Employed in 1968 but not in 1972								
Total number	233							
Total percent [1]	100	29	34	22	9	6	0	0
Less than $2,500	42	17	23	4	0	0	0	0
2,500–4,999	28	4	11	11	2	1	0	0
5,000–7,499	20	6	0	6	2	2	0	0
7,500–9,999	6	0	0	1	4	0	0	0
10,000–14,999	3	0	0	0	0	2	0	0
15,000–19,999	1	0	0	0	0	0	0	0
20,000 or more	0	0	0	0	0	0	0	0
Not employed in 1968 or 1972								
Total number	320							
Total percent [1]	100	84	11	3	1	0	0	0
Less than $2,500	72	68	2	0	0	0	0	0
2,500–4,999	22	13	7	1	0	0	0	0
5,000–7,499	4	1	2	2	0	0	0	0
7,500–9,999	1	0	0	1	0	0	0	0
10,000–14,999	1	0	0	0	0	0	0	0
15,000–19,999	0	0	0	0	0	0	0	0
20,000 or more	0	0	0	0	0	0	0	0

[1] For 1972 total money income, expressed as percentages of total number.

Figure C

A Three Wave Model with Correlated Error



Note: Coefficients for this model are reported in Table 2, following page.

DEFINITIONS OF VARIABLES

A. Satisfaction with Income is an unmeasured construct with three

indicators:

1. SAT69, SAT71, SAT73

Are you satisfied with the way you Does are living?

4 = More than satisfied

3 = Satisfied

2 = Less than satisfied

1 = Very unsatisfied

2. GET 69, GET71, GET73

Ability to get along on income

4 = Always have money left over

3 = Have enough with a little left over sometimes

2 = Have just enough, no more

1 = Can't make ends meet

B. Health is an unmeasured construct with three indicators:

1. LIM69, LIM71, LIM73

   Does health limit the kind of work you do?

   2 = No

   1 = Yes

2. OUT69, OUT71, OUT73

   Are you able to leave the house without help?

   3 = No limitation

   2 = Yes, though health limit work

   1 = No

C. Number of times in hospital (HOS7 is measured with one indicator

D. 1969 household income (INC69) is single indicator of log

income from all sources

from all sources from Campbell and Mutran, 1992.

Reprinted with permission

Table 2

VARIABLES, MEANS, STANDARD DEVIATIONS, VARIANCES, and CORRELATIONS

	SAT69	GET69	INC69	LIM69	OUT69	PUB69	SAT71	GET71	HOS70	LIM71	OUT71	PUB71	SAT73	GET73	LIM73	OUT73	PUB73
SAT69	1.000																
GET69	0.535	1.000															
INC69	0.327	0.436	1.000														
LIM69	0.282	0.298	0.236	1.000													
OUT69	0.213	0.229	0.201	0.882	1.000												
PUB69	0.221	0.237	0.211	0.833	0.741	1.000											
SAT71	0.445	0.411	0.290	0.229	0.173	0.185	1.000										
GET71	0.419	0.590	0.401	0.262	0.208	0.211	0.522	1.000									
HOS70	-0.039	-0.041	-0.020	-0.102	-0.074	-0.079	-0.077	-0.070	1.000								
LIM71	0.238	0.244	0.195	0.543	0.462	0.449	0.263	0.276	0.191	1.000							
OUT71	0.160	0.171	0.158	0.419	0.393	0.367	0.196	0.216	-0.158	0.888	1.000						
PUB71	0.175	0.177	0.159	0.413	0.375	0.379	0.198	0.217	-0.162	0.833	0.279	1.000					
SAT73	0.361	0.363	0.233	0.196	0.166	0.158	0.428	0.412	-0.049	0.216	0.170	0.172	1.000				
GET73	0.374	0.521	0.356	0.245	0.198	0.207	0.408	0.582	-0.069	0.233	0.185	0.816	0.507	1.000			
LIM73	0.199	0.237	0.168	0.459	0.392	0.390	0.214	0.236	-0.144	0.502	0.428	0.418	0.246	0.277	1.000		
OUT73	0.122	0.146	0.126	0.313	0.292	0.279	0.141	0.159	-0.111	0.385	0.382	0.370	0.195	0.215	0.886	1.000	
PUB73	0.128	0.153	0.128	0.308	0.279	0.284	0.143	0.160	-0.104	0.376	0.366	0.371	0.189	0.214	0.833	0.807	1.000
MEAN	2.781	2.451	8.531	1.661	2.663	2.710	2.803	2.508	0.156	1.636	2.679	2.699	2.780	2.465	1.578	2.660	2.672
SD	0.687	0.951	1.086	0.473	0.534	0.485	0.715	0.961	0.487	0.481	0.513	0.511	0.698	0.925	0.494	0.512	0.523
VAR.	0.472	0.904	1.179	0.224	0.285	0.235	0.511	0.924	0.237	0.231	0.263	0.261	0.487	0.865	0.244	0.262	0.273

From Campbell and Mutran, 1982.
Reprinted with permission.

SUMMARY AND CONCLUSIONS

In developing the working paper on longitudinal surveys, the subcommittee found that few of the issues were simple. For each question that was raised there were multiple and sometimes contradictory conclusions encountered in the literature, or in the experience of the subcommittee members. This complicated the task of drawing conclusions about when or how to use longitudinal surveys; what was is clear is that anyone considering a longitudinal survey should remember four general points. These points could apply equally well either to longitudinal or to cross-sectional surveys, but certain aspects are especially important in longitudinal surveys.

First, research goals should be clearly stated and alternative kinds of data collection should be evaluated. Cross-sectional research is not automatically less expensive, and certain research goals cannot be attained with one-time surveys. The evidence seems to indicate that longitudinal surveys are not intrinsically more costly than one-time surveys of comparable scope. In many cases, one longitudinal survey will be more efficient than a series of one-time surveys. However, cost considerations may dictate that neither a longitudinal survey nor a series of one-time surveys could be carried out. Compromises are often made on frequency of interview or sample size to permit some longitudinal data collection.

- For certain research goals, such as identifying the frequency or duration of change, or the causes of change (as in longitudinal surveys of labor force status), only a longitudinal survey will work. For topics that are difficult for respondents to recall, such as attitudes or detailed behavior (as in longitudinal surveys of retirement, or health treatments, or household income), a prospective longitudinal survey is the best choice.

- All other things being equal, a longitudinal survey achieves a given level of precision for measures of change with a somewhat smaller sample than is possible in a series of one-time surveys. In addition, the cost of maintaining contact with a longitudinal sample may be no higher than the cost of selecting and contacting a one-time sample.

- Timing of results plays an important part in the decision to select a longitudinal survey. If early results are needed, then a longitudinal survey is not appropriate. If early waves of a longitudinal survey can be analyzed quickly and provide useful information, then some of the timing problem is dissipated. If the research needs can only be met by a longitudinal survey and those waiting for results clearly understand the timing, longitudinal surveys are clearly superior.

Second, once the decision has been made to conduct a longitudinal survey, the subcommittee recommends that a greater emphasis be placed on the early formulation of clear and specific analysis objectives as the next step in research planning. The failure to formulate detailed analysis early enough explains some of the disappointments that some organizations have experienced with longitudinal surveys.

- As the simplest example, when research objectives are not clearly stated or understood, the longitudinal nature of the data has not always been fully exploited in analysis.

- Many of the operational features of longitudinal surveys should only be selected after the development of clear and specific plans for analysis. Even such seemingly unrelated factors at the interval between interviews may be determined by analysis plans. For example, discrimination between some simple stochastic models is ruled out if data collection intervals are constant. Other examples are given in Singer and Spilerman's study of longitudinal analysis (1976).

- A clear statement of specific research goals, including analysis plans, reduces the likelihood that a project will require unanticipated funding extensions or auxiliary sponsors for completion. Comprehensive planning ensures that a survey will appeal to a wide constituency, and reflect the research goals of an adequate sponsorship base.

- Fully developed research objectives make it less likely that a need for different -- or additional -- data will become apparent part way through the survey.

Third, longitudinal surveys can easily incorporate features that facilitate the evaluation of internal data quality, and that compare the effectiveness or cost of alternative methods. Repeated data collection makes this possible in ways that are beyond the scope of a one-time survey.

- Any longitudinal survey that varies data collection mode while maintaining a constant questionnaire can be a vehicle for studying the impact of mode of interview on response. Evaluations have indicated that the NLS obtained comparable results by using personal or telephone interviews after the first interview, for example.

- Data from longitudinal surveys can be used to understand the impact of nonresponse on the representativeness of a sample. The characteristics of nonrespondents in a later wave can be studied through what is known about them from the first interview, or from later follow-ups in which they do respond. In the NLS, each extension of the survey has been preceded by evaluations of the impact of attrition through comparisons with population controls developed in the first wave of interviews.

- The effect of continued participation on response can be evaluated each time new persons are brought into the sample or interviewed for the first time. The original HS+B survey program, for example, provided for an additional sample; a group from the original sample to be interviewed only in the later waves, specifically to evaluate panel effects.

- Alternative methods for simulating complete response from incomplete data (such as imputing from other cases, or from what was reported in another interview, or by increasing the weight of completed interviews) can be evaluated using a longitudinal file. The final comparisons have to wait until all the waves of a longitudinal survey are completed, but preliminary results can be used in earlier waves, and a variety of procedures can be compared at the end of the program in order to select the most effective method.

- Data from longitudinal questionnaires can and should be compared to the results from comparable questions asked of similar respondents in one-time surveys. The results of NLS labor force questions were constantly evaluated against cross-sectional labor force surveys. This provides ongoing information on sampling error, and on the impact of questionnaire design on response.

- Data from a longitudinal survey, from related administrative records,and from comparable surveys of one-time samples can be compared to estimate the impact of recall periods, or the interval between interviews, or the effect of bounding interviews. The Income Survey Development Program demonstrated the importance of just such an exhaustive testing program which accompanied planning for SIPP.

- The costs of alternative data collection strategies should be recorded, along with the operational considerations and the impact on data quality. This information will be invaluable when the most efficient methods must be chosen for other surveys.

- The costs and effects of alternative data processing strategies should be recorded to allow comparisons, such as the costs and benefits of matching longitudinal records through characteristics or through unique identification codes for sample persons and households. Early tests such as these led to the development of the case-linking strategy selected for SIPP.

These and many other comparisons are possible with longitudinal surveys, because so many materials, respondents and operations vary throughout the course of the survey. With minimal additional efforts toward record-keeping and control, most longitudinal operations can provide important data for evaluating internal data quality and to guide future survey designers.

Fourth there are many measurement error problems that exist with any kind of survey, some of which are exacerbated by a longitudinal design. So far, the research on many of these methodological problems has not been definitive, so choices are made based on cost and intuition. There is a rich field for investigations and those seeking to do longitudinal surveys should strive to include some methodological elements. Some of this kind of research has been carried out, as described above, but more is needed.

- Time-in-sample bias permeates every survey that requires repeated interviewing. It is not limited to one particular kind of variable or one mode of data collection. As a result there is a systematic bias in the data that shows up when data are compared by the number of interviews a respondent has had. No one knows which set of data are more accurate, those from earlier or those from later interviews. People make judgments based on little or no data, and the topic needs careful investigation.

- Response errors have the effect of exaggerating change. People do forget and change their minds, and different household respondents give different answers to the same questions. The length of time between interviews also influences answers. More work needs to be done to separate real from spurious change.

- Attrition is a serious problem in longitudinal surveys. Many longitudinal surveys are able to keep 90 to 95 percent of their respondents on each interviewing wave, but even low nonresponse mounts over time. Although compensation strategies look promising, it is troublesome to realize that for some variables, a quarter to one-half of the data are not given by respondents.

- There has been little research on the best length of time to allow between interviews. Decisions are based mainly on cost, yet we know that the longer the interval, the less that is reported, and the more that is reported in the wrong time periods. Work needs to continue on this aspect.

- It is known that the questions on a survey are not processed one by one by respondents. The presence of questions on other topics affects responses to questions on variables of interest. This happens whether the additional questions precede or follow the main questions. However, the tendency is to keep adding new topics. We may be causing a deterioration of data quality by doing this.

Longitudinal surveys are increasingly being used as the basis for policy decisions by the Federal government. In our review, we have become convinced that for some research goals there is no alternative to longitudinal data collection. However, before agencies make the decision to conduct a longitudinal survey, they should carefully consider the important operational, management, and statistical problems associated with them.

CASE STUDY 1

SURVEY OF INCOME AND PROGRAM PARTICIPATION

I. Purpose of the survey

In October 1983, the Bureau of the Census conducted the first interviews of the Survey of Income and Program Participation (SIPP). The SIPP is a nationally representative household survey intended to provide detailed information on all sources of cash and noncash income, eligibility and participation in various government transfer programs, disability, labor force status, assets and liabilities, pension coverage, taxes, and many other items. Data from the survey will provide a multiyear perspective on changes in income, and their relationship to participation in government programs, changes in household composition, and so forth. In general, the SIPP data system is designed to measure elements of the federal tax and transfer system in a comprehensive data base.

SIPP began in response to the recognition that the principal source of information on the distribution of household and personal income in the United States -- the March Income Supplement of the Current Population Survey (CPS) had limitations which could only be rectified by making substantial changes in the survey instrument and procedures. For example, the CPS does not provide monthly income, monthly household composition or detailed asset information. These deficiencies became especially serious when the scope of policy analyses was broadened during the 1960's and early 1970's as public assistance programs were expanded and reorganized. Model-builders were forced to make many assumptions and impute intrayear data using CPS data to carry out their activities. In this environment, with analysts requiring more detailed data and improved measures of cash and noncash income, the Income Survey Development Program (ISDP) was established.

The purpose of the ISDP, authorized in 1975, was to design and prepare for a major new,survey, the Survey of Income and Program Participation (SIPP). The ISDP developed methods intended to overcome the three principal shortcomings of the CPS for analyses of income: 1) the under reporting of property income and other irregular sources of income; 2) the underreporting and misclassification of participation in major income security programs and other types of information that people generally find difficult to report accurately (for example, monthly detail on income earned during the year); and 3) the lack of information necessary to analyze program participation and eligibility (annual income estimates were available, but eligibility for most Federal programs is based on a monthly accounting period).

Four experimental field tests were conducted to examine different concepts, procedures, questionnaires, and recall periods. Two of the tests were restricted to a small number of geographic sites, the other two were nationwide. The largest test, conducted in 1979, was also the most complex. Although used primarily for methodological purposes, the nationally representative sample of 8,200 households was sufficiently large to provide reliable national estimates of many characteristics. More detailed discussions of the ISDP and its activities are provided in Ycas and Lininger (1981) and David (1983).

Because the ISDP was the predecessor to SIPP, it is not surprising that many characteristics of the ISDP are reflected in the SIPP design, including many elements of the survey's design, content, and questionnaire format.

II. Sponsors

The ISDP development effort was directed by the Office of the Assistant Secretary for Planning and Evaluation in the Department of Health and Human Services and was carried out jointly with the Bureau of the Census, which assisted in the planning and carried out the field work, and the Social Security Administration (SSA), which administers the major cash income security programs. In late 1981 virtually all funding for ISDP research and planning for the ongoing SIPP program was deleted from the budget of the Social Security Administration. The loss of funding for fiscal year 1981 brought all work on the new survey to a halt. Then in fiscal year 1983, money for the initiation of the new survey was allotted in the budget of the Bureau of the Census.

In planning the content, procedures, and products of the SIPP, the Census Bureau works closely with a SIPP Interagency Advisory Committee, established and chaired by the Office of Management and Budget (OMB). The committee consists of individuals representing the following departments and agencies: the Departments of Labor, Education, Defense, Commerce, Agriculture, Health and Human

Services, Treasury, Housing and Urban Development, and Justice; Energy Information Administration; National Science Foundation; Council of Economic Advisors; Congressional Budget Office; Bureau of Labor Statistics; Bureau of Economic Analysis; Veterans Administration; Internal Revenue Service; and the Office of Management and Budget.

III. Sample Design

SIPP started in October 1983 as an ongoing survey program of the Bureau of the Census with one sample panel of approximately 21,000 households in 174 primary sample units (PSU's) 1/ selected to represent the noninstitutional population of the United States. The sample design is self-weighting; that is, each unit selected in the sample has the same probability of selection.

In February 1985 and every February thereafter, a new, slightly smaller panel of 15,000 households is introduced. This design allows cross-sectional estimates to be produced from the combined sample from both panels. The overlapping panel design enhances the estimates of change, particularly year-to-year change. Since portions of the sample are the same from one year to the next, year-to-year change estimates can be based in part on a direct comparison across 2 years for the same group of households.

To facilitate field operations, the sample is divided into four approximately equal subsamples, called rotation groups; one rotation group is interviewed in a given month. Thus, one cycle or "wave" of interviewing takes 4 consecutive months. This design creates manageable interviewing and processing workloads each month instead of one large workload every 4 months; however, it results in each rotation group using a different reference period.

Data collection operations are managed through the Census Bureau's 12 permanent regional offices. Interviewers assigned to these offices conduct one personal visit interview with each sampled household every 4 months. At the time of the interviewer's visit, each person 15 years old or older who is present is asked to provide information about himself/herself; a proxy respondent is asked to provide information for those who are not available. The average length of the interview is about 30 minutes. Telephone interviewing is permitted only to obtain missing information or to interview persons who will not or cannot participate otherwise.

An important design feature of SIPP is that all persons in a sampled household at the time of the first interview remain in the sample even if they move to a new address. For cost and operational reasons, personal-visit interviews are only conducted at new addresses that are in or within 100 miles of a SIPP primary sampling unit (persons moving outside that limit are contacted by telephone if possible). After the first interview, the SIPP sample is a person-based sample, consisting of all individuals who were living in the sample unit at the time of the first interview -- these people are labelled original sample persons". Individuals aged 15 and over who subsequently share living quarters with the original sample people are also interviewed in order to provide the overall economic context of the original sample persons. Changes in household composition caused by persons who join or leave the household after the first interview are also recorded. These individuals are interviewed as long as they reside with an original sample person. More information about these procedures can be found in Jean and McArthur (1984).

IV. Survey Design and Content

Each person in the SIPP sample is interviewed once every 4 months for 2 2/3 years to produce sufficient data for longitudinal analyses while providing a relatively short recall period for reporting monthly income. The reference period for the principal survey items is the 4 months preceding the interview. For example, in October, the reference period is June through September; when the household is interviewed again in February, it is October through January. This interviewing plan will result in eight interviews per household.

An important design feature of SIPP is the assignment of an individual identification number. Each sample person is assigned a unique fourteen-digit identification (ID) number at the time he/she enters the sample; an additional two-digits code is assigned if the person moves to a new address. A master list of identification numbers is used by the regional offices to monitor the status of interviewing each month after Wave 1. The regional offices keep track of each number on the list representing all the persons assigned for interview in a month; each must be accounted for with a completed questionnaire or a reason for noninterview. The list is updated regularly to account for persons who are added or deleted from the sample.

The ID helps to link information about an individual across time; it identifies which household each person is a member of at any point in the panel. Through the ID system, data can be linked from all persons ever associated with a given household throughout the 2 2/3-year duration of a panel.

The survey consists of three major components: (1) the control card, (2) the core data, and (3) topical data. The control card is used to obtain and maintain information on the basic Characteristics associated with households and all household members and to record information for operational control purposes. These data include the age, race, ethnic origin, sex, marital status, and educational level of each member of the household, as well as information on the housing unit and the relationship of the householder to other members. A household respondent provides this information, which is updated at each interview. The control card is also used to keep track of when and why persons enter and leave the household, thereby providing enough information to compose monthly household and family groups. There is also space to record information that will improve the interviewer's ability to follow persons who move during the survey. In addition, after each visit, data on employment, income, and other information are transcribed from the core questionnaire to the control card so the data can be used in the next interview as a reference for the interviewer and thus shorten succeeding interviews.

A questionnaire is filled for each household member who is 15 years or older. The questionnaire consists of a "core" of labor force and income questions asked during each interview and a set of topical modules which are scheduled during the life of the panel. The core labor force and income questions are designed to measure the economic situation of persons in the United States. These questions expand the data currently available on the distribution of cash and noncash income and are repeated at each interviewing wave. SIPP core data build an income profile of each person aged

15 and over in a sample household. The profile is developed by determining the labor force participation status of each person in the sample and asking specific questions about the types of income received, including transfer payments and noncash benefits from various programs for each month of the reference period. A few questions on private health insurance coverage are also included in the core.

Persons employed at anytime during the 4-month reference period are asked to report on jobs held or businesses owned, number of hours and weeks worked, hourly rate of pay, amount of earnings received, and weeks without a job or business in addition to questions about labor force activity and the earnings from a job, self-employment, or farm, the core includes questions related to nearly 50 other types of income as well as the ownership of assets which produce income.

The SIPP has been designed to provide a broader context for analysis by adding series of questions on a variety of topics not covered in the core section. These questions are labelled "topical modules" and are assigned to particular interviewing waves of the survey. If more than one observation is needed, a topical module may be repeated in a later wave.

The survey design allows for the inclusion of these special modules because less time is required in later waves to update the core information collected in the first interview. The subjects covered do not require repeated measurement at each interview and, therefore, may use a reference period longer than the period used for the core information. Examples of topical modules include health and disability, work history, assets and liabilities, pension plan coverage, tax-related information, marital history, fertility, migration, household relationships, child care arrangements, and pension plan coverage. For more information about the SIPP design refer to Nelson, McMillen, and Kasprzyk (1984).

V. Survey Response Rates

The first SIPP interviews were conducted in October 1983. At this time, cumulative household noninterview rates are available for the first six waves of SIPP, that is, through August 1985. Sample loss through the sixth wave has been 19 percent, of which 15 percent was due to refusals and other situations in which the interviewer was unable to make contact with the household, and 4 percent was due to movers that the interviewer was not able to contact again.

Survey nonresponse rates for persons are discussed in McArthur and Short (1985). In this work they characterize the population that is leaving the sample; comparing these persons' characteristics to those of persons continuing to be interviewed in the survey. At the end of the third wave of interviewing, combining all reasons for noninterview -- including refusals, institutionalization, move s to unknown addresses, persons who were temporarily absent, and so on -- 10.5 percent of all persons who were interviewed during the first wave had left the sample. There is some indication that those noninterviewed persons are different from persons who continue to be interviewed. Noninterviews are more likely to be renters rather than homeowners, to live in large urban areas, and to have reported their marital status to be single or separated.

Coder and Feldman (1984) found that imputation for a selected group of items was quite small. In this analysis item nonresponse rates on labor force, income recipiency, and income amounts are examined. They also discussed the impact of self or proxy respondents on nonresponse rates. Lamas and McNeil (1984) discussed the quality of data measuring household wealth in the survey. The nonresponse rate was low for all asset types (1.4 percent) for all persons asked about asset ownership. They found that nonreponse rates varied by type of asset -- lowest for rental property and highest for certificates of deposit -- and by age and education levels of the respondents -- higher nonresponse for older persons and higher nonresponse with greater educational attainment. McMillen and Kasprzyk (1985) used counts of amputations made for each person as the measure of item response rates. The maximum number of amputations that could have been made for an individual was 83. They found that in the first two waves of interviews, 86 percent of the persons had no imputation at all. In Waves 1 and 2, respectively, 87 percent and 92 percent of the cases with some imputation had no more than 3 items imputed. More work planned to study nonresponse is discussed in the research section.

VI. Survey Evaluation Work

SIPP evaluation work is in an early stage; the Census Bureau and other users of the SIPP data are developing appropriate methods of evaluation. For example, research is being carried on for three types of nonresponse -- unit nonresponse defined as nonresponse to all waves of the survey, wave nonresponse defined as nonresponse to a particular wave interview, and item nonresponse defined as nonresponse to a particular item -- and their patterns of occurrence.

Another area of useful evaluation work combines survey data with administrative record data. The SIPP was developed as an integrated data system in order to use combined information sources to validate and supplement information collected in the survey. An internal Census Bureau committee is assessing the potential uses of administrative data linkages and identifying content and availability of administrative record systems for use in demonstration studies. One record linkage project which is currently under development will match SIPP survey data for individuals to their administrative records at the state level. Various federal record systems which may also be brought into this project are also being investigated. At this time both the number of states and the number of records systems involved is limited.

A discussion of the quality of the income data collected as of each wave of the SIPP is contained in an appendix to each SIPP quarterly report (U.S. Bureau of the Census). The appendix supplies information on the nonresponse rates for selected income questions, the average amounts of income reported in the survey or assigned in the imputation of missing responses, and the extent to which the survey figures underestimate numbers of income recipients and amounts of income received. For example, in the report for the third quarter of 1983 (P70, no.1) nonresponse rates range from a low of about 3 percent for Aid to Families with Dependent Children (AFDC) and food stamp allotments, to about 13 percent for self-employment income. The report states that survey underestimates of income recipients ranged from about 21 percent for AFDC to about 1 percent for Social Security recipients, and the survey estimate of persons receiving state unemployment compensation payments was about 103 percent of the independent estimate. The underreporting for AFDC is-related to misclassification of this income type as other types of public assistance or welfare.

Evaluation of the ISDP is relevant to work in the SIPP. For example because of its design, SIPP has a potential for missing and inconsistent data problems from wave to wave. One area of current research is the phenomenon of significant income changes and program turnover occurring between waves more often than within waves. Some analysis of this phenomenon using data from the 1979 ISDP Panel is presented in Moore and Kasprzyk (1984). Continuing this area of research using data from SIPP, Burkhead and Coder (1985) looked at gross changes in income recipiency from month to month over a period of one year, the first three waves of SIPP. Their examination indicated that change in recipiency statuses was significantly higher for the months that spanned successive interviewing reference periods, that is between the last reference month for one interview and the first from the next interview. Vaughan, Whiteman, and Lininger (1984) also discussed the quality of income and program data in the ISDP. They discuss numbers of income recipients and program participants, and amounts of income and benefits in comparison to independent sources and the CPS. Other relevant studies are: Ferber and Frankel (1981), studying the reliability of the net worth data in the 1979 panel of the ISDP; Feldman, Nelson and Coder (19801, evaluating the quality of wage and salary income reporting in the 1978 ISDP; and U.S. Bureau of the Census (1982).

VII. Survey Data Products and Research Activities

A number of publications and public-use data files are being generated from the information collected in SIPP. Both publications and data files are identified by whether they are cross-sectional or longitudinal. Two types of cross-sectional reports are planned by the Census Bureau: 1) a set of quarterly reports that focus on core information; and 2) periodic or onetime reports that use the detailed data from the topical modules.

The quarterly cross-sectional reports show average monthly labor force activities, income, and program participation statistics. The first quarterly report was issued in fall 1984 (U.S. Bureau of the Census, 1984) and contains data referring to the Third Quarter of 1983. The report covering the Fourth Quarter of 1984 was released in November 1985. The periodic and single-time reports will use the detailed data from the topical modules (for example, disability and earnings, health insurance coverage and household net worth). These reports may also use a combination of the core and topical module data.

Plans for longitudinal data reports are under discussion, but they are expected to concentrate on data that can be used to examine trends and changes over time. This may include analyses of the dynamic aspects of the labor force or the effect of changes in household composition on economic status and program participation. Examples of reports under consideration in this series are: economic profile reports, presenting yearly aggregates of monthly data on individuals; comparative profile reports, presenting annual comparisons of the economic activity of individuals; transition reports, providing changes in income and program participation status between two points in time: longitudinal family and unrelated individual reports, presenting the characteristics of longitudinal family units defined in SIPP (see McMillen and Herriot.(1984) for more information on this topic); and special event reports, providing data preceding and/or following a particular event, such as marriage, divorce, separation, the birth of a child, a return to school, a move to a new address, or a job change.

SIPP cross-sectional data files are issued on a wave-by-wave basis. 2/ Each file includes person, family, and household information collected in the survey wave. Virtually all data obtained on the core questionnaire are included on the files; certain summary income recodes are also included. Data that might disclose the identity of a person are excluded or recoded in accordance with standard Census Bureau confidentiality restrictions. Wave files are edited, imputed, and weighted in a manner consistent with their use for cross-sectional analysis. A unique identification number is included to allow users to merge two or more SIPP files. However, since the processing of wave files is independent, wave-to-wave data inconsistencies will occur and the user must be prepared to resolve them.

Data files containing topical module information will be released together with the core data that were collected at the same time. Identifiers will be included on the file to allow linkage to other topical module files.

Plans for producing public-use files designed for longitudinal analysis are now under discussion. The first longitudinal file For SIPP will be a research file containing twelve months of core income data; this is essentially the first three SIPP interviews.

A SIPP working paper series has been established as a mechanism to provide timely and widespread access to information developed as part of the SIPP. Papers in the series will cover a broad range of topics including: procedural information on the collection and processing of data; survey methodology research; and preliminary substantive results, such as the measurement of household composition change over time

The 1984 and 1985 meetings of the American Statistical Association were used to bring the research community up-to-date on a variety of SIPP-related research issues. A wide range of topics, both methodological and substantive, were covered in sessions organized under the auspices of the Social Statistics and Survey Research Methods Sections. Papers presented in 1984 have been compiled by Kasprzyk and Frankel (1985) and the 1985 papers have been compiled by Frankel (1985).

A number of other research projects are underway at the Census Bureau and at independent research centers such as the Survey Research Center/University of Michigan. These projects are vital to the understanding, use, and future development of the SIPP. This work includes studies of longitudinal imputation and weighting strategies; characteristics of persons who become nonrespondents: composite estimation; potential for use of data base management systems; linkage of administrative records and economic data from other census files to SIPP results, see Sater (1985). The American Statistical Association (ASA)-National Science Foundation (NSF)-Census research fellow program has been expanded reidentify explicitly SIPP-related research activities.

1/ A primary sampling unit consists of a county or a group of contiguous counties.

2/ For information about the SIPP public use files, please call the Data Users Services Division at (301) 763-4100 and ask for the "Data Developments" for SIPP.

CASE STUDY 2

CONSUMER PRICE INDEX

I. Purpose

The Consumer Price Index (CPI) is a measure of price change for a fixed quantity and quality of goods and services purchased by consumers. The CPI is used most widely as an index of price change. During periods of price increases, it is an index of inflation and services as an indicator to measure the effectiveness of Government economic policy.

The CPI is used also as a deflator of other economic series, that is, to adjust other series for price changes and to translate these series into inflation-free dollars. These series include retail sales, hourly and weekly earnings, and some personal consumption expenditures used to calculate the gross national product (GNP) - all important indicators of economic performance.

A third major use of the CPI is to adjust income payments. More than 8.5 million workers are covered by collective bargaining contracts which provide for increases in wage rates based on increases in the CPI. In addition to workers whose wages or pensions are adjusted according to changes in the -- PI, the index now affects the income of more than 50 million persons, largely as a result of statutory action: Almost 31 million social security beneficiaries, about 2« million retired military and Federal Civil Service employees and survivors, and about 20 million food stamp recipients. Changes in the CPI also affect the 25 million children who eat lunch at school. Under the National School Lunch Act and the Child Nutrition Act, national average payments for those lunches and breakfasts are adjusted semi-annually by the Secretary of Agriculture on the basis of the change in the CPI series, "Food away from home".

Also, the official poverty threshold estimate, which is the basis of eligibility for many health and welfare programs of Federal, state and local governments, is updated periodically to keep in step with the CPI. Under the Comprehensive Employment and Training Act of 1973, the "low income" criterion for distribution of revenue-sharing funds, is kept current through adjustments based on the index.

In addition, the Economic Recovery Tax Act of 1981 provides for adjustments to the income tax structure based on the change in the CPI in order to prevent inflation-induced tax rate increases. These adjustments, designed to offset the phenomenon called "bracket creep", are to be calculated initially in 1984 and reflected in the 1985 tax schedules.

II. Sponsors

The CPI is collected, analyzed and published monthly by the Bureau of Labor Statistics. The Census Bureau under contract to BLS collects two surveys, the expenditure survey and the Point of Purchase survey which are used to construct sampling frames for selecting the item and outlet sample for the CPI.

III. Sample Design - General

The most recent major revision of the CPI was completed in 1978. This revision introduced probability sampling procedures at all levels of sampling including within outlet selection of items. It incorporated new expenditure weights from the 1972-73 Consumer Expenditure Survey, new retail outlet samples from the 1974 Point of Purchases Survey, and population data from the 1970 census. It also introduced a second index, the more broadly based CPI for All Urban Consumers (CPI-U) , which took into account the buying patterns of professional and salaried workers, part-time workers, the self-employed, the unemployed, and retired people, in addition to wage earners and clerical workers. The two indexes differ chiefly in the weighting used.

In January 1983, the BLS changed the way in which homeownership costs are measured. A rental equivalence method replaced the asset price approach to homeownership costs for the CPI-U. In January 1985 the same change will be made in the more narrowly defined index constructed for the Wage earners and clerical workers (CPI-W). The central purpose of the change was to separate shelter costs and the investment component of homeownership so that the index would reflect only the cost of shelter services provided by owner-occupied homes.

Several key concepts indicate the nature of the Consumer Price Index and guide the way in which it is calculated.

1. Prices and Living Costs. The CPI is based on the prices of food, clothing, shelter and fuels, transportation fares, medical services, and the other goods and services that people buy for day-

to-day living. It is constructed in accord with statistical methods that make it representative of the prices of all goods and services purchased by consumers in urban areas of the United States. Price change is measured by repricing essentially the same market basket of goods and services on monthly or bimonthly time intervals and comparing aggregate costs with the costs of the same market basket in a selected base period. The longitudinal aspect of the survey is the month to month linkage of the sample of item/outlet specifications (quotes) and their price, size and quantity for the given quote.

2. Weights and relative importance. The weight of an item in the index is derived from a survey of consumers which provides data about the dollar amount spent for consumer items during the survey year. In a fixed weight index, such as the CPI, the implicit quantity of any item used in calculating the index remains the same f rom month to month (for example, the number of gallons

of gasoline) . This should not be taken to mean that the relative importance, of gasoline in the average consumer's budget remains the same. Relative importances change over time because they reflect the effect of price change on expenditures. Items whose prices rise faster than the average become relatively more important.

3. Sampling. Since it is impossible to obtain prices for all expenditures by all consumers, the CPI is constructed from a set of samples not all of which are longitudinal in nature:

a. A sample of areas selected from all U.S. urban areas.

b. A sample of families within each sample area for expenditures of consumers, this sample need not be longitudinal, but linkage of records from a series of interviews was used.

c. A sample of outlets from which these families purchase goods and services. A household survey which is used to identify and construct the sampling frame of outlets is not longitudinal, however, the sample of outlets selected from this frame is longitudinal.,

d. A sample of items for the goods and services purchased by these families. This is the primary longitudinal component of the CPI.

It is from these samples that weights are developed and data are obtained for the monthly calculation of the index. Specifics for each sample or sampling stage are described as follows:

A. CPI Area Design

Pricing for the CPI is conducted in 87 sample geographic areas. Eighty five strata were defined by combining similar PSU's according to the following 1970 Census characteristics:

1. region, population size, SMSA versus non-SMSA
2. percent population increase from 1960 to 1970
3. major industry
4. percent nonwhite
5. percent urban

This area design resulted in 29 strata with one pricing area per stratum and 58 non-selfrepresenting strata. Twelve publication areas consisting of three city-sizes (non-selfrepresenting SMSA's of over 388,000 population, SMSA's less than 388,000 population, and non-SMSA urban areas) crossed by four Census regions were defined along with the 29 local areas to provide estimated indexes for all urban areas of the country. Each of the twelve region, city size publication areas contained four, six or eight strata. In addition special supplementation was made to support publication for Denver.

B. Expenditure Survey Sample Design

In 1972-73 two household surveys, a Diary and an Interview Survey were conducted by the Census Bureau for BLS to collect expenditure information for consumer units. The sampling unit for these surveys was a housing unit. The reporting unit was a consumer unit which was defined to be (1) a group of two or more persons, usually living together, who pool their income and draw from a common fund for their major items of expense, or (2) a person living alone or sharing a household with others, or living as a roomer in a private home, lodging house, or hotel, but who is financially independent-that is, income and expenditures not pooled with other residents. Never married children living with parents always were considered members of the consumer unit. The eligible population included the civilian noninstitutional population of the United States as well as that portion of doctors' and nurses' quarters of general hospitals. Armed forces personnel living outside military installations were included in the coverage while armed forces personnel living on post were excluded. Also excluded from eligibility were persons living in college dormitories, fraternity or sorority houses, prisons, monasteries, aboard ships, or in other quarters containing five or more unrelated persons.

The first component was a Diary Survey completed by respondents for two consecutive one week periods. The objective of the Diary

Survey was to obtain expenditure data on small frequently purchased items which are normally difficult to recall. These items include expenditures for food and beverages, natural gas and electricity, gasoline, housekeeping supplies, non-prescription drugs, medical supplies, and personal care products and services. Consumer units were asked to list all expenses during the survey period. Data on income and family characteristics also were collected. The sample of housing units was balanced across areas and time of year. The records of the two consecutive one week periods for each consumer unit were linked to create two week levels of expenditure.

The second component of the CE, called the Interview Survey, was a panel survey in which each consumer unit in the sample was interviewed every three months over a fifteen month period. This survey was designed to collect information on major items of expense as well as on income and family characteristics. Items reported on the interview survey included expenditures for the following: housing, household equipment, house furnishings, vehicles, subscriptions, insurance, educational expenses, clothing, repair and maintenance of property, utilities, fuels, vehicle operating expenses and expenses for out of town trips. The final interview in the fifth quarter provided the regularly recorded expenses plus information on homeownership costs, work experience, changes in assets and liabilities, estimates of consumer unit income and other selected financial information. The quarter records for each consumer unit were linked to form annual records for each consumer unit. Only consumer units responding in at least the fifth interview were used to form these "linked" records of annual expenditures for estimation.

The samples of consumer units for the CE were selected as follows. For both the diary and interview survey the nation was stratified into 216 geographic strata using stratification variables defined for the Current Population Survey of the Census Bureau. Thirty of these areas were designated as selfrepresenting. Half of the housing units in each self-representing area were covered in the first survey year and half in the second survey year. The 186 equal sized non-self- representing areas were divided into two 93-area groups. One sample area from each of the 93 groups was in sample in each of the two survey years. Each sampling area was randomly selected proportional to population from each of the 186 strata.

1. Interview Survey

The universe for sample selection was the 1970 Census 20% sample data file. A sample of 12,613 housing units was designated for the 1972 Interview Survey component, and 13,014 housing units for the 1973 Interview Survey. For the first year 11.1 percent were vacant, nonexistent or ineligible and the refusal rate was 10.3

percent of the designated sample. Interviews were completed in 9914 units. For the second year 12.9 percent was vacant, nonexistent and ineligible with a refusal rate of 9 percent. Interviews were completed in 10158 units.

At the time of selection, housing units for the Interview Survey within a PSU were distributed by month within the quarter to allow for data collection throughout each quarter. Each sample unit was visited once each quarter, at

approximately the same time in the quarter, and each consumer unit within the household was interviewed. Data from previous quarters were available for the interviewer to use in bounding expenditure reporting. Bounding is an interviewing technique which unduplicates expenditures reported in the previous interview from the current interview. The type of expenditures reported during each interview varied since the recall periods varied from three months to one year. Housing, major equipment, automobiles, subscriptions and insurance were annual recall items. A semi-annual recall period was used for minor equipment, house furnishings, renting and leasing of vehicles, and education. The following sections were covered each quarter: repair, alterations, and maintenance of owned property; utilities, fuel, and household help; clothing and household textiles; equipment repairs; vehicle operating expenses; and out-of-town trips. Interviewing was conducted with any person available in the consumer unit; no attempt was made to interview all persons in the consumer unit, that is proxy responses within a consumer unit were used. Proxy responses for persons away at school was the source for some of the college members of a consumer unit.

2. Diary Survey

Again the universe for sample selection was the 1970 Census 20% sample data file. A sample of housing units was selected from this Census file for each year of the diary survey. Approximately 14,590 housing units were designated and 12,661 eligible for the 1972 Diary component, and about 15,210 designated and 12,999 eligible for the 1973 Diary component. These numbers included an augmented sample of households which were to be visited during the four week period preceding the end of the year holidays. Each housing unit was visited twice, once at the end of each week of the two week survey period. For the first year the eligible response rate was 80.1% and 89.9% for the second year.

IV. CPI Survey Design and Content

The primary longitudinal samples for the CPI is the sample of item/outlet specifications and their respective prices, which are obtained every month. BLS collects prices for the Food, Commodities and Services, Rent, and Property Tax components of the CPI. These prices are collected monthly or bimonthly in all 87 areas. Each one of these components has a separate survey with its own sample design. Data used for the Mortgage Interest and House Prices components of the CPI are not collected by the Bureau but are obtained from outside sources such as FHA and FHLBB.

The Point of Purchase Survey (POPS) is the source of the outlet sampling frames for about 60% of the CPI items by expenditure weight. The items not covered by the POPS are grouped together under the heading non-POPS and include rent, property tax, mortgage interest, house prices, utilities, transportation, insurance, and several miscellaneous categories. These sample designs are not described here except rent.

1. Point of Purchase Household Survey - Frame Source

In the spring-summer of 1974 a household survey, the Point of Purchase Survey, was conducted by the Census Bureau for BLS to provide the sampling

frame of outlets for food and most commodities and services to be priced in the CPI and to provide demographic data for classification of the households reporting an expenditure for an outlet. The survey was conducted in the 85 PSU's defined for the CPI. The commodities and services for which sampling frames were developed in each PSU included food, apparel, drugs, personnel care items, household furnishings and housekeeping supplies, beverages, most medical services, sports equipment, gasoline and automobiles, and automotive parts and services. Expenditures, name, and location of the place of purchase were collected for approximately 100 relatively broad categories of expenditures with reference periods of one week to two years depending on the expected frequency of reporting. To control the expected number of responses received from a household and minimize respondent burden two groups of categories were defined; one set given to 1/4 of the sample households and the second set given to 3/4 of the sample households. The combination of sample size of the households asked a category and the reference period for a given POPS category was designed to generate approximately 6 to 12 not necessarily unique outlets reported for a given PSU/POPS category.

For POPS the national sample size was 23,000 designated housing units. Since separate frames of outlets were required for individual CPI pricing area (PSU's), the sample is not self-weighting across PSU's, but within a PSU, the households are selected with a uniform probability.

2. CPI Outlet Sampling Procedures

When a sample ELI was selected a specific POPS category was identified for outlet selection. In self-representing areas, sample households were divided into two independent groups by the first stage order of selection. This defined two frames of outlets for outlet selection to support variance estimation. The following approach was used for outlet selection for frames developed from the POPS and CPOPS Survey.

A systematic selection of outlets reported for a given POPS category for the W population was made where the measure of size for each outlet was proportional to the average daily expenditure reported for the outlet by all consumer units in the W population.

Before January 1982, the outlets for the U population were then selected using a conditional probability technique to maximize the overlap between outlets. The sample outlets for the U population were then selected by a repeat of the systematic selection using the new measures of size. After January 1982 the collection of prices for the W population was discontinued. The sample outlets are now selected systematically with probability proportional to average daily expenditure of the U population.

All outlets reported by CPOPS sample families in any sample area are eligible for pricing. However, BLS restricts pricing of outlets to be within a 25 mile radius of a given sample PSU unless 10 or more designated items are identified in some clustered area beyond the mileage limitation. If this is the case, there is no mileage limitation and all items in the clustered area are priced.

The non-POPS categories were excluded from the POPS either because existing sampling frames were adequate, or it was felt the POPS would not yield an adequate sampling frame.

Each non-POPS commodities and services item has its own sample design. For each item, the frame consisted of all outlets providing the commodity or service in each sample area. A measure of size was associated with each outlet on the sampling frame. Ideally, this measure of size was the amount of revenue generated by the outlet by providing the item to the !CPI U population in the sample area. Whenever revenue was not available, an alternate measure of size, such as, employment, number of customers, or quantity of sales was substituted. Since no measures of size could be determined strictly for the w population, a single sample of outlets and quotes was selected for estimating the index for each population. All samples were selected using the systematic sampling technique with probability proportional to the measure of size available.

a. CPI Sample Items

The basic CPI item structure is an follows: The seven, major groups (food, housing, apparel, transportation, medical care, entertainment and personal care) are broken into 68 expenditure classes (ECIS) (such as auto repair). Within each EC, expenditures are grouped into one or more item strata (such as body work, power plant repair, component repair, and maintenance and service). There are a total of 265 item strata within each item strata, one or more substrata, called Entry Level Items (ELI's) are defined. There are a total of 382 ELI's. ELI's are the ultimate sampling units for items as selected in the BLS Central Office. They are used in the field by the data collectors as their initial level of item definition within an outlet. An ELI is assigned to one and only one POPS or Non-POPS outlet category.

Four regional market basket universes were tabulated into the item strata structure from the Diary and Interview surveys to reflect regional differences within each of the four regions (Northeast, North Central, South, and West) eight independent samples of ELI's were selected for each item stratum. Thus, eight samples of ELI's were selected for each region and for each population-thirty-two sample selections nationally for each population. Each CPI PSU was assigned one or two of the eight item samples from the corresponding region for pricing. Self-representing published areas were assigned two independent item samples and each non-self-representing area was assigned one item sample. These independent item samples were designed to accommodate variance estimation for the CPI. A given item sample for all item strata assigned to a given PSU is called a half-sample. The sample of ELI's and appropriate POPS categories are merged to create specific outlet/item samples.

b. Within Outlet Selection for Specific items

For each ELI, whether in a POPS or Non-POPS category, the selection of a specific store item by a data collector is performed using multi-stage probability selection techniques with measures of size proportional to percentages of dollar sales usually provided by the respondent for the outlet.

To perform this operation, the data collector is provided with a checklist that includes all the descriptive characteristics which are believed to identify the items of the ELI and determine or explain price differences for all items defined within the ELI. In addition, the data collector is given the definition of the ELI, suggested stages of groupings of items to aid in

quickly selecting a specific store item and a series of worksheets on which to define the categories of items, post the probabilities and identify the next category within which to select the specific store item by use of the random number table on the worksheet.

In developing this procedure, it was necessary to provide the data collector with several alternative methods for defining the categories and obtaining the percentage of dollar sales or approximations to those sales. The procedures developed to obtain the proportion of sales were:

a. Obtaining the proportions directly from a respondent.

b. Ranking the categories by importance of sales and then obtaining the proportions directly or using preassigned proportions.

c. Using shelf space to estimate the proportions where applicable.

d. Using equal probability if all else fails.

To define the categories, direct responses from the respondent as to what he sells or an inventory technique was used.

The procedures make possible an objective probability sampling of items throughout the CPI. They also allow broad definitions of ELI's so that the same tight specification need not be priced everywhere. The wide variety of specific items greatly reduces the within item component of variance, reduces the correlation of price movement between areas, and allows a substantial reduction in the number of quotes required to obtain the same precision as the pre-1978 index. A second important benefit from the broader ELI's, along with the POPS categories, is a significantly higher probability of finding a priceable item within the definition of the ELI within the sample outlet. Procedure a) was used approximately 60% of the time, procedure b) was used about 30% of the time, procedure c) about 7% and procedure d) the remainder.

Once the sample of items in the sample PSU's are identified, the price for the specification which define the items within the sample outlets are priced on a monthly or bimonthly basis. This continues for a minimum of a 5 year period and is the basis for measuring price change for the CPI. This time series for each individual specification is the longitudinal element of the CPI.

C. Sample Maintenance

Since 1977, the Bureau has sponsored a Continuing Point of Purchase Survey (CPOPS) also conducted by the Census Bureau. This survey is aimed at producing current data on outlets. The CPOPS has been expanded from the original 100 categories of expenditures included in the POPS to 134 categories of which 102 categories are asked from each of two equal size panels. This survey is conducted each year in one fifth of the 87 PSU's on a rotating basis. From the results of this household survey, new samples of outlets and item specifications are rotated into the CPI data collection to replace the old sample of outlets and items priced for the CPI in a given area.

d. Response Rates

A sample of 24,278 outlets were designated from the original

POPS survey for CPI pricing. The out-of-scope response rate was 12.6 percent. There were

1,649 with non-responses resulting from no contact, refusals, or temporary agencies. This non-response rate for designated sample units was 6.8 percent. Thus the response rate was 93%. Each year one-fifth of the sample areas have all of the outlets reselected for repricing. Approximately 7300 outlets are selected of which 11.8% are out of scope and the response rate has been 95% from those outlets which have sample items available to price. An annual attrition rate for outlets has been 3.3%. In addition for the outlets which remain in sample, the average annual item substitution rate for items within outlets has been 6.2%. Substitution occurs because an item selected for sample is modified or no longer available and the field representative obtains a description and price for an item most similar to the original item selected from the outlet.

V. Rent Survey

A. Sample selection

The current CPI rent index is based on a sample of approximately 23,000 rental units, allocated among the 87 PSU's. The units were selected from two universes, a stratified multistage, systematic, self-weighting area sample of housing units built before 1970 and a continuously updated sample of newly constructed units. The Bureau of the Census provides the sample of new construction units from building permits. Approximately 2,000 units have been obtained from this source as of 1982.

Using an area segment sampling approach, 19,000 rental units were selected from 6,422 area segments. There has been an attrition of about 2,000 units due to conversions to owner housing. This sample has been augmented with approximately 1,500 new segments and 4,000 rental units to minimally support the rental equivalency concept of homeownership. This augmentation followed a process similar to the original area segment sampling approach.

B. Data Collection

In order to collect the monthly information necessary to calculate the rent index, the sample is divided into six panels of approximately 3,800 units each. The units in each panel are visited twice a year on a six month cycle. The information collected includes the rents paid for the current month and the previous month, information on extra charges and reductions, a description of the unit, and the facilities included in the rent. The latter questions are used to make quality adjustments to the calculated rents in order to assure that the rent change measured is for a set of units of a consistent quality. Data collection is by personal visit or telephone to tenants or property managers.

For the CPI Rent sample the response rate for occupied in scope units is 88 percent.

VI. Scope and Calculation

A. Index and Non-Rent Estimation

Prices used in calculating the index are collected in 87 urban areas across the country from about 24,000 retail establishments.

Prices of food, fuels, and a few other items are obtained every month in all 87 locations. Prices of most other commodities and services are collected every month in the five largest urban areas and every other month in other areas. Prices of most goods and services are obtained by personal visits. Some repricing for selected easily identified commodities are obtained by telephone and a mail questionnaire is used to obtain electricity rates.

In calculating the index, price changes for the various item strata in together with urban area weights which represent each market basket are averaged sent their importance in the spending of the appropriate population group. Local data are then combined to obtain a U.S. average. Separate indexes are also compiled by size of city, by region of the country, for cross-classifications of regions and population-size classes, and for 29 local areas. The estimation for monthly item strata level price relatives ($R_{t,t-1}$) is the ratio of two long term relatives for time t and t-1.

$$R_{t,t-1} = \frac{R_{t,O}}{R_{t-1,O}}$$

Each long term relative is calculated as a weighted sum of individual items price relatives

$$R_{t,O} = \sum_{i=1}^{m} \frac{W_i}{M} \frac{P_{ti}}{P_{Oi}}$$

where

R.t,o is the long term estimate of price change for a set of items representing the item strata

P.ti is the price at time t for item i

P.Oi is the price at time 0, the base period, for item i

W.i is an estimate of expenditures for the ELI contained in the item strata for which the items are a sample

M is the number of eligible sample prices in the ELI

The index each month is a weighted average of the price relatives divided by a base expenditure (C.O). The weights (C.t-1,i) of the index are estimates of expenditure for each item stratum which reflect buying patterns of a given reference period and all price change up to the previous month:

$$I_{t,0} = \sum_{i=1}^{m} \frac{C_{t-1,i} R_{t,t-1,i}}{C_{0}}$$

B. Rent Estimation

Estimates of the monthly rent price relatives for each market basket are calculated using special cost weights and 1- and 6-month, estimates of rates of change.

Let S.1 be the set of units interviewed in time t in a market basket which has rent values for time t and t-1, and S.6 be this set of units interviewed in time t in a market basket which has rent values for times t and t-6. The rents for the ith unit in a market basket for the given time period are represented by r.iT where T-t, t-1, or t-6. The l- and 6-months rates of change,

R.t,t-1 and R.t,t-6, are calculated by:

```
                 ä     r.it   W.i                  ä   r.it    W.i
               ieS.1               and R.t,t-1    ieS.6
R.t,t-1   =    -----------------                  -----------------
                 ä   r.it-1  W.i                   ä   r.it-6  W.i
               ieS.1                              ieS.6
```

where W.i reflects the probability of selection adjusted for nonresponse.

Using R.t,t-1 and R.t,t-6' a composite estimate is made of a current month's cost weight CW.t for the market basket:

$$CW.t = P \quad R.T,T-1 \ \ CW.t-1 + (1 - P)R.t,t-6 \quad CW.t-6'$$

where P = .65. The value of P was based on simulations of weighted averages of 1- and 6-month rent relatives designed to minimize variances.

A final 1-month estimate of rent price change for the particular market basket is

$$R_{t,t-1} = \frac{CW_t}{CW_{t-1}}$$

C. Rental Equivalency

In January 1983, BLS will begin measuring the housing component of the CPI-U using the rental equivalency method which assumes the cost of homeownership is the amount which would be paid to rent an equivalent home. Rental equivalency will be measured using a sample of rental units with new weights assigned to each rental unit which reflect the number of homeowner units in the universe for which the rental unit is equivalent. The rent component of the CPI will continue to be measured in the usual way.

After 1986 rental equivalency will be measured using a sample of owned units. Rent change will be determined for these units by matching the owned units to equivalent rental units based upon unit and neighborhood characteristics. Using estimated owners rents, monthly change for rental equivalency will be calculated in a fashion similar to that used to calculate the current rent index.

VII. Data Products and Analysis

The monthly CPI is first published in a news release during the fourth week following the month in which the data are collected. (The index for January is published in late February.) The release includes a narrative summary and analysis of major price changes, short tables showing seasonally adjusted and unadjusted percentage changes in major expenditure categories, and several detailed tables. Summary tables are also published in the Monthly Labor Review the following month; shortly thereafter, a great deal of additional information appears in the monthly CPI Detailed Report.

Seasonally adjusted data are presented in addition to unadjusted data because they are preferred for analyzing general price trends in the economy. They eliminate the effect of changes that normally occur at the same time and in about the same magnitude every year, such as price movements resulting from changing climatic conditions, production cycles, model changeovers, holidays, and sales. Seasonal factors used in computing the seasonally adjusted indexes are derived by the X-11 Variant of the

Census Method II Seasonal Adjustment Program and are reevaluated annually.

The data collected is item descriptive data plus the price, size and quantity of the item being priced. Longitudinal analysis is specifically related to determination of degree of price change and trend for a given commodity sector and explaining the reasons for the change for both the short and long term by examining the micro data and ancillary information for the locale and the nation. In addition, studies are conducted to assess the impact of government policy changes or changing economic conditions on the index. The techniques used are regression, distribution analysis and simulation.

VI. Limitations of the Index

The CPI is not an exact measure of price change. It is subject to sampling errors which may cause it to deviate somewhat from the results which would be obtained if actual records of all purchases by consumers could be used to compile the index. These estimating or sampling errors are limitations on the precise accuracy of the index rather than mistakes in the index calculation. The accuracy could be increased by using much larger samples, but the cost is prohibitive. Furthermore, the index is believed to be sufficiently accurate for most of the practical uses made of it.

Another kind of error occurs because people who give information do not always report accurately. The Bureau makes every effort to keep these errors to a minimum, obtaining prices wherever possible by personal observation, and corrects errors whenever they are discovered subsequently. Precautions are taken to guard against errors in pricing, which would affect the index most seriously. The field representatives who collect the price data and the commodity specialists and clerks who process them are well trained to watch for unusual deviations in prices which might be due to errors in reporting.

The CPI represents the average movement of prices for two specified populations but not the change in prices paid by any one family or small group of families. The index is not directly applicable to nonurban workers and others not included in the samples. The index measures only the change in prices and none of the other factors which affect family living expenses, such as changes in the size of the family or changes in buying patterns. Nor does it reflect consumption, such as fringe benefits.

Area indexes do not measure differences in the level of prices among cities; they only measure the average change in prices for each area since the base period.

Although the CPI has been called a cost-of-living index and used at times as if it were one, there are important conceptual differences between a price index and a cost-of- living index. A true cost-of-living index would take into account not only price changes but also changes in the market basket as consumers adjust their purchases to changes in the relative prices of what they buy. Thus,, during a period of rising prices, a cost-of-living index might rise more slowly than a price index if consumers substitute cheaper items for more expensive ones, or generally reduce expenditures on higher priced items in their budget. However, an index such as the CPI' does not directly reflect such consumer behavior, since the quality and the implicit quantity weights of the items represented in the CPI remain constant. The index indicates what it would cost to maintain the same level of living, not what consumers actually spend on their living costs. What consumers actually spend may reflect a decision to accept a lower standard of living in order to keep living costs from rising.

There are other differences between the two types of index. For example, the CPI includes only the cost of sales and excise taxes that are included in the purchase price o f goods and services, but not income taxes, whereas a cost-of-living index would include both sales and income taxes.

EMPLOYMENT COST INDEX CASE STUDY

I. Purpose

The Employment Cost Index (ECI) measures change in total employee compensation and has been designed as a principal Federal economic indicator by the Office of Management and Budget. The ECI is used in monitoring the effects of monetary and fiscal policies by enabling analysts and policymakers to assess the impact of labor cost changes on the economy, both in the aggregate and by sector. The limitations of the index must be kept in mind. Because the ECI is an index, it only measures change in employee compensation; the index is not a measure of the total cost of labor. Not all labor cost (e.g., training expenses, retroactive pay, etc.) fall under the ECI definition of compensation.

II. Sponsors

The Bureau of Labor Statistics developed the ECI in 1975 to provide a comprehensive measure of employee compensation. The initial design was started in the early 70's by the Office of Wages and Industrial Relations and the Office of Survey Design of BLS. All data collection and data processing is provided by Bureau staff.

III. Sample Design

A. Private Sector Sample Design

A principle concern of the ECI sample design is to provide an ongoing sample that in some sense represents an outgoing current universe. ECI accomplishes this with what is called replenishment groups. A replenishment group is an establishment sample of SICs which replaces a segment of the current sample. A new replenishment group is introduced each quarter until the entire sample has been replaced; after which, the cycle is repeated (currently every four years). The quarterly replenishment groups each have, approximately, an equal number of establishments. This equality reduces the disruption in the quarterly estimates and is within resource constraints. A replenishment group collection cycle begins every three months and the new sample is introduced into the ECI estimates after the section update.

1) Description of the Private Section Establishment Selection

Each replenishment sample is composed of a number of related two-digit SIC subsamples. Within each SIC, the frame (Unemployment Insurance File) may be sorted by Census Region, employment or establishment name. A sample of 450 establishments is selected probability proportionate to employment for the entire replenishment group. Systematic samples of about 300 establishments comprise the main replenishment sample. The remaining 150 establishments are selected for several supplemental groups. The supplemental.

groups are held in reserve in case additional sample is required if a larger than expected number of out-of-scope is obtained. To enable variance estimation by replication techniques, the establishments are assigned to two half-samples.

2) Description of the Occupation Selection

To measure Major Occupation Group (MOG) compensation change, the Occupational Universe (currently based on the 1970 Census occupations) is partitioned into the MOGS, such as professionals, technical workers, etc. Each MOG may be further partitioned into Entry

Level Occupations (ELOs), such as Teachers.

There are usually 9 to 13 ELOs, which represent all occupations within an SIC. For each ELO found in the establishment, data is collected to represent that ELO. During the initial visit to a sample establishment each detailed establishment occupation is matched into one of the ELOs. Then a probability proportionate to employment selection is made within each ELO, selecting one specific occupation. Data for wages and benefits is then collected for each of the selected detailed establishment occupations.

B. Public Sector Sample Design

The public sector sample has been fixed since June 1981, when it was introduced. There is no public sector replenishment system because of the lack of updated frame. An easily accessible frame does not exist for State and local governments.

1) Public Sector Establishment Sample Design

The public sector frames were divided into four parts: schools, hospitals, State and large local governments (all SICs except schools and hospital), and small local governments.

a. Schools:

The public elementary and secondary schools frame, (SIC 821) as well as the higher education (SIC 822) frame, came from 1973-74 National Center for Education Statistics (NCES) listing of all State and local schools.

Establishments were stratified by 3-digit SIC; then a sample was selected with probability of selection proportionate to enrollment within the school. A first phase mail survey was conducted to determine ELO employments for the selected schools. Using these ELO employments to obtain measures of size, the second stage sample of 206 establishments employing a two-way controlled selection technique controlling on respondent burden and the number of designated quotes within each selected ELO was selected.

b. Hospitals:

The hospital frame was the 1976 Health, Education and Welfare (HEW) list of public hospitals. The hospital survey design did not include a first phase occupational survey. Public hospitals were stratified by Census region and ownership and selected systematically using probability proportionate to employment. The occupation selection was essentially a systematic sample (equal probability) within each establishment. The 106 establishments in the final sample were then requested to supply data from the appropriate occupations.

c. State and Large Local Governments

No universe listing of establishments was available for State and large local governments. A refinement survey was used to develop a sampling frame. The local government jurisdictions in the refinement survey (cities, counties, special districts, etc.) were selected from 1972 Census of Government file provided by the Bureau of the Census. Only jurisdictions with more than 100 employees were included in the refinement survey (see "small local governments" below). The 3,729 local jurisdictions were stratified into size class/Census region strata. Forty-six jurisdictions were selected probability proportionate to employment.

In addition, sixteen States were selected probability proportionate to employment land included in the Refinement Survey.

Once the refinement was completed, a probability proportionate to employment sample of 780 refined units were selected for a first phase occupational employment survey. Occupational employments were requested for nine occupational groups within each of the 780 units. The final sample includes 350 units.

d. Small Local Government

Due to their small size (units with less than 100 employees), no refinement or first phase survey was done for small local governments. Instead, the list of small local governments was stratified by Census Region and then a probability proportionate to employment sample of 30 units was selected. Any refinement required was accomplished by BLS field representatives at the time of collection.

IV. Survey Design and Content

A. Design

1) Reporting Unit

The ECI reporting unit is the physical location of a business (establishment). Sometimes data can only be collected for a unit which is larger than the original designated establishment, Usually this is acceptable and a weighting adjustment is made later. It is also possible that data is much more accessible at a finer level than an establishment; in this case, subsampling procedures are available to randomly select a subunit.

2) Following Movers

If the collection unit is essentially unchanged after a physical move, then it is followed provided it remains within the same State.

3) Weighting

The weights for each establishment/ELO is the reciprocal of the selection probability times the ELO employment. There is also a nonresponse adjustment factor applied to the weight.

4) Interview Schedule

Each establishment reports wage and benefit data four times a year (March, June, September and December). The typical private sector establishment will be included in the survey for a four year period, at which time the sample is replaced.

Currently, there is no definite date when the public sector sample will be replaced.

5) Interview Mode

The initial data collection is always a personal visit. During subsequent quarters a mail update form is used. When necessary, telephone calls are made to obtain required data.

6) Questionnaire

There are two basic types of ECI collections -- initiation and quarterly update collections. During the initiation, the field representative selects a detail establishment occupation to represent each ELO. Once the establishment occupation is selected, benefit usage, benefit plan, wage and work schedule data are collected for each selected detail establishment occupation.

During the quarterly update, wage data and benefit plan change data are collected. When a benefit plan changes, the new plan is incorporated into the database using the initiation usage.

B. Content

The Employment Cost Index is a relatively new Bureau of Labor Statistics survey measuring the change in the employer cost of employing workers. When the ECI first started its publication in December 1975, it measured quarterly wage change covering the private non-farm sector, excluding Alaska, Hawaii and private households. Publications included overall National, Major Industry Division (MID) like wholesale trade, manufacturing and services; Major Occupation Group (MOG) like Professionals, Managers and Clerical Workers; Census Region (Northeast, South, North Central and West); Union/Non Union and Metropolitan/Non-Metropolitan Area quarterly change numbers. Currently, the ECI is an index measuring total compensation change covering the total non-farm civilian sector excluding private household and the federal government. Compensation is composed of wages and twenty-three benefits (hours related benefits, such as vacation; supplemental pay, such as shift differentials, insurance, such as health benefits; pension and legally required, such as social security).

The National series (Overall National, MID, MOG indices) use Laspeyres (fixed weight)/industry/occupation estimates. For each of the non-National series.1/ (Census Region, Union/Non Union and Metropolitan/Non-Metropolitan), estimates (e.g.,union/industry/occupation) are obtained by allocating the fixed weight industry/occupation estimates using current sample data; so that the non-national series cannot be considered Laspeyres.

V. Response

A. Determination of Private Sector Replenishment Cycle

Assuming the sample is completely replaced after n, 2n, 3n,..., quarters and that the response and attrition rates are equal across replenishment, then the response rate obtained after n quarters should be maintained each quarter thereafter. We call this the maintainable response rate. The determination of the appropriate time length for the complete replenishment cycle can be made by computing the maintainable response rates for various cycles and comparing the rates.

To compute the maintainable response rate, the following wage information from the original sample is used:

proportion of initial sample in scope, 0.85;

proportion of initial in scope sample responding, 0.82;

proportion of sample remaining each quarter, 0. 98; and

.1/ For an economic interpretation of the non-national estimates see:

Estimation Procedures for the Employment Cost Index, G. Donald Wood, Jr., Monthly Labor Review, May 1982.

number of establishments required at the end of

the replenishment cycle 2000.

Using the above information the following table on quarterly sample size and maintainable response rate is determined.

Replenishment Cycle (Years)	Number of units initiated per quarter	Estimated Maintainable response rate (wages)
2	385	0.76
3	267	6.74
4	208	0.71

Considering the initial work required introducing an establishment into the survey, a two year cycle was not considered cost effective. A 0.71 wage response rate with a four year cycle is lower than desired considering the fact that the benefit response rate would be closer to 0.6 than to 0.7. A three year cycle would keep respondents in the survey for a reasonable length of time and provide a benefit response rate at least close to 0.65. Therefore, the initial decision was to proceed with a three year cycle.

After the first year of replenishment samples, it became apparent that field resource constraints would not allow a three year cycle. We are currently working on a four year cycle.

B. Public Sector

The Public Sector does not have a replenishment system in place at this time. The initial response rate, in June 1981, was 81%. Since then the attrition rate has averaged 0.3% each quarter. These numbers are considerably better than the private sector. Even though there is no replenishment system, the response rate does not decrease quickly. In addition, the number of establishment births and deaths within the public sector should be much less than the number within the private sector. The universe, therefore, should remain relatively stable until 1990.

C. Imputation Schemes

There are three levels of imputation in the ECI. The first level is a weight adjustment to compensate for the initial nonresponse. The second level is an imputation for temporary nonrespondents. (Those establishments that will respond next quarter, but for some reason cannot respond this quarter). This imputation is done at the item level. Its purpose is to serve as a link for periods when there is a response. The third level of imputation is at the estimation cell level, whenever there is no data for the entire estimation cell. This imputation assures that the same cells are being compared each quarter.

VII. Data Product

At the present time, no public use tapes of micro ECI data are available. The only data available to researchers are that contained in the quarterly news release which is available on Labstat. The feasibility of developing a public use tape is being explored.

CASE STUDY 4

NATIONAL LONGITUDINAL STUDY

OF THE HIGH SCHOOL CLASS OF 1972

I. Purpose:

The basic purpose of NLS-72 is to provide data on the experiences that affect the development and attainment of a current generation of young people. Specifically, this study provides data on:

- the transition of young people from high school to postsecondary education
- the transition from high school to the world of work,
- persistence in postsecondary education (as opposed to dropping out),
- the transition from postsecondary education to the world of work.

II. Sponsor

NLS-72 has, since its inception, been sponsored by the National Center for Education Statistics (NCES) within the U.S. Department of Education.

The principal contractors who have played major roles in WLS-72 are:

1. Education Testing Services (ETS) -- Base-year survey in

1972.

2. Research Triangle Institute (RTI) -- First four follow-up surveys 1974, 1975, 1977, and 1980.

3. National Opinion Research Center (NORC) -- Fifth follow-up survey and Postsecondary Transcript Study in 1984-85.

III. Sample Design

The sample design for NLS-72 is a stratified multistage probability sample of students from all schools, public and private, in the 50 states and the District of Columbia, which contained a 12th grade class. Stratification variables were: type of control (public vs. private), geographic, region, enrollment size, proximity to a college, percent minority, income level of community, and urbanicity.

The original sample design for the base-year survey called for selecting a probability sample of 1,200 schools from the population of schools with a 12th grade, and within each school random selecting 18 seniors. Since 231 of these schools refused to participate and 21 had no seniors enrolled, the number of schools actually participating was 948. The number of students participating was 16,683.

At the time of the first follow-up, in 1974, 205 of the nonparticipating schools were induced to participate and former seniors from those schools were administered retrospective surveys. Ultimately the reconstituted base-year sample consisted of 22,652 students from 1,318 schools.

IV. Survey Design and Content

In the base-year survey, questionnaires and cognitive tests were administered to groups of students in each participating school. Information on courses taken and grades earned was extrated from school records.

Follow-up surveys have been conducted primarily by mail but when repeated reminders failed to elicit a response, resort was had to personal interviews, either by telephone or face-to-face. About one third of the mail respondents in each follow-up survey were telephoned to resolve response inconsistencies.

The fifth follow up, which is now in the field-test stage, is being funded by NCES with the help of a consortium of interested agencies. It will also be conducted primarily by mail. To reduce costs only a subsample of the original sample will be used.

The various questionnaires tap numerous content areas, including: background characteristics, cognitive ability, socioeconomic status, home background, community environment, relative importance of significant others, current and planned educational and occupational activities, school characteristics, performance in school, work performance and satisfaction, goal orientations, marriage and family, opinions of school, et al. A more detailed listing of survey content areas is displayed in the attached "Table 2."

The content areas for the fifth follow-up survey are being reduced somewhat in order to make room for certain new topics. Education and work history items are retained, however. In addition, special new questionnaires are included to be filled out by hose respondents who have become teachers, or parents.

V. Response Rates

As a result of extraordinary tracking efforts and intensive data collection activities, the response rate to the various student questionnaires' has remained quite high over the 12 years of RLS-72 operation. Student responses rates for each of the surveys thus far completed were:

Base year	87.8%*
1st FU	94.2%
2nd FU	92.1%
3rd FU	88.7%
4th FU	82.2%

* This figure is the percentage supplying data, based on all targetted students in participating schools in the original base-year survey. The corresponding figure for the reconstituted sample was 73.6%.

VI. Evaluations

To maximize the validity and reliability of the data, several procedures were followed:

1. For each of the surveys thus far completed, the student questionnaire was first pretested on a sample of 1971 seniors. (This will not be possible for the 5th follow up because tracing efforts for those students were not adequate to retain a sufficiently large subsample).

2. For the base-year survey, a reliability check was conducted in which 500 respondents were asked to reanswer 10 questions 3 months later.

3. For the base-year survey, a validity check was conducted by asking the parents of 500 students to confirm or correct the student's report of family income.

4. To improve the quality of mail responses, all questionnaires were checked for completeness and consistency. Respondents whose forms failed these edit checks were telephoned for clarifications.

VII. Data Products and Analysis

NCES makes all NLS-72 data files available to the public at cost. As each new data file becomes available, an Announcement to that effect is widely disseminated to potential users.

As of 1981, over 320 research reports based on NLS-72 data had been published. These are listed and annotated in the following publication: National Longitudinal Study of the High School Class 1972; Study Reports Update: Review and Annotation by M. E. Taylor, C. E. Stafford, and C. Place. Research Triangle Institute, June 1981.

Table 2. Content areas in each survey

Content category	Spring 1972 base-year survey	Fall-winter 1973-74 first follow-up survey	Fall-winter 1974-75 second follow-up survey	Fall-winter 1976-77 third follow-up survey	Fall-winter 1979-80 fourth follow-up survey
Constitutional factors	Sex, birth date, physical handicap	Sex, birth date	Sex, birth date	Sex, birth date	Sex, birth date
Ability	SAT, ACT, and test scores (vocabulary, reading, math, letter groups, mosaic comparison, picture-number)				Retest (vocabulary, mathematics)
Socioeconomic status	Parental income, education, occupation	Parental education, occupation			
Home background	Number of persons dependent on parents, number of siblings in college, objects in home, language at home, parental expectation	Parental encouragement	Birth order, number of siblings		
Community environment	Type of community, distance of home from postsecondary schools	Type of community where individual lives	Type of community where individual lives, mobility and reasons	Type of community where individual lives, mobility and reasons	Type of community where individual lives, mobility and reasons
Ethnicity	Race, religion		Race		
Significant others	Relative importance of family, peers, school personnel to secondary program, postsecondary plans, quality of counseling services				
Activity status	Activity plans for Fall 1973	Activity status in October 1972, October 1973	Activity status in October 1974	Activity status in October 1975, October 1976	Activity status in October 1977, October 1978, October 1979 and supplemental data for activity status 1972-76

(continued)

Table 2. Content areas in each survey — continued.

Content category	Spring 1972 base-year survey	Fall-winter 1973-74 first follow-up survey	Fall-winter 1974-75 second follow-up survey	Fall-winter 1976-77 third follow-up survey	Fall-winter 1979-80 fourth follow-up survey
Educational attainment	Educational plans (entry, financial support arrangement), factors interfering with attainment, school choice	Educational status (degree, certificate received), factors interfering with attainment	Educational status (degree, certificate received), factors interfering with attainment	Educational status, graduate school application and entry, factors interfering with attainment	Educational status, graduate school application and entry, factors interfering with attainment
School characteristics	Secondary school size, student-counselor ratio, ability grouping or tracking, racial composition, college-going ratio, school SES, teacher qualifications, school facilities, counseling services	Postsecondary school control and type	Postsecondary school control and type	Postsecondary school control and type	Postsecondary school control and type
School experience	Time spent on homework, program of study participation in remedial and special services programs, school quality, courses taken, instructional strategies	Program type, major study area, full-time, financial aid programs, program duration	Program type, major study area, full-time, financial aid programs, program duration	Program type, major study area, full-time, financial aid programs, program duration	Program type, major study area, full-time, financial aid programs, program duration
School performance	Grade average, extra curricular activities	Grade average, dropout, transfer, satisfaction with schooling, total credits earned	Grade average, dropout, transfer, satisfaction with schooling, total credits earned	Grade average, dropout, transfer, satisfaction with schooling, total credits earned	Grade average, dropout, transfer, satisfaction with schooling, total credits earned
Work status	Type of work, hours of work, work plans for years after graduation	Resources used for job hunting, work type, hours/week, reasons for not working	Resources used for job hunting, work type, hours/week, reasons for not working	Resources used for job hunting, work type, hours/week, reasons for not working	Work type, hours/week looking for work
Work performance and satisfaction	Factors in career selection	Income, pay, and work conditions, satisfaction	Income, pay, and work conditions, satisfaction, application of job training	Income, pay, and work conditions, satisfaction, supervision, application of job training	Factors in career selection, income, pay, work conditions, satisfaction, supervision, application of schooling, training
Noncognitive traits	Self-concept, locus of control	Self-concept, locus of control	Self-concept, locus of control, maturity scale	Self-concept, locus of control	Self-concept, locus of control
Goal orientations	Work and educational aspirations, expectations, and plans; life goals	Work and educational aspirations, expectations, and plans; life goals	Work and educational aspirations, expectations, and plans; life goals	Work and educational aspirations, expectations, and plans; life goals	Work and educational aspirations, expectations, and plans; life goals, satisfaction with progress
Marriage and family	Plans to be a full-time homemaker, number of dependents	Marital status, number of dependents, income number of children had and expect to have	Marital status, number of dependents, spouse's education and occupation, income, number of children, items and home owned	Marital status, number of dependents, spouse's education and occupation, income, number of children had and expect to have, items and home owned	Marital status, number of dependents, spouse's education and occupation, income, number of children had and expect to have, items and home owned, their value
Opinions	Ratings of high school	Usefulness of specialized training in high school	Participation in political activities, consumerism, quality of life	Political activities, sex-role orientation, sex and race biases, rating of high school, satisfaction with postsecondary education or training	Political activities, sex-role orientation, satisfaction with education or training
Military	Plans for military service	Type, training, duration, satisfaction, plans	Type, training, duration, plans	Type, training, duration, plans	Type, training, duration, plans

CASE STUDY 5

HIGH SCHOOL AND BEYOND

I. Purpose:

High School and Beyond is a longitudinal study of a nationally representative sample of 1980 high school sophomores and seniors in the United States. Its basic purpose is to replicate, eight years later,the National Longitudinal Study of the High School Class of 1972. Specifically HS&B; would provide updated information on:

- factors influencing persistence vs. dropping out of high school or college,
- the transition of young people from high school to postsecondary education or to the world of work,
- persistence in postsecondary education,

- the transition from postsecondary education to the world of work
- courses taken and grades received, both at the high-school and the college level.

II. Sponsor

Since its inception HS&B; has been sponsored by the National Center for Education Statistics (WCES) within the U.S. Department of Education.

The principal contractor who has been primarily responsible for the details of research design and for data collection, coding, and storage,has been the National Opinion Research Center (NORC).

III. Sample Design

HS&B; employs a two-stage, highly stratified sample design. In the first stage 1,122 schools that had either 10th or 12th grade students (or both) were drawn. To make the sample more useful for policy analysis, the following types of schools were oversampled: alternative public schools, public schools with high percentages of Hispanic students, Catholic schools with high percentages of minority group students, and high performing private schools. In the second stage, 36 sophomores and 36 seniors were randomly selected, school size permitting, yielding total samples of 30,030 sophomores and 28,240 seniors.

In the first follow-up survey, conducted in spring 1982, all sophomore cohort members who were still in the same schools were included with certainty, as were all dropouts and other subgroups of policy interest, yielding a sophomore cohort sample size of 29,737. Of these, a subsample of 18,000 was selected for a detailed study of high school transcripts.

In the first follow-up survey a subsample of 11,995 of the 1980 senior sample were selected.

The second follow-up survey took place in spring, 1984. At that time, samples of 15,000 members of the sophomore cohort, and 11,995 members of the senior cohort were selected for further data collection.

IV. Survey Design and Content

In the base-year survey, questionnaires and cognitive tests were administered to groups of students in each participating school. The administrator in each school filled out a questionnaire about the school; teachers in each school were asked to comment on students in the sample; and a sample of parents of sophomores and seniors (about 3,600 for each cohort) was surveyed primarily for information about their plans for financing their child's postsecondary education.

The first follow-up survey of the sophomore cohort took place in spring 1982 when most respondents were seniors. Questionnaires and tests were group administered to all base-year sample members still attending the same school. Dropouts, and transferees were contacted by mail or as a last resort, by personal interview.

For the second follow-up of the sophomore cohort and for all follow-ups of the senior cohort, contact was by mail or, when necessary, by personal interview.

The student questionnaires cover a large number of content areas, including: school work, gainful employment, demographic characteristics, physical condition, parental characteristics, social relations, and life plans. Marital and fertility history are also covered in the follow-up questionnaires.

V. Response Rates

A total of all (72 percent) of the 1,122 eligible schools selected for the base-year survey actually participated. Of the 311 schools that were unable or unwilling to participate, 204 were replaced with schools which matched them with regard to geographical area, enrollment size, community type, and other characteristics. This brought the total number of participating schools to 1,015, or 90 percent of the 1,122 target.

The student-level base-year response rate within participating schools was 85 percent. The first follow-up survey response rate was about 94 percent for each cohort.

Response rates for the second follow-up survey were 92 percent and 91 percent for the sophomore and senior cohorts, respectively.

VI. Evaluations

To maximize the validity and reliability of the data, several steps were taken:

(1) all data collection instruments were pretested on a group of respondents similar to those who would participate in the main survey.

(2) Ambiguous or inconsistent responses to mail questionnaire items were clarified by means of telephone calls.

(3) A special analysis was performed by NCES to compare the estimates of family income given by the students with those given by the parents.

VII. Data Products and Analysis

NCES makes all HS&B; data files available to the public at cost. As each new data file becomes available, an Announcement to that effect is widely disseminated to potential users.

As of summer 1984, over 150 different research studies based on HS&B; data had been published. The principal contractor of HS&B;, NORC, is developing a computerized bibliography of all HS&B-based; publications.

CASE STUDY 6

NATIONAL LONGITUDINAL SURVEYS OF LABOR MARKET EXPERIENCE

I. purpose

The National Longitudinal Surveys of Labor Market Experience (NLS) were designed to identify factors that influence the labor market behavior and experience of a group of workers (Parnes:12). Five cohorts were selected to represent workers with labor market problems of special concern to national policy makers.

The NLS was the first national survey of employment-related phenomena to focus on individual labor market behavior through time. Since 1940, cross-sectional data on labor force participation had been available from the Current Population Survey.

Since the 1950's, information on earnings and employer characteristics had been available from the Continuous Work History Sample, based on a sample of the Social Security Administration's records. Longitudinal data on associated topics is available from the Panel Survey on Income Dynamics, the Longitudinal Retirement History Study, and the Continuous Longitudinal Manpower Survey of CETA participants. None of the other surveys, however, has provided data like those from the NLS on individual gross flows linked to attitudes and experience.

II. Sponsors

In 1965 the U.S. Department of Labor's Manpower, Development and Training Administration (now the Employment and Training Administration) undertook a series of longitudinal studies of the labor force. The Department of Labor (DOL) set up a contract with the Ohio State University Center for Human Resource Research (OSU) under which OSU was responsible for planning and analyzing the surveys. The DOL set up a separate contract with the U.S. Bureau of the Census for data collection for the original cohorts. Data collection for the new youth cohorts was subcontracted to the National Opinion Research Center (NORC).

III. Sample Design

Respondents in the original four cohorts were selected from an area probability sample of the non-institutionalized civilian U.S. population. Primary Sampling Units were selected on the basis of the 1960 Census. For each cohort reliable statistics for Whites and Blacks were ensured by selecting about 1,500 Black respondents and 3,500 White respondents in each cohort. This was accomplished by classifying enumeration districts by race, and using a sampling rate between 3 and 4 times higher in predominately Black ED's.

Forty-two thousand housing units were contacted for screening interviews in early 1966. From these, interviewers identified just over 22,000 eligible respondents in 13.500 households. (A number of households contained more than one respondent, sometimes belonging to more than one cohort.)

The new youth cohort selected in 1979 is arranged in 8 strata, by race, ethnicity, income, age and sex. For these cohorts, the Census Bureau drew a sample from an area probability sample of the U.S. stratified so as to produce segments of varying size but equal with respect to the characteristics of the target sample (OSU,1979:11). Seventy five thousand addresses were selected for screening interviews and from these the WORC identified a final sample of about 12,000 respondents between 14 and 21 years of age.

The new young men's cohort includes respondents who are serving in (or returned from) the armed forces. The Department of Defense provided lists of persons on active military duty to NORC for sample selection. In the first stage a sample of military units was drawn, then within these units separate samples of males and females were selected, including some respondents not living on military bases.

IV. Survey Design and Content

A. Design

1. Respondent Rules

Proxy responses are only accepted from relatives or other members of a sample person's household, if the sample person is temporarily incapable of answering questions. Specific questions eliciting opinions or attitudes are excluded from proxy interviews.

2. Reporting Units

Separate questionnaires are completed for each respondent in a multiple respondent household. Separate household record cards are also prepared, but data from one may be transcribed to another by the interviewer.

Household composition is recorded at certain interviews. CPS definitions are used for "household members." Household characteristics are tabulated as respondent attributes at each wave. OSU has prepared special tabulations of multiple respondent households, such as a fathers-and-sons tape, a siblings tape, etc.

3. Following Movers

Local government agencies, the Postal Service, neighbors and relatives, and others recorded at the first interview as knowledgeable about the respondent's whereabouts, are among the contacts that may be questioned to obtain the current address of a sample person who has moved. Respondents who have moved are contacted through the field office closest to their new location.

4. Weighting

The basic weight for each sample case is a reciprocal of selection probability, and reflects the differential sampling ratio by race. The samples have been weighted so that the characteristics for each wave match the known distribution of the characteristics in the population.

5. Interview Schedule

The original NLS plan called for annual interviews of each cohort for five years. To reduce costs, after 1968 the cohorts of adult men and adult women were interviewed only every other year.

In 1972 all four cohorts were extended by including two annual telephone surveys and a personal interview on the tenth anniversary (1976-77). The entire survey was extended an additional 5 years in 1977, on the recommendation of a group of analysts and data users convened by the department of labor. After 1983 the older and younger men's cohorts were dropped, and the older and younger women's cohorts were extended 5 years (along with the new youth cohorts).

6. Interview Mode

For the original four cohorts the first and final waves consisted of personal interviews. Four of the intervening waves were conducted by telephone (5 for mature women), and one mail questionnaire was sent in 1968. The interview, schedule for the new youth cohort called for persona interviews in each year from

1979 to 1964.

B. Content

The NLS was originally composed of 4 separate longitudinal cohorts: Adult men, adult women, young men and young women. The cohorts represent four groups important to policy makers: men in the years leading to retirement (between 45 and 59 years old in 1966); women likely to be re-entering the labor market (between 30 and 44 years old in 1967); and young men and young women likely to be finishing their education and entering the labor market (boys between 14 and 24 years old in 1966 and girls between 14 and 24 years old in 1968).

The longitudinal survey of adult men was planned to answer specific research questions about retirement decisions, about skill obsolescence, about the duration of unemployment in this age group, and about the relationship between health and labor market experience.

The sample of adult women was designed to study women's entry or re-entry into the labor force after a period spent primarily in raising children. Special attention was paid to attitudes toward employment in general and towards the propriety of labor market activity for women in particular.

The cohorts of young men and young women were planned to provide information on the extent of occupational knowledge among teenagers, and on attitudes toward education and toward employment experiences. The new youth cohort was developed in 1979 to study employment patterns in low income and minority groups, and to look at changes since 1960.

Many of the interviewing procedures and labor force concepts used in the NLS were similar to those used in the Current Population Survey (CPS) and the Census Bureau's CPS interviewers were often assigned to do NLS

interviewing as well. Coding of occupation and industry continue to conform to the definitions used in the 1960 Census. Although for most recent 1980 codes are used as well.

Older Men's Cohort:

In each wave data were collected to measure employment and unemployment. For all jobs held since leaving school, the interviews collected occupation, industry, location and duration of employment. In addition, annual income and earnings were collected for each job, along with measures of job satisfaction.

Mature Women's Cohort:

The surveys of adult females contained similar questions about background and labor force participation. But in place of questions about retirement, there were questions designed to study the process of leaving and re-entering the labor force.

Background questions for women were designed to distinguish labor market participation before and after any interregnum that began with marriage. A large number of questions dealt with household structure and responsibilities for dependents,, including attitudes toward child care, costs and preferences for child care, the husband's health limitations, and husband's attitudes toward women working.

Young Men's and Young Women's Cohorts:

The questionnaires for the original youth cohorts Were similar in most ways the adult questionnaires. Among the unique variables were an inventory of current job characteristics which included variety and autonomy of tasks, feedback from supervisors, and opportunities for contact and friendships on the job. Union membership was measured in several waves, and a large number of questions measured educational performance and experiences. These included curriculum preferences in high school and college, college finances, and reasons for leaving school.

For young men, only, retrospective data on military service were collected, including military job series. For the young women's cohort, questions were asked relating to household dependents and child care responsibilities. These were identical to questions asked in the survey of adult females,, including the repeated measures of attitudes toward women working.

Intermittent Questions:

For the adult males, questions were asked in some waves pertaining to physical health, retirement plans, and attitudes toward women working. In other waves questions were asked about commuting times and costs, collective bargaining coverage, training after leaving school, spouse's health limitations, and military service. In two-waves there were questions calling for retrospective evaluations of career experiences, including perceptions of age, sex and race discrimination, perceptions of individual career progress, and perceptions about job pressures.

For the adult women's cohort, there were questions in some waves about volunteer activities, and questions on attitudes toward women working were repeated at intervals.

A number of attitude measures were collected intermittently for the young men's cohort. In the first interview a score for occupational knowledge was compiled, and in the final interview a standard index of job satisfaction was derived for young men. Questions were asked at intervals to evaluate job aspirations and expectations about education and training.

Data from Administrative Records:

For the adult cohorts, the size of the local area labor force, and the annual local unemployment rate were recorded in each file at each wave. In addition, for the adult female cohort, an index of local demand for female labor was also included.

For the youth cohorts, a standard IQ test was administered once to each respondent. The presence of an accredited college in the local area was recorded in each file during the first interview. An index of local demand for female labor was included in six waves for young women. For all the youths, background data were collected on the quality and curriculum of the schools that the respondents were attending at the tine of their selection for the sample.

V. Response

The possibility of sample attrition worried the designers of the NLS, but it does not appear that any major attrition biases have detracted from the reliability of generalizations about the populations which the MLS cohorts represent (OSU, 1982).

Over all, after 12 years of the survey, an average 80 percent of the eligible respondents were still being interviewed (U.S.:321). When a 5 year extension was considered for the original 4 cohorts, the Census studied the known characteristics of non-respondents, and concluded that after 15 years those still being interviewed were not significantly different from those who had dropped out of the survey, judging by most socio-demographic characteristics (OSU, 1982).

The attrition rates have differed by cohort. Three years after the first interview for adult males, almost 5 percent of these respondents were no longer eligible (through death or institutionalization) and about 92 percent of the remainder were interviewed.

The worst attrition has been in the original young men's cohort, perhaps due to the exclusion of those serving in the armed forces (Parnes:25). Of those interviewed in 1966, 1.4 percent were dead or institutionalized in 1968, and an additional 12.4 percent were out of scope because they were in the armed forces. Just under 89 percent of the remainder were interviewed.

The figures for women and girls were slightly better. One percent of the women were ineligible after 2 years, and almost 94 percent of those eligible were interviewed. For girls, over 93 percent of the eligible respondents were interviewed in 1970, 2 years after selection.

To monitor sample attrition in the four original cohorts, every 5 years the distribution of such characteristics as occupation, educational attainment, age and marital status was compared to national estimates. To compensate for attrition, interviews and non-interviews are stratified by race, education, and residential mobility, and the weight of interviews in each cell is adjusted for the proportion of non-interview cases in each wave. A final adjustment is made for the re-entry of young men serving in the armed services during the year the sample was selected (1965-66).

There are no allocations or amputations for missing data to prevent inconsistencies with data from other waves. Only when missing data are clearly due to a record-keeping error are data from one item used to replace those from another.

In 1982, the characteristics of respondents still in the sample were compared to the characteristics of the sample interviewed in the initial year. Age, race, educational attainment, employment status, industry, occupation, marital status, SMSA, and annual income were all compared. For most cohorts, the differences in distribution of characteristics between the 2 samples were less than 2 percent. It was concluded that attrition had not seriously distorted the representativeness of the cohorts, and that any potential bias could be dealt with through weighting (Rhoton:7).

VI. Evaluations

To reduce attrition in the new youth cohort, several procedures were modified, based on experience with the 4 original cohorts. First, some questionnaire items that had caused response problems were changed. Second, more information was collected at the first contact that could be used in tracing mobile respondents. Third, more information about the NLS was provided to respondents, both before and after the interviews, and a newsletter is mailed to respondents to report on survey results. Finally, the NORC traced and contacted persons who were non-respondents in earlier waves. (Previously nonrespondents were dropped from the sample after 2 years of noninterviews.) This tracing was successful in over one-third of attempted cases (Rhotor :2-12).

VII. Data Products and Analysis

The Ohio State University makes WLS data files and documentation available to other researchers at cost. By 1979, data files were available for adult males 1966-76, for adult females 1967-76, for young men 1966-75. and for young women 1968-75. The data at any release point are composed of the entire longitudinal record, and include revisions to remove errors found in previous releases.

CASE STUDY 7

RETIREMENT HISTORY STUDY

I. Purpose

The Social Security Administration's Retirement History Study (RHS) is a multiwave panel survey designed to address a number of policy questions relating to the causes and consequences of retirement. Among these questions are: Why do individuals retire before age 65? How well does income in retirement replace proretirement earnings? What happens to the standard of living after retirement How do Social Security and other laws affect retirement patterns?

Until the RHS was undertaken, data bearing on these Issues were based on retrospective questions from cross sectional surveys. A prospective longitudinal study permits accurate analyses of the factors influencing the retirement decision and an accurate description of the complex of personal adjustments required during preretirement and postretirement years.

II. Sponsors

The RHS was sponsored by the Social Security Administration under direction of staff in the Division of Retirement and Survivor Studies, Office of Research and Statistics. Early consultation was provided by an outside advisory committee.

Data Collection was performed by the Bureau of the Census.

III. Sample Design

The original sample of 12,549 persons was a multi-stage area probability sample selected from members of households in 19 retired Current Population Survey rotation groups. The sample was nationally representative of persons age 58 through 63 in 1969. The sample included men of all marital status categories and women with no husband in the household. Married women were excluded because they were found in early pretests to have no independent retirement plans. Institutionalized persons were also excluded from the original sample.

IV. Survey Design and Content

A. Design

1. Respondent Rule

Proxy responses were accepted only for that part of the questionnaire dealing with spouse's labor force history. Sample persons who were not interviewed in the first wave (1969) were dropped from the survey. Respondents who were institutionalized 90 days or more at the time of subsequent waves were kept in the sample. All other noninterviews in later waves were dropped from the sample.

2. Reporting Units

The reporting units were designated sample members (individuals) only.

3. Following Movers

A year before each interview (after the first) the SSA provided the Census with current address listings for all sample persons and/or Spouses who were benefit recipients. In addition the Census checked all previous addresses with the post office to identify movers. Both these procedures reduced the number of unanticipated movers (especially between data collection regions) encountered at the time of interviewing. All movers were followed except those who emigrated or who lived more than 50 miles from any PSU.

4. Weighting

The weighting procedure began with a basic weight based on factors relating to the original CPS rotation groups and was followed by several stages of ratio estimation. Weighting for noninterviews was adjusted after 1969. No further weighting adjustments were made because by 1979 SSA ha,d determined that the differences between weighted and unweighted estimates were too small to justify the procedures.

5. Interview Schedule

Initial interviews were conducted in 1969 and then in alternate years through 1979. In each wave the interviews were conducted over a 3 to 4 month schedule (usually February to-June).

6. Interview Mode

The interview mode was personal and face-to-face. At each wave contact began with a letter from the Census informing the sample of the upcoming interview. Interviewers were encouraged to use telephone contacts to schedule their visits, but all interviews were by personal visit. Questionnaires with missing information could be completed by telephone.

B. Content

The interview schedule was designed to elicit a wide range of Information about preretirement lives and attitudes of sample members. The schedule was divided into six sections:

(1) respondent's labor force history; (2) preretirement and retirement plans; (3) health; (4) household, family and social activities; (5) income, assets and debts for respondent, spouse and children under age 18; and (6) spouse's labor force history. Base-line labor force history was collected only in the first

interview (1969). This explains why all noninterviews In the 1969,wave were dropped from the sample. By collecting labor force history for the sample person's spouse, longitudinal data was available If a surviving spouse later replaced a deceased sample person as respondent. Survey data were also supplemented with individual Social Security earnings and benefit records, yielding information on the continuity of work history and the amount or benefits to which the workers were entitled.

V. Response

Of the original sample Of Just over 12,500 selected in 1969, 8,700 were interviewed in 1977. This included over 1,000 surviving spouses who were eligible to serve as respondents after the death of a sample person. At each wave nonresponse (composed of refusals, no contact, and persons institutionalized) seldom rose over 4 percent. The remaining attrition was caused by deaths among the sample. The low nonresponse rate was in part attributable to efforts made to contact respondents: no limits Were placed on the number of attempts interviewers should make. Some refusals were related to the length of the interviews. The first averaged an hour and 15 minutes long. In subsequent years the length of the Interview was the Most frequently cited reason for refusal to respond.

VI. Evaluation

(Unknown)

VII. Data Products and Analysis

Most of the published analyses have been organized into a series of reports that are available from the Social Security Administration.

CASE STUDY 8

WORK INCENTIVE EXPERIMENTS

I. Purpose

Section 505(a) of the "Social Security Disability Amendments of 1980" (Pub. L. 96-265) directs the Secretary of Health and Human Services (HHS) to develop and carry out experiments and demonstration projects designed to encourage disability insurance beneficiaries to return to work and leave the benefit rolls. The objectives of these experiments, specified in the law, are to generate long-range savings to the trust funds and to facilitate the administration of title II of the Social Security Act. Section 505(a) itself contains several suggestions for experimental variables, specifically

- Benefit reductions based on amount of postentitlement earnings.

- Lengthening the trial work period.

- Altering the 24 month waiting period for Medicare benefits.

  - Changing the manner in which the program is administered.

The language in section 505(a) states explicitly that the experiments should be carried out in a way that permits thorough and complete evaluation and on a large enough scale so that the results may be generalized reliably to the future day-to-day operation of the disability program. In addition, the report of the House Ways and Means Committee indicates Congress' desire that no individual be disadvantaged compared to existing law.

II. Sponsors

This project, mandated by law (Pub. L. 96-265), directs the Secretary of HHS to carry out the experiments. Planning the experiments has been delegated to SSA. The law authorizes the use of disability insurance trust fund monies to pay for the experiments and authorizes the Secretary to waive the present benefit and eligibility requirements of titles II, XVI and XVIII to the extent necessary to carry out the experiments.

---

*Since its mandate in the Disability Insurance Amendments of 1980 (Pub. L. 96-265), the Social Security Administration disability program work incentive experiments have undergone a number of designs. Due to a number of administrative problems and the imminent deadline of the legislative mandate no experimental plan has yet to be implemented. Legislative extension of the experimental authority is now under consideration. For expository purposes the plans developed in the Fall of 1982 are presented.

III. Sample/Experimental Design

A. The Study Population

The study population for the WIE consists of all newly awarded beneficiaries except those who fall in one of the following categories:

- Under age 18 or over age 59 at time of award.
- Residing outside the 48 contiguous States or in an institution.
- Received a closed period award.

- Previously entitled to DIB.

- Dually entitled to DI and to title II auxiliary benefits.

- Statutorily blind.

- Career railroad case certified to the ERB for payment.

B. Experimental Design

1. Programmatic Changes

Sample sizes for each experimental group and the control group have been determined in an attempt to insure the ability to measure important increases in the proportion of work recoveries. Our best estimate is that under current law about three percent of a newly awarded beneficiary cohort will have their benefits terminated after successful completion of a trial work period. We estimate that for the proposed experimental alternatives, a one percentage point increase in the recovery level (that is, a change from three to four percent) would yield significant trust fund savings, on the order of $100 million per year or larger. Thus, the sample sizes we choose insure a good chance of detecting a one percentage point change if this change occurs in any experimental group. The required sample size total 21,000 cases, including 3,000 for each of the five experimental groups and 6,000 for the control group.

Schematically, the design of the WIE and the sample sizes and allocations can be pictured as follows:

116

Medicare extension

Medicare extension

Benefit/Earnings offset		NO	YES
	NO	T1 - Control Group (present law) 6,000 beneficiaries	T2 - Medicare extension only 3,000 beneficiaries
		T3 - Extension of trial work period and shortening of Medicare waiting period 3,000 beneficiaries	
	YES	T4 - 50% earnings offset (with additional year of eligibility) 3,000 beneficiaries	T5 - 50% earnings offset (with additional year of eligibility and Medicare extension) 3,000 beneficiaries
		T6 - 100% earnings offset (additional year of benefit eligibility) 3,000 beneficiaries	

Group T.1 represents a control group operating under the provision of the current law. For each of the experimental groups, T.2-T.6 inclusive, only the programmatic change(s) specified applies.

2. Administrative Changes

Two administrative changes will be instituted to assure that the WIE operates effectively and efficiently. These changes are (1) a face-to-face interview at the start of the experiments that explains the experimental changes to the participating beneficiaries, and (2) use of a quarterly report of work and earnings. With these up-to-date reports it is Possible to minimize the problems benefit overpayments. These changes in themselves may alter beneficiary behavior. The experiments is therefore designed to test whether these administrative changes have a direct effect an recovery.

The following experimental group make up this portion of the WIE experimental design:

Quarterly Work Report		Face-to-Face Contact: Yes	Face-to-Face Contact: No
	Yes	T_1 Control Group for WIE 6,000	$T_{1.1}$ Quarterly Report Only 3,000
	No	$T_{1.2}$ Face-to-Face Contact Only 3,000	$T_{1.3}$ Control 3,000

This scheme takes advantage of the 6,000 cases that will already serve as the control group for the WIE. As a result, only an additional 9,000 cases would be required to study the impact of the two administrative changes being tested.

The considerations used in determining sample size and the allocation of cases among the four test groups involved in this portion of the experiment are essentially the same as those discussed for the programmatic revisions. It should be pointed out that none of these cases (the 6,000, as well as, the additional 9,000) will involve either increased benefit payments or Medicare reimbursements. They all operate under present program provisions.

C. Sample Design

1. Stratification

In order to improve the efficiency of the experimental design the award population will be Stratified by two factors -- age and medical diary status. Since younger beneficiaries are more likely to return to work and leave the benefit rolls, they are likely to take advantage of the experimental provisions than older beneficiaries. Beneficiaries who are scheduled for medical reexaminations might be less likely to be granted trial work periods because they are judged more likely to recover.

118

The following table defines four age/diary strata.

Medical

Stratum	Age	diary
S1	18-44 (young)	Yes
S2	18-44	No
S3	45-59 (old)	Yes
S4	45-59	No

Taking these strata into account, the full experimental design has the following dimensions:

Experimental group	Total	Stratum S1	S2	S3	S4

Total	36,000	3,600	7,200	3,600	21,600
T1	6,000	600	1,800	600	3,000
T2	3,000				
T3	3,000				
T4	3,000				
T5	3,000	300	900	300	1,500
T6	3,000				
T7	3,000				

T8	3,000				
T9	3,000				
T10	6,000	600	1,800	600	3,000

The allocation to stratum will be roughly proportionate to size.

Note that an additional experimental group, T.10, is shown. This group represents a "silent" control group. The beneficiaries in this group do not receive any program or administrative changes, as is the case for group T.9. The beneficiaries in T.10, however, will not be processed by the WIE review unit. This allows us to test the experimental effect of establishing the special unit itself through comparison of T.9 and T.10 outcomes. Thus, the total number of beneficiaries with any involvement in the WIE is now 36,000.

2. Geographic Clustering and Stratification by Dat a of Award

In order to limit the impact of the face-to-face treatment application on SSA field staff and Costs, SSA's Office of Field Operations has asked that WIE sample cases be in no more than 200 SSA districts. (A district is defined to be an SSA district office and its associated branch offices.) We, therefore, group the WIE population into clusters of SSA districts. The selection of a sample of clusters is the first stage of selection for the WIE sample.

The size of these clusters depends on a number of interrelated requirements. The first requirement is our desire to put a full replicate of the experimental design (or multiples thereof) into each cluster of districts as indicated in the following table:

Experimental group	Total	Stratum S1	S2	S3	S4
Total	120	12	24	12	72
T1	20	2	6	2	10
T2	10				

T3	10				
T4	10				
T5	10	1	3	1	5
T6	10				
T7	10				
T8	10				
T9	10				
T10	20	2	6	2	10

One hundred and twenty cases is the minimum number required to simultaneously satisfy the allocations discussed above among the strata and among the experimental groups .

Placing a full replicate In each cluster induces orthogonality between treatment (and strata) and cluster and facilities the analysis of experimental results. In particular, under the assumption of no interaction between treatment and cluster in producing experimental outcomes, the association between treatment and outcome can be measured by tabulating treatment (and strata, if necessary) results alone essentially ignoring geographic effects. The ability to display the results of the experiments in an uncomplicated manner is of great importance in presentations to those persons responsible for program and operating policy.

The second aspect of the determination of minimum cluster size is that each cluster should have a high probability of providing the necessary number of sample cases in each stratum to complete the design; that is, 12 cases for S1, 24 for S2,, 12 for S3 and 72 for S4. It turns out that a population of 250 will yield the needed cases with a probability greater than .998.

The third aspect to be, considered is that the number of districts in the sample must not exceed 200. This constraint has implications for the length of the sampling period. There are about 614 districts contained In the 48 contiguous States with an average of about 350 new awards per district per year. Since the sample each cluster will require 250 awards to achieve a 120 case replicate, about 75,000 awards will have to be available to obtain the full 36,000 case sample. Since 200 districts can supply about 70,000 cases a year, the 200 district constraint implies the need for a 1 year sampling period.

The 1 year sampling frame will be divided into 6 bimonthly sampling periods, with a full 120 case replicate of the design going into each cluster of districts in each sampling period. Each cluster will need to supply 1,350 awards in each year. Since each cluster supplies 720 (120 times 6) sample cases, 50 clusters are required for the sample to complete the design in 1 year (50 x 720 = 36,000).

IV. Survey Design and Content

A. Design

1. Respondent rule.

No proxy responses are accepted.

2. Reporting units.

Individual beneficiary and spouse.

3. Following movers.

All movers will be followed.

4. Weighting.

The basic weight for each sample case will be the reciprocal of estimation is the probability of selection. No need for ratio anticipated.

B. Interview Schedule/Mode and Content

In addition to data from administrative records, a baseline questionnaire and followup mail questionnaire will be administered.

At the start of the experiment, field personnel will contact all persons (except those in T1.1 and T1.3 and the silent control group) to explain to them in person. At that time the interviewer will administer a short questionnaire designed to obtain data on demographic characteristics, family composition, amount and source of family income and private disability insurance benefits. The questionnaires will be mailed to members of groups that are not contacted for face-to-face interviews.

A supplemental mail questionnaire will be sent every 6 months over 4 years to a subsample of 10,000 beneficiaries. The questionnaire will be designed to elicit information that will update the baseline interview and describe how beneficiaries find jobs and the factors involved in the success or failure Of sustained work.

V. Response

Since all participants will be tracked through administrative records, there will be no actual attrition from the study.

Response to the supplemental questionnaires is expected to be high because they will be administered in conjunction with required administrative reports.

VI. Evaluation

None planned.

VII. Analysis Plans (See text discussion.)

NATIONAL MEDICAL CARE EXPENDITURE SURVEY

I. Purpose

The National medical care Expenditure Survey (NMCES) was designed to assess the use of health care services and to determine the patterns and character of health expenditures and health insurance for the U.S. noninstitutionalized civilian population in 1977. The survey was conducted by the National Center for Health Services Research (NCHSR), as part of a landmark study, the National Health Care Expenditures Study (NHCES), which is providing information on a number of critical issues of national Health policy. Topics of particular interest to government agencies, legislative bodies, health professionals, and others concerned with health care policies and expenditures include:

- The cost, utilization, and budgetary implications of changes in federal financing programs for health care and of alternatives to the present structure of private health insurance.
- The breadth and depth of health insurance coverage.
- The proportion of health care costs paid by various insurance mechanisms.
- The influence of Medicare and Medicaid programs on the use and costs of medical care.
- How and why Medicaid participation changes over time.
- Patterns of use and expenditures as well as sources of payment for major components of care.
- The cost and effectiveness of different federal, state, and local programs aimed at improving access to care.
- The loss of revenue resulting from current tax treatment of medical and health insurance expenses, particularly with regard to the benefits currently accruing to

different categories of individuals and employers, and the potential effects on the federal budget of proposed changes to tax laws.

- How costs of care vary according to diagnostic categories and treatment settings.

The data for these studies were obtained from the National Medical Care Expenditure Survey (NMCES), which has provided the most comprehensive statistical picture to date of how health services are used and paid for in the United States. The survey was completed in September, 1979.

Data were obtained in three separate, complementary stages. About 14,000 randomly selected households in the civilian, noninstitutionalized population were interviewed six times over an 18-month period during 1977 and 1978. This survey was complemented by additional surveys of physicians and health care facilities providing care to household members during 1977 and of employers and insurance companies responsible for their insurance coverage.

II. Sponsors

Funding for NMCES was provided by National Center for Health Services Research, which co-sponsored the survey with the National Center for Health Statistics. Data collection for the survey was done by Research Triangle Institute, NC, and its subcontractors, National Opinion Research Center of the University of Chicago, and Abt Associates, Inc., of Cambridge, MA. Data processing support is being provided by Social and Scientific Systems, Inc. of Washington, D.C.

III. Sample Design

The survey sample was designed to produce statistically unbiased national estimates that are representative of the civilian noninstitutionalized population of the United States. To this end, the study used the national multi-stage area samples of the Research Triangle Institute and the National Opinion Research Center. Sampling specifications required the selection of about 14,000 households. Data were obtained for about 91 percent of eligible households in the first interview and 82 percent by the fifth interview.

The NMCES area sampling design can be characterized as a stratified three-stage area probability design from two independently drawn national area samples. The fourth stage involved the selection of ultimate sampling units (e.g., housing units and a special class of group quarters). An essential ingredient of this design is that each sample element has a known, nonzero selection probability. Also, the national general purpose area samples from the Research Triangle Institute (RTI) and the National Opinion Research Center (NORC) used in the survey are similar in structure and, therefore, compatible. Except for difficulties associated with survey nonresponse and other nonsampling errors, statistically unbiased national and domain estimates can be produced from each sample or from the two samples combined.

The first stage-in both designs consists of primary sampling units which are counties, parts of counties, or groups of contiguous counties. The second stage consists of secondary sampling units which are census enumeration districts or block groups (Bureau of the Census, 1970). Smaller area segments generally consisting of at least 60 housing units constitute the third stage in both designs; a subsample of households was randomly selected from each of these segments in the final stage of sampling. Combined stage specific sample sizes for the two designs were 135 primary sampling units (covering 108 separate localities), 1,290 secondary sampling Units, and 1,290 segments. Here, the number of separate primary areas is less than the sum of the number of primary sampling units in the two national primary samples since units from some of the large Standard Metropolitan Statistical Areas (SMSAS) were selected in both samples. Selection procedures for the fourth stage included a disproportionate sampling scheme to obtain a target of 3,500 uninsured households.

IV. Survey Design and Content

As noted, about 14,000 households participated in six separate rounds of interviews during 1977 and early 1978. The first interviews began in mid January 1977; subsequent rounds of interviews were conducted at intervals of about three months. The first, second, and fifth rounds of interviews were

conducted in person, as were about 20 percent of the third and fourth rounds and about half of the sixth round; the remainder were conducted by telephone.

During each of the first five rounds of interviews, information was obtained on use of medical services, charges for services and sources of payment, numbers and types of disability days, and status of health insurance coverage. Data collected during the first interview covered the period from January 1, 1977, through the date of interview. Data collected during the second, third, and fourth rounds covered the period from the immediately preceding interview through the date of the current interview. The fifth interview covered the period from the previous interview through December 31, 1977.

Beginning in the second round of interviews and continuing through the fifth, the household respondent was asked to review a computer-generated summary of data previously reported on health care services received and costs. This review permitted a check for accuracy and completeness and provided the necessary information to check continuity among the interview rounds for such data as health insurance coverage and charges for multiple services.

The sixth round of interviews consisted of a series of supplemental questions covering limitations of activity, status of income tax filing, and the amount of itemized medical deductions. Supplemental questions also were asked during the second through fifth round interviews. These questions covered employment, health insurance, access to health care, barriers to care, ethnicity, and income and assets.

In addition to answering questions, each survey participant was asked to sign a permission form so that each physician or facility that had been reported as providing medical care during 1977 could release information about the patient. In cases where a person had not reported receiving medical care in 1977 from his usual source of medical care, a permission form for his usual source of medical care was requested. Persons with health insurance policies were asked to sign a permission form authorizing release of information by the employer, union group, or insurance company. When employed persons reported no health insurance coverage, they were asked to sign a permission form authorizing the employer to provide information about the insurance coverage that was available. These forms were collected at various times during the survey and provided data which was the basis for the subsequent surveys of medical providers and health insurers.

V. Response Rates

Data were obtained for approximately 91 percent of eligible households in the first interview and 82 percent by the fifth interview. Of 38,815 participants in the NMCES, 4146, or 10.7 percent failed to respond for the entire time period of 1977 for which they were eligible to respond. For example, a person could have refused participation after initially cooperating -in the first interview by not responding for the remainder of the interviews. Similarly, the inability to reestablish contact with a participant after change of residence would result in this type of nonresponse. This problem of partial nonresponse is not limited or unique to the NMCES, but characteristic of national panel surveys in general.

VI. Evaluation Component

The NMCES used several methodological innovations to insure data reliability. During each round of interviews, respondents were asked to report the diagnosis, total charge and sources of payment for each inpatient hospital stay, medical provider visit, dental visit, prescription drug, or purchase of eyeglasses or other medical equipment. In addition, respondents were asked to provide information about their health insurance coverage. Data on health care use and expenditures were updated each round through the use of a computerized summary of the information reported in the previous interview. Respondents were asked to review this information and make any needed additions or corrections. In particular, the summary was expected to allow respondents a means to provide more complete charge and payment data at a later date if it was unknown at the time of the interview. All respondents were asked to complete the summary. Approximately 32 percent of household survey respondents were also included in the medical provider survey. The medical provider survey (MPS) was a record check or verification procedure to obtain expenditure and diagnostic data from physicians and hospitals who treated a sample of household respondents during the year. Thus, for each person in the household survey the data obtained from the questionnaire was checked in a subsequent interview through the summary mechanism and in about a third of the cases, subjected to verification through the MPS. In addition, household data on health insurance coverage was verified through the Health Insurance/Employer Survey (HIES) which collected, for each private health insurance plan reported in the household survey, data from employers, insurance carriers or other insuring organizations.

VII. Data Products and Analysis

NCHSR has developed National Medical Care Expenditures Survey data files and documentation for public use. As of spring 1985, over 100 different research studies based on NMCES data had been published. A detailed Annotated Bibliography of Studies from the

National Medical Care Expenditure Survey is available from the National Center for Health Services Research.

CASE STUDY 10

NATIONAL MEDICAL CARE UTILIZATION AND EXPENDITURE SURVEY

I. Purpose

The National Medical Care Utilization and Expenditure Survey (NMCUES) was designed to collect data on health, access to and use of medical services, charges and sources of payment for medical services, and health insurance coverage for the U.S civilian noninstitutionalized population during 1980. NMCUES was developed from a series of surveys concerning health, health care, and expenses for health carp. However, NMCUES drew most heavily from two surveys -- the National Health Interview Survey (HIS) and the National Medical Care Expenditure Survey (NMCES).

The HIS is a continuing survey that began in 1957 and is conducted by the National Center for Health Statistics (NCHS). Its primary purpose is to collect information on illness, disability, and use of medical care. Although some medical expenditure and insurance information has been collected in the HIS, a cross-sectional survey design was inefficient for obtaining complete and accurate Information of this type. It was concluded that a panel survey procedure would be required, and a pilot survey was conducted for the NCHS by the Johns Hopkins University Health

Services Research and Development Center and by Westat Research, in 1975-76.

Based on information obtained during the pilot study, the National Center for Health Services Research (NCHSR) and NCHS cosponsored the National Medical Care Expenditure Survey in 1977 - 78. This was a panel survey for which households were interviewed six times to obtain data for 1977.

NMCUES was similar to the NNCES in survey design and questionnaire wording, to allow analysts of change during the 3 years between 1977 and 1980. Both NMCUES and NMCES are similar to the HIS in terms of question wording in areas common to all three surveys. However, each survey is different with special emphasis on different areas. Together they provide extensive information on illness, disability, use of medical care, costs of medical care, sources of payment for medical care, and health insurance coverage at two points in tine.

II. Sponsors

NMCUES was cosponsored by NCHS and the Health Care Financing Administration (HCFA). Data collection was provided under contract by the Research Triangle Institute (RTI) of Research Triangle Park, North Carolina, and its subcontractors, National Opinion Research Center (NORC) of Chicago, Illinois, and SysteMetrics, Inc., of Santa Barbara, California. The contract was awarded in September, 1974.

III. Sample Design NMCUES utilized two frames, the first to provide a national household sample and the second to provide a State Medicaid household sample. The process of selecting each sample was different, and is described separately.

A. The National Household Sample:

The NMCUES sample of dwelling units is derived from two independently selected national samples; one provided by RTI and the other by NORC. The sample designs used by RTI and NORC are quite similar with respect to principal design features. Both can be characterized as self-weighting, stratified, multistage area probability designs. The principal differences between the two designs are the type of stratification variables and the specific definitions of sampling units at each stage.

B. The State Medicaid Household Sample:

The November, 1979 Medicaid eligibility files in California, Michigan, New York and Texas Were used as frames to select a sample of cases for the State Medicaid household component of the survey. A case generally consisted of all members of a family receiving Medicaid within the same category of aid. The State aid categories were collapsed into three or four strata, depending on the State. These were: (1) aid to the blind and disabled; (2) aid to the elderly (those with Supplementary Security Income); (3) Aid to Families With Dependent Children (AFDC);and (4) State only aid in California, Michigan, and New York, which provided some Medicaid coverage without Federal reimbursement. Cases in other Federal aid categories were excluded from the target population because the counts were too few to permit separate stratification. Approximately equal numbers of cases were selected from each stratum, and cases were clustered by zip codes for ease of interviewing. The lack of a central automated eligibility file in New York State (outside of the five New York City boroughs and a few other counties) required selection of counties before stratification. Within many of these counties, the lack of automation also required cases to be selected without consideration of zip codes.

C. Links to Administrative Records:

In addition to the data collected during interviews with sample households, another phase of data collection occurred after the final round of household interviewing was completed. Medicaid and Medicare numbers provided by the household were used to extract data from the Medicaid files of the Federal government. Data from the administrative records were merged with the household data to increase the analysis capabilities of the data.

IV. Survey Design and Content

A. Design

1. Respondent Rules --

The respondent for the interview was required to be a household member, 17 years of age or older. A non-household proxy respondent was permitted only if all eligible household members were unable to respond because of health, language, Or mental condition.

2. Following Movers --

The rules for following movers were slightly different for the national household samples and the State Med4caid sample. First, for the national household survey all persons living in the housing units or group quarters at the tine of the first interview contact became part of the sample. Unmarried students 17 - 22 years of age who lived away from home were included in the sample if the - parent or guardian was included in the sample. In addition, persons who died or were institutionalized between January 1st and the date of first interview were included in the sample if they were related to persons living in the sampled housing units or group quarters. All of these persons were considered "key" persons, and data were collected for them for the full 12 months or 1980 or for the proportion of time they were part of the U.S. civilian noninstitutionalized population. In addition, babies born to key persons were also considered key persons, and data were collected for them from the time of birth.

Relatives from outside the original population (i.e., institutionalized in the Armed Forces, or outside the United States between January 1 and the first interview) who moved in with key persons after the first interview also were considered key persons, and data were collected for then from the time that they joined the key person. Relatives who moved in with key persons but were part of the civilian noninstitutionalized population on January 1, 1980, were classified as "non-key" persons. Data were collected for,non-key persons for the time that they lived with a key person. Because non-key persons had a chance of selection in the initial sample, their data will not be used for general analysis. However, data for non-key persons are used for family analysis because they do contribute to the family's utilization of and expenditures for health care during the time that they are a part of the family.

For the State Medicaid sample, interviewers obtained information for each eligible member of each case. Case members who d4ed before January 1, 1980, or who were continuously institutionalized between January 1, 1980 and the first interviewer contact, were excluded from the survey. Any related person living with a case member when the interviewer contacted the Household also was designated a key person, and was tracked for the complete year.

In addition, babies born to key persons were considered key persons, and data were collected for them from the time of birth. Relatives outside the U.S. noninstitutionalized population between January 1 and the date of the first interview who moved in with a key person after the first interview also were considered key persons. Data were collected for them for the remainder of 1980. Persons who were part of the U.S. noninstitutionalized population on January 1, 1980 and who moved in with a key person after the first interview, were classified as non-key persons; data were collected only for the time that non-key persons lived with a key person. These non-key persons are included only in family analysis.

3. Weighting --

For the analysis of NMCUES data, sample weights are required to compensate for unequal probabilities of selection, to adjust for the potentially biasing effects of failure to obtain data from some persons or households (i.p., nonresponse), and failure to cover some portions of the population because the sampling frame did not include them (i.e., undercoverage).

Basic Sample Design Wieghts' -- Development of weights reflecting the sample design of NMCUES was the first step in the development of weights for each person in the survey. The basic sample weight for a dwellIng unit is the product of four weight components which correspond to the four stages of sample selection. Each of the four weight components is the inverse of the probability of selection at that stage (when sampling was without replacement), or the inverse of the expected number of selections (when sampling was with replacement and multiple selections of the sample unit were possible).

- Two Sample Adjustment Factor -- As previously described, the NMCUES sample is comprised of two independently selected samples. Each Sample, together with its basic sample design weights, yields independent unbiased estimates of population parameters. As the two NMCES samples were of approximately equal size, a simple average of the two independent estimators was used for the combined sample estimator. This is equivalent to computing an adjusted basic sample design weight by dividing each basic sample design weight by two. In the subsequent discussion, only the combined sample design weights are considered.

Ratio Adjustment (Household Level) -- The basic sampling weights were adjusted decrease sampling variation and to compensate for household level nonresponse and undercoverage. In total there were 63 ratio adjustment cells which were formed by cross-classify4ng race, age, and type of household head and size of household. Estimates from the 1980 CPS were used for population controls.

- Ratio Adjustment (Person Level) -- The household level adjusted weights were further ratio adjusted at the person level. A total of 59 ratio adjustment cells (based on age, race and sex) were utilized. Population controls, which were provided by the U.S. (Census bureau, were based on projections from the 1980 Census.

4. Interview Schedule

The sample dwelling units were interviewed at approximately 3 month intervals beginning in February, 1980 and ending March, 1981. The core questionnaire was administered during each of the five interview rounds to collect data on health, health care, health care charges, sources of payment, and health insurance coverage A summary of responses was used to update information reported in previous rounds. Supplements to the core questionnaire were used during the first, third, and fifth interview rounds to collect data that did not change during the year, or that were needed only once.

b. Interview Mode

Approximately 80 percent of the third and fourth round interviews were conduct by telephone; all remaining interviews were conducted in person.

6. Survey Costs

The basic survey design and data collection contract with RTI and NORC cost approximately $18.9 million dollar.

B. Content:

1. Core and Intermittent Questions --

The repetitive core of questions for NMCUES included health insurance coverage episodes of illness, the number of bed days, restricted activity days, hospital admissions, physician and dental visits, other medical care encounters, and purchase of prescribed medicine. For each contact with the medical care system, data were obtained on the nature of the health conditions, characteristics of the provider, services provided, charges, sources, and amounts of payment. Questions asked only once included data on access to medical care services, limitation of activities, occupation, income, and other sociodemographic characteristic.

2. Cross-wave Controls

Collection of data from the households was facilitated by the use of a calendar and a summary. At the time of the first interval, the household respondent was given a calendar on which to record information about health problems and health services utilization, and to assemble physician and other provider bills between interviews. Following each household interview, information about health provider contracts and the payment of charges associated with them was used to generate a computer summary of information provided. This summary was then printed out in a simple format and mailed to the household for review of its accuracy and completeness prior to the next interview. At the subsequent interview, the interviewers reviewed this information with the household respondent to ensure accuracy and to obtain information not available during a previous interview.

V. Response

A. Survey Nonresponse

Response rates for households and persons in the NMCUES were high, with approximately 90 percent of the sample households agreeing to participate in the survey, and approximately 94 percent of the individuals in the participating house

holds supplying information. Even though the overall response rates are high, survey based estimates of means and proportions may be biased if nonrespondents tend to have different health care experiences than respondents, or of there is a substantial response rate differential across subgroups of the target population. Furthermore, annual totals will tend to be underestimated unless allowance is made for the loss of data cue to nonresponse.

Two methods commonly used to compensate for survey nonresponse are data imputation and,the adjustment of sampling weights. For NMCUES, data imputation was used to compensate for attrition and for item nonresponse, and weight adjustment was used to compensate for total nonresponse. The calculations of the weight adjustment factors were discussed previously in the section on sampling weights.

1. Attrition Imputation --

A special form of the sequential hot deck imputation method was used for attrition Amputation. First, each sample person with incomplete annual data (referred to as a "recipient") was linked to a sample person with similar demographic and socioeconomic characteristics who had complete annual data (referred to as a "donor"). Secondly, the time periods for which the recipient had missing data were divided into two categories: Imputed eligible days and imputed ineligible days. The imputed eligible days were those days for which the donor was eligible, in scope) and the imputed ineligible days were those days for which the donor was ineligible (i.e., out of scope).

The donor's medical care experiences such as medical,provider visits, dental visits, hospital stays, etc., during the imputed eligible days were imputed into the recipient's record for those days. Finally, the results of the attrition imputation were used to make the final determination of a person's respondent status. If more than two-thirds of the person's total eligible days (both reported and imputed) were imputed, then the person was considered to be a total nonrespondent and the data for the person was removed from the data file.

2. Item Nonresponse and Imputation --

Among persons who are classified as respondents, there is still the possibility that they may fail to provide information for some or many items in the questionnaire. In the NMCUES, item nonresponse was particularly a problem for expenditures for health care, income, and other sensitive topics. The extent of missing data varied by question, and imputation for all items in the data file would have been expensive. Imputations were made for missing data on key demographic, economic, and expenditure items across the five data files in the Public Use Data Tape. Table 1 (page 13) illustrates the extent of the item nonresponse problem for selected survey measures which received imputations in the four data files used in this report.

Demographic items tend to require the least amount of imputation, some at insignificant levels such as for age, sex, and education. Income items had higher levels of nonresponse, and for total personal income, which is a cumulation of all earned income and 11 sources of unearned income, nearly one-third of the persons required imputation for at least one component. The bed disability days, work loss days, and cut down days have levels of imputation that are intermediate between the

demographic and income items.

The highest levels of imputation occurred for the important charge items on to various visit, hospital stay, and medical expenses files. Total charges for medic visits, hospital stays, and prescribed medicines and other medical expense records were imputed for 25.9, 36.3, and 19.4 percent of the events, respectively. Among the source of payment data, the imputation rates for the source of payment were small, but the rates for the amount paid by the first source of payment was genera subject to high rates of imputation. Nights hospitalized on the hospital stay file was imputed at a rate comparable to the first source of payment.

The methods used to impute for missing items were diverse and tailored to the measure requiring imputation. Three types of imputation predominate: Editing or logical amputations; a sequential hot deck; and a weighted sequential hot deck.

The imputation process will be described for two items t o illustrate the nature of imputation for the NMCUES. For Hispanic Origin, two different imputation procedures were used; logical and sequential hot deck. Since Hispanic Origin was not recorded during the interview for children under 17 years of age, a logical Amputation was made by assigning the Hispanic Origin of the head of the household to the child. For the remaining cases which were not assigned a value by this procedure, the data were grouped into classes by race of the head of the house-hold, and within classes the data were sorted by household identification number, primary sampling unit, and segment. An unweighted sequential hot deck was used to impute values of Hispanic Origin for the remaining cases with missing values.

The imputations for medical visit total charge were made after extensive edit, had been done to eliminate as many inconsistencies as possible between sources of payment data and total charge. The medical visit records were then separated into three types: Emergency room, hospital outpatient department, and doctor visit Within each type, the records were classed and sorted by several measures which differed across visit types prior to a weighted hot deck imputation. For example, for doctor visits the records were classified by reason for visit, type of doctor seen, whether work was done by a physician, and age of the individual. Within the groups formed by these classing variables, the records were then sorted by type of insurance coverage and the month of visit. The weighted hot deck procedure was then used to impute for missing total charge, sources of payment, and sources of payment amounts for the classified and sorted data file.

Since amputations were made for missing items for a large number of the important items in the NMCUES, they can be expected to influence the results of the survey in several ways. In general, the weighted hot deck is expected to preserve the means of the nonmissing observations when those means are for the total sample or classes within which amputations were made. However, means for other, subgroups, particularly small subgroups, may be changed substantially by imputation.

In addition, sampling variances can be substantially underestimated when impute values are used in the estimation process. For a variable with one-quarter of its values imputed, for instance, sampling variances based on all cases will be based on one-third more values than were actually collected in the survey for the given item. That is, the variance would be too small by a factor of one-third, at least. Finally, the strength of relationships between measures which received imputations can be substantially attenuated by the imputation.

VI.Analysis and Evaluation

Since 1980 NCHS has awarded a number of contracts for the review and analysis of NMCUES data to evaluate the quality of the data and the data collection and processing methods. This includes a contract with Westat (of Rockville, Maryland) to evaluate NMCUES data collection,and data processing and a series of 3 contracts with the University of Michigan to analyze findings related to physicians charges, patient expenditures and sources of payment. Another contract, with Applied Management Sciences, examined family characteristics and expenditures for healthcare.

VII. Data Products

Data from the NMCUES are available with documentation on public use tapes from the National Technical Information Service, a division of the Department of Commerce in Springfield, Virginia. Additional information concerning the public use tapes is available from the Utilization and Expenditure Statistics branch, NCHS.

Findings from the survey were presented in official publications primarily from the government's Public Health Service and Health Care Financing Administration 1983 - 85. A number of analyses of NMCUES appeared in a Working Paper series published by the NCHS which now has over 20 titles, as well as in professional journals dealing with public administration and public health.

Table 1. Percent of Data Imputed for Selected Survey Items in Four of the NMCUES Public Use Data Fi1es

Tape Location	Survey Item	Percent Imputed
Person File	(n = 17,123)	
	Age	0.1
	Race	20.1
(1)		
	Sex	0.1
	Highest Grade Attended	0.1
	Perceived Health Status	0.8
	Functional Limitation Score	3.2

	Number of Bed Disability Days	7.9
	Number of Work Loss Days	8.9
	Number of Cut Down Days	8.2
	Wages, Salary, Business Income	9.7
	Pension Income	3.5
	Interest Income	121.6
	Total Personal Income	30.4(2)
Medical Visit		
File	(n = 86,594)	
	Total Charge	25.9
	First Source of Payment	1.8
	First source of Payment Amount	11.6

Hospital Stay

File	(n = 2,946)		
		Nights Hospitalized	3.1
		Total Charge	36.3
		First Source of Payment	2.2
		First Source of Payment Amount	17.6

Medical Expenses

File	(n = 58,544),		
		Total Charge	19.4
		First Source of Payment	2.9
		First Source of Payment Amount	10.0

(1) Race for Children under 14 imputed from race of head

(2) Cumulative across 12 types of income

CASE STUDY 11

LONGITUDINAL ESTABLISHMENT DATA FILE

Historically the economist has relied upon aggregate economic information from various sources (including the Census of Manufactures and Annual Survey of Manufactures (ASM) programs) to investigate the changing structure of the manufacturing sector of the United States economy. It has not been possible to observe the variations in behavior among establishments (plants) or to determine how changes in the behavior of individual establishments affected the enterprise (firm) or the aggregate statistical totals. The Census Bureau has developed a Longitudinal Establishment Data

(LED) file which, when coupled with recent advances in econometric computer software, makes possible a wide range of empirical analysis at the manufacturing establishment level.

The LED file was developed in cooperation with the National Science Foundation under the general direction of Nancy and Richard. Ruggles of Yale University. The LED file is a time series of economic variables collected from manufacturing establishments in the Census of Manufactures and Annual Survey of Manufactures programs. The LED file contains establishment level identifying information; basic information on the factors of production (inputs, such as levels of capital, labor, energy and materials) and the products produced (outputs); and other basic economic information used to define the operations of a manufacturing plant. The LED file resides in a random access database environment which facilitates immediate access to individual data values.

History

The ASM program was initiated in 1949 and provides detailed economic information on the functioning of manufacturing plants in intercensal years. Since the inception of the ASM program the Census Bureau has understood the potential of linking establishment records across ASM survey years to create a longitudinal micro level data file suitable to perform time series analysis. The Ruggles' were particularly interested in developing such a file for various types of macroeconomic studies.

The first real attempt at creating such a file was undertaken in the late 1950's using the 1954 Census of Manufactures as a starting point. This first attempt tried to match establishments across time using survey identification numbers as keys. While a significant portion of the establishments had retained their identification numbers for several years, many identification numbers had been changed and no audit trail was maintained. There was really no way of linking such establishments except by laborious search of the name and address records in the mailing directory. In those days, shuttle forms were used and thus the linkage of identification numbers in different years was not critical in order to measure year-to-year change in manufacturing establishments.

This first attempt at a matching of identification numbers required a labor intensive effort to ensure accurate matches. This experience led to modifications in the ASM processing that placed greater responsibility on the directory to document identification number changes and to link old and new identification numbers. It also led to the introduction of the concept of the permanent plant number that would be assigned to an establishment throughout its life in the ASM program. This permanent identification number became critical not only to the directory controls but also to new methods of editing and tabulation.

Considerable staff and computer time were expended on this first effort and a large segment of the ASM file was successfully matched for the years 1954-1962. However, since the computer record for many establishments did not include all corrections resulting from the survey review, and because many nonmatches were left unresolved, the file was not developed to the extent necessary to be usable for a wide variety of longitudinal studies.

The first effort at creating a time series file of establishment level microdata was discontinued in 1968 because of budget restrictions. However, the experience gained from the first effort added significantly to the directory, editing and tabulation techniques used in the ASM; specifically the computer edit of the Census and ASM programs were modified to incorporate more year-to-year analysis.

During the 1970's several major advances were made at the Bureau which made it possible to renew the effort to develop a longitudinal establishment file. First, the Industrial Directory was started in 1972 which solved the problems of linkage of identification numbers due to changes in ownership. Second, the establishment correction system introduced into the Census and ASM programs in 1979 assures that all corrections made by the staff during the review of the data are applied to the data records.

Prior to 1975, budgetary constraints prevented the complete correction of the computer data files, although the corrected data were included in the official published statistics.

The current effort to develop the LED file was undertaken as a joint effort by the Census Bureau and Richard and Nancy Ruggles of Yale University, with funding provided by the NSF and the Small Business Administration. The Census Bureau has created a longitudinal data file of individual manufacturing establishment data from the Census of Manufactures and ASM for the years 1972 to 1981. This process required the linkage of establishment level records based upon identification numbers. This linkage process was complicated by the numerous plant closings, plant openings, mergers and acquisitions that transpired during the decade covered by the file.

A computer match was performed to link establishment records over time, linkage problems were resolved by the data analysts so that a consistent series of economic surveys is available for each establishment in operation during the period covered. The linked data were reformatted into a data structure suitable for such a file and extraction routines were developed so that data can be removed from the file.

Contents of the File

The basic unit of collection for the Census of Manufactures and the ASM is the manufacturing establishment. Thus the establishment is the basic unit of data storage in the LED file. An establishment is defined as a single,physical location engaged in one of the categories of industrial activity in the Standard Industrial Classification (SIC) system. The SIC system is used in the classification of manufacturing establishments by type of activity in which they are engaged; it facilitates the collection, tabulation, presentation and analysis of census data relating to establishments.

The data are stored as a time sequence of survey responses for establishments rather than as a time series of annual observations for variables. The data are sorted by a permanent establishment identification number and survey year.

The data for a particular year are stored in modular sets of fixed length records; data for a module (a set of variables) have a consistent format for all years.

The variables available from the LED file are presented in Table 1, the LED Directory. As this table indicates, basic economic information on the factors of production (inputs) such as employment, payrolls, supplementary labor costs, worker hours, cost of fuels and electricity, cost of materials, capital expenditures, rental payments, inventories and on the products produced (outputs), such as value of shipments and value added, are available for all years. In recent years, a number of new items have been added, including the consumption of specific types of fuels, methods of valuation or inventories, purchases of used structures and machinery, retirements, and depreciation. The detailed information obtained in census years on materials consumed and on products shipped are not available from the ASM, thus a continuous time series is not available for those variables.

Methodological Problems

Data Comparability through Time:

The main objective of survey processing is to identify "significant errors", i.e., those that affect the quality of the aggregate data or the test for confidentiality. We cannot afford the cost of cleaning up "insignificant" data errors. Therefore, we do not always insist on complete and correct data for each establishment, even in a sample, and rely instead on our computer edit to maintain the completeness of the record, to "estimate" data for establishments that fail to report, and to identify "significant" errors (edit failures) that are referred to the analysts for review. This means that some data errors remain in the records of the individual establishments. It should be noted that data "flags" included in the longitudinal file will indicate which cells have been computer changed or analyst corrected.

Most importantly, because of cost, we have concentrated on year-to-year comparisons of establishment data. Our computer edit has been designed to work with only two periods of data; current year and previous year. Our aggregate review focuses on two years of data, current and previous, although trends are also considered. For economic research purposes, where micro data for several years are needed this type of editing and review may not be sufficient. Different problems will come into focus when establishment data are edited and reviewed over a long-period of time as compared to using only two years.

Another factor that affects data comparability over time involves the errors that are identified during the survey processing, but which are not carried back to the file because of cost considerations. As noted earlier, this situation was virtually eliminated with the introduction of an establishment correction system for the 1975 ASM. For the, 1972 Census and the 1973 and 1974 ASM, this system was not available, but efforts were taken to assure that most of the corrections were carried back to the file. Therefore for these years a tabulation of the computer file will yield results very close to the publication totals.

Data comparability over time may also be affected by two other factors. The first involves a change in the definition of an individual item. An example of this will occur for the 1982 Census of Manufactures in regard to inventories.

Prior to 1982, information on the book value of inventories was Collected. Investigations of methods used by individual companies to compile inventories indicates that the best way to obtain consistent data among different companies and even among individual establishments of the same company is to request LIFO (last-in-first-out) inventories before the application of the LIFO adjustment or reserve. Therefore, the inventories inquiry has been revised for 1982 ,to collect data on a pre-LIFO basis (i.e., gross value before any LIFO reserve or adjustment). However, since we will be requesting additional information including the amount of the LIFO reserve, we will be able to "estimate" book value for 1982.

The second factor that would affect data comparability involves modification of the computer editing procedure used for a particular item. An example occurred in the 1977 census when the addition of retirements and detailed capital expenditure items to the report form resulted in a complete change of the editing procedure used for the assets-expenditures-retirements complex. Assets data continue to be collected as in the past, but the new computer editing procedure probably resulted in a "break" in the series for a few establishments whose assets data were edited differently for 1972 through 1976 as compared to 1977 and subsequent years.

Availability of "Processed" rather than "Raw" data:

In analysis of an establishment file, some researchers feel that the actual data reported by the respondent are preferable to the data that have been edited and changed (without verification by the company). However, the data files used for the development of the time series file include a mixture of "raw" (originally reported) and computer-corrected data. The "raw" data are no longer available for all establishments.

Therefore, researchers who advocate economic research based only on "raw" microdata will find the Census/ASM LED to be of limited use. We have already noted that data "flags" included in the longitudinal file will indicate which cells have been computer changed or analyst corrected. As a result, researchers may choose to isolate only the "raw" microdata that remain unchanged as a result of Census Bureau processing procedures.

Disclosure

The last problem to be discussed, and the most complex, involves disclosure implications. Data collected by the Bureau of the Census are protected by Title 13 of the U.S. Code from disclosure to outside parties. All tabulations and analysis of longitudinal data must be analyzed to ensure that no individually identifiable confidential data are released to outside users. Bureau of the Census policy also requires that the Center for Economic Studies prevent actual estimation or close approximation of individual confidential data from released statistics. This is accomplished by applying the Census Bureau's respondent and concentration rules, which may require suppression of individual data cells. Additional suppression of nondisclosure cells may be required in cross-tabulations to avoid complementary or indirect disclosure of confidential data.

After a request for tabulation or analysis is received by the Center, a comprehensive analysis of possible disclosure of sensitive information will be performed . The user will be notified of Possible disclosure which would require

the suppression of information. Due to the complex nature of the LED file, each disclosure analysis will be handled on a case by case basis. Under no circumstances will the Bureau release names or addresses of establishments in the file. Also the Bureau will not release microdata in any format which would allow identification of individual establishments.

The results of each project must be carefully scrutinized in terms of disclosure implications before the data can be released to the researchers. The effects of ownership changes, industry changes, corrections made as a result of reviewing the establishment data, and so forth, must be taken into consideration. Furthermore, if the time-series data are subject to regression analysis or other mathematical analysis, interesting questions are raised on what information can be released. Finally, the results of each project must be compared against the results of previous studies in order to avoid complementary disclosure problems. This is quite an undertaking, and, at present, a systematic approach to handling disclosure problems has not been developed.

How will the File be Used

Users of the LED file will work through the staff of the Center for Economic Studies (CES). A major purpose of the CES is to make industrial data available to the data user community of economic policymakers and researchers to facilitate analysis and research. The result of that analysis and research will then help the Bureau to improve its economic measurement programs. The Census confidentiality policies and the U.S. Code limit direct access to individual establishment data to Census employees who have sworn to protect their confidentiality. This regulation precludes direct access to the LED data by outside researchers only sworn Census employees will have direct access to the LED file.

The CES will act as the interface between the data user community and the LED file by processing requests by outside researchers for tabulations and analyses of the LED file. The CES is creating a computer environment that will permit low-cost expeditious processing of user requests. It will be possible for an outside analyst to request cross-tabulations of aggregate statistics, estimations of econometric models, and other economic and statistical relationships based on the establishment level data. These tasks will be performed on a cost-reimbursable basis.

The types of tasks that can be performed using the LED file include:

1. Analysis of a wide range of issues from the field of industrial organization including diversification, concentration, ownership patterns and changes, and monopolistic and oligopolistic industries.

2. Analysis of productivity, technological change and efficiency and their diffusion within and across establishments, enterprises and industries.

3. A wide range of descriptive statistics such as cross-tabulation of important variables (productivity value added, wage rates) by size of establishment or enterprise, by industry or by geographic area.

4. A wide range of studies of various economic surveys by comparing detail and summary statistics across surveys.

5. Analysis of the sources and nature of productivity growth, including geographic, size and industry dimensions.

6. Analysis of geographic patterns in input markets, especially labor and energy markets.

7. Analysis of energy use in manufacturing establishments.

8. Analysis of the geographic dimensions of, for example, labor and energy markets.

The data user/research community benefits by analysis of a rich longitudinal data base for manufacturing establishments and (through integration with other economic survey results) whole enterprises. The Bureau's economic survey programs will benefit from validation and evaluation studies through time and across economic surveys. Feedback on the scope of the surveys, uses of the data, and data anomalies discovered during analysis will improve both the content and the quality of the survey data and statistical products based on theory. Also generalized data manipulation and analysis software produced for analytical uses of the file can be made available for use in the economics division for their use in production processes.

Table 1. The Longitudinal Establishment Data File Directory

X = data available   na = not available

Description	Years									
	72	73	74	75	76	77	78	79	80	81
Identification										
(a) SIC - 4-digit industry	X	X	X	X	X	X	X	X	X	X
(b) Establishment ID number	X	X	X	X	X	X	X	X	X	X
(c) Permanent plant number	X	X	X	X	X	X	X	X	X	X
(d) Sample weight	X	X	X	X	X	X	X	X	X	X
(e) Employment size code	X	X	X	X	X	X	X	X	X	X
(f) Primary product class code	X	X	X	X	X	X	X	X	X	X
(g) Percent specialization in industry	X	X	X	X	X	X	X	X	X	X
(h) Percent specialization in primary product class	X	X	X	X	X	X	X	X	X	X
1. Location										
(a) State	X	X	X	X	X	X	X	X	X	X
(b) SMSA	X	X	X	X	X	X	X	X	X	X
(c) County										
(d) Place	X	X	X	X	X	X	X	X	X	X
2. Number of Employees, Total	X	X	X	X	X	X	X	X	X	X
(a) Production workers, average	X	X	X	X	X	X	X	X	X	X
(1) March	X	X	X	X	X	X	X	X	X	X
(2) May	X	X	X	X	X	X	X	X	X	X
(3) August	X	X	X	X	X	X	X	X	X	X
(4) November	X	X	X	X	X	X	X	X	X	X
(b) All other employees	X	X	X	X	X	X	X	X	X	X
3A. Payrolls, Total	X	X	X	X	X	X	X	X	X	X
(a) Production workers	X	X	X	X	X	X	X	X	X	X
(b) Other employees	X	X	X	X	X	X	X	X	X	X
3B. Supplementary Labor Costs, Total	X	X	X	X	X	X	X	X	X	X
(a) Legally required	X	X	X	X	X	X	X	X	X	X
(b) Voluntary	X	X	X	X	X	X	X	X	X	X
3C. First quarter payroll	na	X	X	X	X	X	X	X	X	X
4. Worker-hours of Production Workers, Total	X	X	X	X	X	X	X	X	X	X
(a) January-March	X	X	X	X	X	X	X	X	X	na
(b) April-June	X	X	X	X	X	X	X	X	X	na
(c) July-September	X	X	X	X	X	X	X	X	X	na
(d) October-December	X	X	X	X	X	X	X	X	X	na

Table 1. The Longitudinal Establishment Data File Directory
Continued--

	Years									
Description	72	73	74	75	76	77	78	79	80	81
5. Cost of Materials and Services, Total	X	X	X	X	X	X	X	X	X	X
(a) Materials, parts, etc.	X	X —	X	X	X	X	X	X	X	X
(b) Resales										
(c) Fuel	X	X	X	X	X	X	X	X	X	X
(d) Electricity	X	X	X	X	X	X	X	X	X	X
(e) Contract work	X	X	X	X	X	X	X	X	X	X
6. Quantity of Electricity										
(a) Purchased	X	X	X	X	X	X	X	X	X	X
(b) Generated	X	X	X	X	X	X	X	X	X	X
(c) Sold	X	X	X	X	X	X	X	X	X	X
7. Inventories (beginning and end-of-year), Total	X	X	X	X	X	X	X	X	X	X
(a) Finished products	X	X	X	X	X	X	X	X	X	X
(b) Work-in-process	X	X	X	X	X	X	X	X	X	X
(c) Materials and supplies	X	X	X	X	X	X	X	X	X	X
8. Depreciable Assets										
(a) Gross book value (beginning-of-year), total	X	X	X	X	X	X	X	X	X	X
(1) Structures	X	X	X	X	X	X	X	X	X	X
(2) Machinery	X	X	X	X	X	X	X	X	X	X
(b) New capital expenditures, total	X	X	X	X	X	X	X	X	X	X
(1) Structures	X	X	X	X	X	X	X	X	X	X
(2) Machinery	X	X	X	X	X	X	X	X	X	X
(c) Used capital expenditures, total	X	X	X	X	X	X	X	X	X	X
(1) Structures	na	na	na	na	na	X	X	X	X	X
(2) Machinery	na	na	na	na	na	X	X	X	X	X
(d) Retirements, total	na	na	na	na	na	X	X	X	X	X
(1) Structures	na	na	na	na	na	X	X	X	X	X
(2) Machinery	na	na	na	na	na	X	X	X	X	X
(e) Gross book value (at end-of-year), total	X	X	X	X	X	X	X	X	X	X
(1) Structures	X	X	X	X	X	X	X	X	X	X
(2) Machinery	X	X	X	X	X	X	X	X	X	X
9. Depreciation Charges, Total	na	na	na	na	na	X	X	X	X	X
(a) Structures										
(b) Machinery	na	na	na	na	na	X	X	X	X	X
10. Rental Payments, Total	X	X	X	X	X	X	X	X	X	X
(a) Structures	X	X	X	X	X	X	X	X	X	X
(b) Machinery	X	X	X	X	X	X	X	X	X	X

Table 1. The Longitudinal Establishment Data File Directory
Continued—

Description	Years 72	73	74	75	76	77	78	79	80	81
11A. Purchased Fuel (Quantity, Cost, and Stock*)	na	X	X	X	X	X	X	X	X	X
(a) Coal	na	X	X	X	X	X	X	X	X	X
(b) Coke	na	X	X	X	X	X	X	X	X	X
(c) Distillate fuel oil	na	X	X	X	X	X	X	X	X	X
(d) Residual fuel oil	na	X	X	X	X	X	X	X	X	X
(e) Natural gas	na	X	X	X	X	X	X	X	X	X
(f) Liquefied petroleum gas	na	na	na	na	na	na	X	X	X	X
(g) Other fuels	na	X	X	X	X	X	X	X	X	X
11B. Nonpurchased Fuels Used										
(a) Type of fuel	na	na	na	na	na	na	X	X	X	X
(b) Percent of total fuel used	na	na	na	na	na	na	X	X	X	X
12. Methods of Inventory Valuation	na	na	na	X	X	X	X	X	X	X
13A. Status of Establishment	X	X	X	X	X	X	X	X	X	X
13B. Legal Form of Organization	X	na	na	na	na	X	na	na	na	na
14. First Year of Operations	na	na	na	X	na	na	na	na	na	X
15. Unfilled Orders (Single-Units Only)	na	na	na	na	X	X	X	X	X	X
16. Consistency Checks (On Form but not Keyed)	X	X	X	X	X	X	X	X	X	na
17. Detailed Materials Consumed and Water Usage	X	na	na	na	na	X	na	na	na	na
18A. Products and Services										
(a) Product class code	na	X	X	X	X	na	X	X	X	X
(b) Product (7-digit) code	X	na	na	na	na	X	na	na	na	na
(c) Value of Shipments	X	X	X	X	X	X	X	X	X	X
18B. Value of Shipments										
(a) To other plants of same company	na	na	na	na	X	X	X	X	X	X
(b) For export	na	na	na	na	X	X	na	na	X	X
18C. Other Receipts										
(a) For work or services performed	X	X	X	X	X	X	X	X	X	X
(b) Resales	X	X	X	X	X	X	X	X	X	X
(c) Miscellaneous	X	X	X	X	X	X	X	X	X	X
19. Special Inquiries for Selected Industries	X	na	na	na	na	X	na	na	na	na

* = Included for 1978-1981 only.

CASE STUDY 12

STATISTICS OF INCOME PROGRAM

I. Purpose

The internal Revenue Service, in addition to its primary mission of enforcing the Federal tax laws, is also charged with publishing statistics on the operation of the tax laws. The data, based on tax returns, are released in a series of reports called Statistics of Income (SOI).

The SOI reports from the very beginning (1916) have been used extensively for tax research and for estimating revenue, especially,by officials in the Department of the Treasury. The main emphasis of the annual statistics has always been individual and corporation income tax data. Other subjects based on other types of returns for which data have been tabulated either annually or periodically have been partnerships, estates and gifts, fiduciaries, farmers' cooperatives, and foundations and other tax exempt organizations. Data are also published on the international income and taxes of U.S. persons and corporations.

Traditionally, the SOI Program has been based on cross-sectional samples. However, these statistics told very little about the relationships between events that were being described. For example, was it the people who moved who achieved increases in income? Did people whose tax rates went down give more or less to charitable organizations? Only with longitudinal studies has IRS been able to relate status at one point in time to status at another. This is done by focusing on specified observational units in one Year, and following their status through successive (or preceding) years. In addition, when dealing with attitudes, such as the response of taxpayers to tax law and economic changes, longitudinal samples are as close as SOI can come to performing controlled experiments.

Most of the longitudinal studies have been panel studies. The same variables are measured for the same observational units at different periods in time. This is done by creating a file of individual tax return data for a group of taxpayers for each of a succession of years. The IRS has also done transtemporal studies, in which different variables have been measured in different years for the same taxpayers. An example would be the matching of individual income tax returns filed during a taxpayer's lifetime with the estate tax return (which indicates the taxpayer's wealth) filed after his or her death, A third type of longitudinal study is the non-identical study, in which one set of variables is measured for one set of observational units at one time, and another set of variables is measured for a related but not identical group of observational units at another. This occurs when the estate tax return of one individual is matched to the income tax returns filed in later years by his or her heirs.

Because IRS is dealing with administrative files, one more set of distinctions deserves to he made. Each of the types of longitudinal studies mentioned above can be either prospective or retrospective in nature. In other words, the historical data can be built by going either backwards

From a paper presented to the American Statistical Association by Robert A. Wilson and John DiPaolo, and a presentation to the Joint U.S. and Canadian Conference on Tax Modelling by Peter J. Sailer.

or forwards in time from the point at which the sample was selected. The SOI Division has created both types of files, as well as hybrids which move in both directions.

II. Sponsorship

The SOI program is the responsibility of the Statistics of Income Division of the IRS Office of Returns and Information Processing. The Statistics of Income Division is responsible not only for SOI, but also for conducting special statistical studies and providing advice on sample designs for use in helping other organizations in IRS to conduct studies of their own.

III. Sample Design

The SOI program has the following basic character. Returns filed with the ten service centers are processed for administrative purposes to determine the correct tax liability. During processing, the returns are entered on tape for eventual posting to the IRS Master File. It is when the return records are on tape that they are designated for SOI After the returns are designated, they are subjected to additional editing and relational testing for the SOI program.

A. Design Problems

The first task is to identify the same observational units. In the case of individual taxpayers, this is not too difficult, at least in theory. All records are identified by social security number (SSN), and most of the electronic files are sorted in SSN order.

There are many reasons, however, which can cause non-matches. Deaths (in the case of prospective studies) and births (in the case of retrospective studies) guarantee that not all records will match to a record for another year. (Births and deaths mean coming into the system or leaving the system. This leads to the phenomenon that a taxpayer can be born into the estate tax system only by dying.) Unfortunately (for the SOI program), many taxpayers show a tendency to die only temporarily, and then to be reborn a few years later.

However, neither processing errors, nor births, nor deaths create as many problems as marriages. When a male in an SOI panel gets married, he will generally start filing a joint return with his wife, using his SSN as the primary SSN on the return. This means that he will still be in the panel but, in contrast to earlier years, he may well have a second persons's income and taxes mixed in with his. On the other hand, when a female gets married, she is generally lost to a panel, especially if the sample selection is performed at the service centers, where secondary SSN's are not always key-entered. No matter how much effort is made to keep all the observational units from one year to the next, the fact remains that it will not be possible to include completely comparable data items, since joint returns always combine data items for both taxpayers.

The problem of marriages is compounded when one is trying to establish a panel of corporations. While multiple marriages do occur among individuals, at least they occur serially. In the case of corporations, the frequent and cumulative merging of observational units often with units from totally unrelated industrial groupings, can wreak havoc with corporation panels. For that reason, corporation panel studies undertaken by the Statistics of Income Division have been confined to very small pilot efforts.

Although setting up a panel file may be much more complicated than simply selecting a series of cross-sectional samples, panel files have one additional benefit. While the sampling variability of the estimates for each year should be about the same as they would be for a cross-sectional sample of the same size for each year, the sampling variability of the changes from one year to the next should be considerably smaller. This happens because the differences between one year and the next truly are differences, not the results of selecting different samples.

IV. Survey Design and Content

A. The 1967-73 Individual SOI Panel

The 1967-73 panel was created by incorporating two four-digit social security number endings in each stratum of each Statistics of Income sample for those years. In other words, anybody whose SSN ended in one of those two combinations of digits was included in the larger, stratified sample selected to produce the annual Statistics of Income report. In theory, at least, this created a general-purpose panel at a very low cost. The cost of abstracting, keying, and testing important data items from selected tax returns was absorbed as part of the regular statistical processing.

One problem arose because an annual 2 percent delinquency rate added tip to quite a few incomplete observational units over a seven-year period -- over 10 percent, as a matter of fact. Further complications arose because of the many tax law changes and consequent redesign of the tax forms over the 7-year period of the panel. Because of these changes, the file format changed considerably over the period, with old items being dropped and new ones added. IRS finally decided to create a completely new file format, which would work for all the years in the panel. Fields were created for all items that existed over the 7-year period, and were filled in for those years for which they existed.

When the completeness of the file was evaluated, going back only one year (i.e., to 1972), returns for 11.7 percent of the taxpayers in the sample were missing. Going back another year, some of the lost taxpayers reappeared, while others dropped out, for a net loss of 18.4 percent. By the time IRS had gone back 6 years to the beginning of the panel, no returns could be found for

32.6 percent of the 1973 taxpayers. The number for which IRS did not have complete records was closer to 50 percent. In spite of its limitations, the file proved useful in studying a number of issues.

B. The Capital Gains Panel

Beginning with Tax Year 1973, the Statistics of Income Division began assembling "capital gains panels." These are 5-year, retrospective/prospective panels, with the base year in the middle. A highly stratified sample of Schedule D returns (Capital Gains and Losses) with sampling rates ranging from 1/48,000 to 1/5, is selected for the middle year. The IRS Individual Master File is then used to locate the returns for the two previous years and, eventually, for the two following years. The returns are pulled, and details on each capital transaction are edited and transcribed.

C. The Estate Collation Study

While a panel of Forms 1040 can provide information about the realization of capital gains, and a panel of Schedule D data can indicate what type of assets have been traded and how long they have been held, neither shows how these relate to the total wealth of the taxpayer. Wealth, in fact, is reported at most once for any given taxpayer -- on Form 706 (Estate Tax Return), by the taxpayer's estate, after he or she has died. The purpose of the SOI's estate collation studies is to establish a connection between the income and the wealth of taxpayers, and to trace the transfer of wealth (and consequent changes in income) when a taxpayer dies. This is done by matching a decedent's estate tax return first to his or her income tax returns prior to death, then to the beneficiaries' income tax returns both before and after the death. In other words, this is a hybrid of every type of longitudinal study mentioned above: a retrospective and prospective, non-identical, transtemporal panel.

For the 1976 Estate Collation Study, IRS matched estate tax returns filed in 1977 with the decedent's income tax returns filed for the two previous years. In addition, IRS matched the income tax returns for nonspousal heirs to whom a bequest of $50,000 or more had been made, obtaining data for the two years before and the three years after the bequest.

D. Taxpayer Migration Data

This project is probably one of the largest panel studies ever undertaken. It is not done by the Internal Revenue Service, but, it involves data files that are provided by IRS to the Bureau of the Census. The Census matches every computer record of individual income tax returns filed from January through September of a given year to the previous year's record. The Census Bureau is given access to return records, among other things, to make intercensal population and income estimates, and to provide county and minor civil division level data to the Treasury Department for the Federal Revenue Sharing program. The matching of return records is in part an operational necessity. Taxpayers frequently use a business or Post Office Box address on their returns. Therefore, the Bureau persuaded IRS to put a question on the return about the exact governmental unit in which a taxpayer lives. However, this is done only once every few years -- the most recent year was 1980.

Among the series of data which Census creates from these files are matrices which show from where to where the population is shifting; and county migration data which show how many taxpayers entered and left each county within a given period of time, how many exemptions they claimed, and, for some years, the amount of income for the in-migrants, the out-migrants, and the non-migrants.

E. Department of Defense (DOD) Salary Study

The DOD Salary Study is the result of a public law passed by the U.S. Congress which requires the Department of Defense to perform an evaluation of the military pay structure at least once every four-years. Part of this study entails following the earnings of persons who leave the Armed Forces-separatees, as DOD calls them -- to learn what the "opportunity costs" are for persons who stay in the Armed Forces.

The sample of separatees is chosen by DOD. New separatees are sampled each year. Once selected for the sample, the individual stays in it forever. DOD gives IRS the social security numbers of the new designees, along with codes indicating their DOD characteristics. By going to Forms 14-2 (Wage and Tax Statements), rather than to income tax returns, IRS gets only the salaries of the individuals in the sample.

Because of the taxpayer's right to privacy, no identifiable data are returned to DOD. All SSN's are removed from the data before they are sent back to DOD. Furthermore, DOD supplies IRS with at least three individuals with any given combination of DOD characteristics codes, so that there will not be any way to match back to the SSN's.

One of the limitations of this panel is that of missing data. There are no indicators on the Form W-2 to indicate whether a person for whom data are missing is self-employed, unemployed, retired, or dead, or whether IRS has made a processing error. At this point, there is no alternative to simply leaving these individuals out of the analysis.

F. The Individual Panel Beginning with Tax Year 1979

The Tax Year 1979 sample was designed to study certain questions related to mortality and morbidity rates by occupation of taxpayer. Funds had been made available for this purpose by the Social Security Administration and the National Cancer Institute. Since future links to certain data items from the Social Security Administration's Continuous Work History Sample (CWHS) were anticipated, five SSN endings were chosen to overlap with the CWHS sample. There is now a 3-year panel of some 45,000 randomly selected tax return records, and a 4-year panel of 9,000 records.

G. Corporation Tax Adjustment Study (CORTAX)

This study is intended to quantify the effects of adjustments (through carrybacks of net operating losses and unused credits, IRS examination activity, etc.) to corporate tax liability after the corporation's original tax return (Form 1120 series) has been filed. By linking SOI corporate sample EIN's to their Business Master File -(BMF) accounts, SOI expects to tabulate these adjustment amounts for all tax years on the BMF extract -- usually the most recent five or so.

For example, CORTAX 86 will commence in 1986 by extracting these adjustment data for Tax Years 1978 - 1982, using the Tax Year 1982 sample file of EIN's as the extract or link variables. While a significant portion of the SOI corporate on sample (like other SOI sampling frames) is already longitudinal , CORTAX will lend an additional longitudinal aspect with its five years of

adjustment data for each CORTAX year's record. In addition, CORTAX will show cumulative adjustment effects (and, thus, annual changes) for certain tax years over time for the longitudinal "core" of records in the SOI corporate samples.

CORTAX 87 is expected to provide tax liability adjustment data for an accounting period range ending with Tax Year 1985, and may expand tabulations to include interest and penalty assessment amounts as well. Thereafter, CORTAX studies are planned for annual occurrence, and should continue to provide Treasury's Office of Tax Analysis and Congress' Joint Committee on Taxation with the supplemental data bases necessary for the development of more current and detailed tax policy/legislation analyses.

V. Future Studies

There is no doubt that longitudinal studies are essential to the IRS mandate to produce statistics on how the internal revenue laws are operating. A new estate collation study is being planned for 1982 decedents. In this new, improved study, wealth transferred to trusts and other estates, as well as to individuals, will be traced. One of the most ambitious plans is the study of Intergenerational Transfers of Wealth. The only time an actual accounting is available for an heir's wealth will be when that heir, in turn, passes away. This is what the study of intergenerational transfers is all about. By linking estate tax returns filed by succeeding generations of heirs a classic non-identical longitudinal study -- it is possible to study changes in the concentration of wealth during the history of the tax system, and the role intergenerational transfers of wealth have played in this process.

Additional plans for the future include improved individual panel studies using data from the Individual Master File of all tax return records, including one in which the postal ZIP code will be used to trace migration patterns; Also planned are additional capital gains panels, and a panel study of large private

foundations.

152

REFERENCES

ANDERSON, T.W.

1957 "Maximum Likelihood Estimates for a Multivariate Normal Distribution When Some Observations are Missing." Journal of the American Statistical Association, 51, 200-203.

ARTZROUNI, MARC

1980 "Tracing Respondents in Longitudinal Surveys: A Bibliographic Overview." Unpublished Ms., U.S. Bureau of the Census, Statistical Research Division.

BARTHOLOMEW, D.J.

1973 Stochastic Models for Social Processes, John Wiley and Sons.

BENUS, J.

1975 "Response rates and data quality." Five Thousand American Families -- Patterns of Economic Progress, Vol. III, Ann Arbor: Institute for Social Research, G.J. Duncan and J. N. Morgan (eds.).

BIDERMAN, A., CANTOR, D. and REISS. A.

1982 "A Quasi-Experimental Analysis of Personal Victimization Reporting by Household Respondents in the National Crime Survey." Paper prepared for the Joint Statistical Meetings of the American Statistical Association, Cincinnati, Ohio.

BISHOP, YVONNE M.M.; FIENBERG, STEPHEN E.; HOLLAND, PAUL W.;

1975 Discrete Multivariate Analysis, The MIT Press.

BLALOCK, H.M.

1970 Causal Models in the Social Sciences, Chicago: Aldine.

BURKHEAD, D., and CODER, J.

1985 "Gross Changes in Income Recipiency from the Survey of In come and Program Participation". Proceedings of the Social Statistics section, American Statistical Association.

BYE, BARRY V. and SCHECHTER, EVAN S.

1980 "Estimating Response Variance from Latent Markov Models: An Application to Self Reported Disability Status", ORS Staff Paper no. 37.

BYE, BARRY V. and SCHECHTER, EVAN S.

1986 "A Latent Markov Model Approach to the Estimation of Response Errors in Multiwave Panel Data", Journal of the American Statistical Association, forthcoming, June, 1986.

CAMPBELL, RICHARD T. and MUTRAN, ELIZABETH,

1982 "Analyzing Panel Data in Studies of Aging, Research on Aging, Vol.4 , no. 1 , 3-41 .

CITRO, C.F.

1985 "Alternative Definitions of Longitudinal Households and Poverty Status in the ISDP", Proceedings of the Survey Methods Research Section, American Statistical Association.

CLOGG, CLIFFORD C.

1979 "Latent Structure Models of Mobility" The Pennsylvania State University.

CODER, J., and FELDMAN, A.;

1984 "Early Indications of Item Nonresponse on the Survey of Income and Program Participation", in Proceedings of the Survey Methods Research Section, American Statistical Association.

COLEMAN, JAMES

1981 Longitudinal Data Analysis, Basic Books Inc.

COOK, MARTIN A., and ALEXANDER, KARL L.

1982 "design & Substance in Educational Research" Adolescent Attainment, A Case in Point" in Sociology of Education: 53 no. 4: 197-202.

COX, B., and BONHAM, G.

1983 "Sources and Solutions for Missing Data in the NMCUES," in Proceedings of the Survey Research Methods Section, American Statistical Association, Washington, D.C.

COX, B. and COHEN, S.

1985 Methodological Issues for Health Care Surveys. Marcel Dekker, New York.

DAVID, M. (ed.);

1983 Technical, Conceptual and Administrative Lessons of the Income

Survey Development Program. New York, Social Science Research Council.

DAVID, M. and LITTLE, R., and McMILLEN, D.

1983 "Weighting Adjustments for Nonresponse in Panel Surveys." Unpublished working paper, U.S. Bureau of the Census.

DAVID, M., and LITTLE, R.

1983 "Concepts and Strategies for Imputation of ISDP and SIPP." Unpublished working paper, U.S. Bureau of the Census.

DUNCAN, G.J., JUSTER, F.T. and MORGAN J.N.

1982 "The role of panel studies in a world of scarce research resources." Paper prepared for Social Science Research Council Conference on Designing Research with Scarce Resources, Washington, D.C.

DUNCAN, G. & KALTON, G.

1985 Issues of Design and Analysis of Surveys Across Time, paper presented to the I.S.I., August, 1985, Amsterdam.

DUNTEMAN, GEORGE H. and PENG, SAMUEL S.

1977 "Some Analysis Strategies Applied to the National Longitudinal Study of the High School Class of 1972," Research Triangle Institute,

ELANDT-JOHNSON, REGINA C., and JOHNSON, NORMAN L.

1980 Survival Models and Data Analysis, John Wiley and Sons.

ERNST, I., HUBBLE, D., and JUDKINS, D.

1984 "Longitudinal Family and Household Estimation in SIPP". Proceedings of the Survey Research Methods Section, American Statistical Association, Washington, D.C.

FELDMAN, A., NELSON, C., and CODER, J.;

1980 "Evaluation of Wage and Salary Income Reporting on the 1978 Income Survey Development Program Test Panel", in Proceedings of the Section on Survey Research Methods, American Statistical Association.

FERBER, R., and FRANKEL, D,

1981 Evaluation of the Reliability of the Net Worth Data in the 1979 Panel: Asset Ownership on Wave 1. Prepared under contract with the Survey Research Laboratory, University of Illinois.

FIENBERG, S.D., and TANUR, J.M.

1983 "The Design and Analysis of Longitudinal Surveys: Controversties and Issues of Cost and Continuity." Technical Report no. 289, Department of Statistics, Carnegie-Mellon

University, Pittsburgh,,Pennsylvania.

FOX, ALAN

1976 "Work Status and Income Change, 1968-72: Retirement History Study Preview," Social Security Bulletin.

FRANKEL, D.

1985 Survey of Income and Program Participation: Selected Papers given at the 1985 Annual meeting of the American Statistical Associating. Las Vegas, Nevada, 1985.

GINSBERG, RALPH B.

1972a "Critique of Probabilistic Models: Application of the Semi-Markov Model to Migration," Journal of Mathematical Sociology, Vol. 2, 63-82.

GINSBERG, RALPH B.

1972b "Incorporating Causal Structure and Exogenous Information with Probabalistic Models" With Special Reference to Choice, Gravity, Migration and Markov Chains," Journal of Mathematical Sociology, Vol. 2, 83-103.

GROVES, R. M. and KAHN, R.L.

1979 Surveys by Telephone: A National Comparison with Personal Interviews. New York: Academic Press.

HAUSER, ROBERT M.

1978 "some Exploratory Methods for Modeling Mobility Tables and other Crossclassified Data", CDE Working Paper 78-19.

HECKMAN, JAMES J. and SINGER, BURTON

1982 "The Identification Problem in Econometric Models for Duration Data." Advances in Econometrics, Cambridge University Press.

HENNESSEY, JOHN C.

1982 "Testing the Predictive Power of a Proportional Hazards Semi-Markov Model of Postentitlement Work Histories of Disabled Male Beneficiaries", Social Security Administration ORS Working Paper no. 29.

JEAN, A. and McARTHUR E.

1984 "Some data collection issues for panel surveys with application to SIPP." Proceedings of the Survey Methods Section, American Statistical Association

JONES. B.

1982 "Development of Sample Weights for the National Household Component of the National Medical Care Utilization and Expenditure Survey." Research Institute, Research Triangle Park, NC. RTI/1815/05-01F.

J™RESKOG, KARL G., and S™RBOM, DAG

1976 "Statistical Models and Methods for Analysis of Longitudinal Data", In D.J. Aigner and A.S. Goldberger (eds.), Latent Variables in Socioeconomic Models, pp. 285-325. Amsterdam, Holland.

1978 LISREL User's Guide: Version IV, International Educational Services.

1979 Advances in Factor Analysis and Structural Equation Models, Abt Books.

KALACHEK, E.

1978 "Longitudinal surveys & labor market analysis" background
paper no. 6. National Commission on Employment A Unemployment
Statistics, Washington-, D.C.

KALTON, G., KASPRZYK, D., and SANTOS, R.

1981 "Issues of Nonresponse and Imputation in the Survey of Income
and Program Participation " in Current Topics in Survey
Sampling. Krewski, D, Platek, R., Rao, J.N.K. (eds), Academic
Press, New-York.

KALTON, G. and LEPKOWSKI, J.

1992 "Longitudinal Weighting in the ISOP? Chapter 12 in David, op.
Cit. Lessons of the ISDP, D,C., SSRC.

KALTON, G and LEPKOWSKI, J.

1983 "Cross-Wave Imputation," in Technical, Conceptual and Administrative Lessons of the Income Survey Development Program (ISDP). M. David, (ed.), Social Science Research Council, New York.

KALTON, G., LEPKOWSKI, J., and LIN, T.

1985 "Compensating for Wave Nonresponse in the 1979 ISDP Research Panel" in Proceedings of the Survey Research Methods Section, American Statistical Association. Washington, D.C.

KALTON, G., LEPKOWSKI, J., and SANTOS, R.

1981 "Longitudinal Imputation." Survey Research Center/University of Michigan, Income Survey Development Program. Unpublished report, of the ISDP. Department of Health and Human Services, Washington D.C.

KASPRZYK, D., and FRANKEL, D.

1985 Survey of Income and Program Participation and Related Longitudinal Surveys: 1984; Selected Papers Given at the 1984 Annual Meeting of the American Statistical Association. Philadelphia, Pa.

KASPRZYK, D., and KALTON, G.

1983 "Longitudinal Weighting in the Income Survey Development Program," in Technical. Conceptual and Administrative Lessons of the Income Survey Development Program (ISDP). M. David (ed.), Social Science Research Council, New York.

LAND, K.C.

1971 "On the Definition of Social Indicators", in American

Sociologist. 6:322.

LANDIS, RICHARD J., and KOCH, GARY G.

(N.D.)"The Analysis of Categorical Data in Longitudinal

Studies of Behavioral Development", (source unknown).

LANDIS, RICHARD J.; STANISH, WILLIAM M.; FREEMAN, JEAN S.; KOCH,

GARY G.

1976 "A Computer Program for the Generalized Chi-Square Analysis,

of Categorical Data Using Weighted Least Squares (GENCAT)",

Computer Programs in Biomedicine, 6: 196-231.

LITTLE, R.

1984 "Survey Nonresponse Adjustments* in Proceedings of the Survey

Research Methods Section, American Statistical Association,.Washington, D.C.

LITTLE, R.

1985 "Nonresponse Adjustments in Longitudinal Surveys: Models for Categorical Data." Paper prepared for the meeting of the International Statistical Institute, August 1985.

MARINI, M., OLSEN, A., and RUBIN, D.

1980 "Maximum-Likelihood Estimation in Panel Studies with Missing Data," in Sociological Methodology. Schuessler,, K.F. (ed.), Jossey-Bass, San Francisco.

McARTHUR, E., and SHORT, K.;

1985 "The Characteristics of Sample Attrition in the Survey of Income and Program Participation" in Proceedings of the Survey Research Methods Section, American Statistical Association.

McMILLE14, D., and HERRIOT, R.A.;

1984 "Toward a Longitudinal Definition of Households", in
Proceedings of the Social Statistics Section, American
Statistical Association.

McMILLEN, D., and KASPRZYK, D.;

1985 "Item Nonresponse in SIPP", in Proceedings of the Survey
Research Methods Section, American Statistical Association.

MOORE, J., AND KASPRZYK, K.

1984 "Month-to-Month Recipiency Turnover in the ISDP", in
Proceedings of the Survey Research Methods Section, American
Statistical Association.

NELSON, D., McMILLEN, K., and KASPRZYK D.

1983 "An overview of the survey of income and program participation." U.S. Bureau of the Census, Washington, D.C.

OHIO STATE UNIVERSITY, The;

1979 The National Longitudinal Surveys Handbook. First Edition.

Columbus: Center for Human Resource Research.

1982 The National Longitudinal Surveys Handbook. Second Edition.

Columbus: Center for Human Resource Research.

PARNES, H.S.

1972 Longitudinal Surveys: Prospects and Problems" in Monthly Labor

Review, 95 no.2:11-15.

RHOTON, P.

1983 "Attrition and the National Longitudinal Surveys of Labor
Force Behavior: Avoidance, Control and Correction".
Unpublished mss.,

RUBIN, D.

1974 "Characterizing the Estimation of Parameters in Incomplete
Data Problems." Journal of the American Statistical
Association, 69, 467-474.

SATER, D.,

1985 "Enhancing Data from the Survey of Income and Program
Participation with Data from Economic Censuses and Surveys".
SIPP Working Paper series no. 8505, Bureau of the Census.

SINGER, BURTON.

1983 "Longitudinal Data Analysis", in N. Johnson and S. Kotz (eds.), Encyclopedia of Statistical Sciences, Vol. IV, John Wiley and Sons.

SINGER, BURTON and SPILERMAN, SEYMOUR

1976 "Some Methodological Issues in the Analysis of Longitudinal Surveys", The Annals of Economic and Social Measurement, Vol. 5 no. 4 Fall, 447-474.

TUMA, NANCY BRANDON

1976 "Rewards Resources, and the Rate of Mobility: A Nonstationary Multivariate Stochastic Model," American Sociological Review, Vol. 41, 338-360.

TUMA, NANCY BRANDON and HANNAN, MICHAEL T.

1984 Social Dynamics: Models and Methods., Academic Press.

U.S. BUREAU OF THE CENSUS;

1982 Wage and Salary Data from the Income Survey Development
Program, Current Population Reports, Series P-2 , no. 11 .
U.S.G.P.O., Washington, D.C.

1983 Economic Characteristics of Households in the United States: -
third Quarter, 1983, Current Population Reports, Series P-70,
no. 1. U.S.G.P.O., Washington, D.C. (This series is published
with quarterly information. no. 5 in the series, containing
data for fourth quarter 1984, was released November, 1985.)

U.S. DEPARTMENT OF COMMERCE;

1978 A Framework for Planning U.S. Federal Statistics for the 80's.
Office of Federal Statistical Policy and Standards.
Washington, D.C.

U.S. SOCIAL SECURITY ADMINISTRATION, ORLANDO, FLORIDA

1982 "disability Insurance Work Incentive Experiments: Project Statement", SSA/OP/ORDS/DDS, March.

U.S. SOCIAL SECURITY ADMINISTRATION

(N.D.)"Retirement History Study Report Series", Social Security Administration Publication no. 73-11700.

VAUGHN, D., WHITEMAN, T., and LININGER, C.;

1984 "The Quality of Income and Program Data in the 1979 ISDP Research Panel: Some Preliminary Findings", in Review of

Public Data Use,"Vol. 12, no. 2, pp. 107-131.

WHITE, G.D. JR., and HUANG, H.

1982 "Mover Followup Costs for the Income Survey Development Program" paper presented at the Joint Statistical Meetings of the American Statistical Association et al., Cincinnati, Ohio, August.

WHITMORE, R. , COX, B. , and FOLSOM, R.

1982 Family Unit Weighting Methodology for the National Household Survey Component of the National Medicaid Care Utilization and Expenditure Survey. Research Triangle Institute, Research Triangle Park, N.C. RTI/1898/06-03F.

WILLIAMS, W.H. and C.L. MALLOWS

1970 "Systematic Biases in Panel Surveys," in JASA 65: 1338-1349.

YCAS, MARTYNAS A.

1982 "survey Design and Panel Attrition". Paper no. 11 in David, Op. Cit. pp. 147-154.

YCAS, MARTYNAS A. and LININGER, C.;

1981 "The Income Survey Development Program: Design Features and Initial Findings" in Social Security Bulletin, vol. 44, no. ii.

Reports Available in the

Statistical Policy

Working Paper Series

1. Report on Statistics for Allocation of Funds; GPO Stock Number 003-005-00178-6, price $2.40

2. Report on Statistical Disclosure and Disclosure-Avoidance Techniques; GPO Stock Number 003-005-00177-8, price $2.50

3. An Error Profile: Employment as Measured by the Current Population Survey; GPO Stock Number 003-005-00182-4, price $2.75

4. Glossary of Nonsampling Error Terms: An Illustration of a

Semantic Problem in Statistics (A limited number of copies are available from OMB)

5. Report on Exact and Statistical Matching Techniques; GPO Stock Number 003-005-00186-7, price $3.50

6. Report on Statistical Uses of Administrative Records; GPO Stock Number 003-005-00185-9, price $5.00

7. An Interagency Review of Time-Series Revision Policies (A limited number of copies are available from OMB)

8. Statistical Interagency Agreements (A limited number of copies are available from OMB)

9. Contracting for Surveys (Available through NTIS Document Sales, PB83-233148)

10. Approaches to Developing Questionnaires (Available through NTIS Document Sales, PB84-105055)

11. A Review of Industry Coding Systems (Available through NTIS Document Sales, PB84-135276)

12. The Role of Telephone Data Collection in Federal Statistics (Available through NTIS Document Sales, PB85-105971)

13. Federal Longitudinal Surveys (Available through NTIS Document Sales, PB86-139730)

Copies of these working papers, as indicated, may be ordered from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 (202-783-3238) or from NTIS Document Sales,

5285 Part Royal Road, Springfield, VA 22161 (703-487-4650).