



**Statistical Policy
Working Paper 26**

**Seminar on Statistical Methodology
in the Public Service
Part 2 of 2**

Federal Committee on Statistical Methodology

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

August 1997

THE FEDERAL COMMITTEE ON STATISTICAL METHODOLOGY
(November 1996)

MEMBERS

| | |
|---|--|
| Nancy J. Kirkendall, Chair Office of Management and Budget | Robert P. Parker Bureau of Economic Analysis |
| Susan W. Ahmed National Center for Education Statistics | Charles P. Pautler, Jr. Bureau of the Census |
| Yvonne M. Bishop Energy Information Administration | David A. Pierce Federal Reserve Board |
| Cynthia Z. F. Clark Bureau of the Census | Thomas J. Plewes Bureau of Labor Statistics |
| Steven Cohen Administration for Health Policy Research | Wesley L. Schaible Bureau of Labor Statistics |
| Lawrence H. Cox Environmental Protection Agency | Rolf R. Schmitt Bureau of Transportation Statistics |
| Zahava D. Doering Smithsonian Institution | Monroe G. Sirken National Center for Health Statistics |
| Gerald Gates Bureau of the Census | Alan R. Tupek National Science Foundation |
| Robert Fay Bureau of the Census | Denton R. Vaughn Social Security Administration |
| Daniel Kasprzyk National Center for Education Statistics | Robert Warren Immigration and Naturalization Service |
| Daniel Melnick National Science Foundation | G. David Williamson Centers for Disease Control and Prevention |

EXPERT CONSULTANTS

| | |
|--|--|
| Robert Groves Joint Program in Survey Methodology | Graham Kalton Joint Program in Survey Methodology |
|--|--|

DEDICATION
to
Maria Elena Gonzalez

The Seminar on Statistical Methodology in the Public Service was dedicated to the memory of Maria Elena Gonzalez, the founder and leader of the Federal Committee on Statistical Methodology (FCSM) from its inception in 1975 until her death in 1996. Maria worked -- often quietly -- but always tenaciously to improve the quality of the statistics our government provides.

No single professional accomplishment is more closely associated with Maria Gonzalez than her stewardship of the FCSM. This novel approach to leveraging the talents of our statistical community was fully Maria's creation. It is a tribute to her vision that the current environment renders interagency cooperation ever more important and relevant. We are pleased to honor Maria by vowing that the FCSM will continue to be a focal point for collaboration among Federal Agencies on matters of statistical methodology and statistical practice.

Each of us is better for having known Maria Gonzalez. Her boundless energy and intellect, exemplified in her leadership of the FCSM, are at once a legacy and a challenge to us.

Katherine K. Wallman
Chief Statistician
Office of Management and Budget

PREFACE

The Federal Committee on Statistical Methodology was organized by the Office of Management and Budget (OMB) in 1975 to investigate issues of data quality affecting Federal statistics. Members of the committee, selected by OMB on the basis of their individual expertise and interest in statistical methods, serve in a personal capacity rather than as agency representatives. The committee conducts its work through subcommittees that are organized to study particular issues. The subcommittees are open by invitation to Federal employees who wish to participate. Since 1978, 26 Statistical Policy Working Papers have been published under the auspices of the Committee.

On November 12-13, 1996, the Council of Professional Associations on Federal Statistics (COPAFS) hosted a "Seminar on Statistical Methodology in the Public Service." Developed to capitalize on work undertaken during the past fifteen years by the Federal Committee on Statistical Methodology and its subcommittees, the seminar focused on a variety of topics that have been explored thus far in the Statistical Policy Working Paper series and on work on statistical standards done by the Statistical Policy Office. The subjects covered at the seminar included:

- Electronic Dissemination of Federal Statistics
- What the Public Needs to Know About Federal Statistics
- Training Federal Statisticians
- Measuring Customer Satisfaction
- One-Stop Shopping for Federal Statistics
- Reviewing and Reporting Quality in Survey Data
- Performance Measurement in Statistical Agencies
- Survey Integration: Initiatives in Health Data
- Sharing Data for Statistical Purposes
- Reinventing Economic Classifications
- Using Administrative Records for Statistical Purposes
- Validation of Cognitive Questionnaire Pretesting Methods

Each of these topics was presented in a two-hour session that featured formal papers and discussion, followed by informal dialogue among all speakers and attendees.

Statistical Policy Working Paper 26, published in 2 parts, presents the proceedings of the "Seminar on Statistical Methodology in the Public Service." In addition to providing the papers and formal discussions from each of the twelve sessions, this working paper includes Victor Cohn's keynote address, "What We the Public Need to Know," and comments by Graham Kalton and Nancy J. Kirkendall at the closing session, "Toward an Agenda for the Future."

We are indebted to all of our colleagues who assisted in organizing the seminar, and to the many individuals who not only presented papers but who also prepared these materials for publication.

Table of Contents

Tuesday, November 12, 1996

Part 1

KEYNOTE ADDRESS

| | |
|---|--|
| WHAT WE, THE PUBLIC, NEED TO KNOW.....1 | |
| Victor Cohn, Reports on Health | |

Session 1 - ELECTRONIC DISSEMINATION OF FEDERAL STATISTICS

| | |
|--|--|
| ELECTRONIC DISSEMINATION OF ENERGY DATA AND ANALYSIS.....9 | |
| Henry Weigel and Mark Rodekoher, Energy Information Administration | |
| DISCUSSION.....13 | |
| Robyn Levine, Congressional Research Service | |

Session 2 - WHAT THE PUBLIC NEEDS TO KNOW ABOUT FEDERAL STATISTICS

| | |
|--|--|
| ORGANIZATIONAL PERSPECTIVES AND THE FEDERAL STATISTICS AGENDA: HOW WHAT WE KNOW REFLECTS WHO WE ARE.....17 | |
| Daniel Melnick, National Science Foundation | |
| IMMIGRATION STATISTICS: NO LONGER NEGLECTED, BUT STILL INADEQUATE.....27 | |
| Jeffrey Passell, The Urban Institute | |
| DISCUSSION.....43 | |
| Constance F. Citro, Committee on National Statistics | |
| DISCUSSION.....52 | |
| TerriAnn Lowenthal, Rothleder-Lowenthal Group | |

Session 3 - TRAINING FEDERAL STATISTICIANS

| | |
|---|--|
| FCSM SUBCOMMITTEE ON TRAINING FEDERAL STATISTICIANS AND STATISTICAL TRAINING.....59 | |
| Cynthia Z.F. Clark, Bureau of the Census | |
| FEDERAL STATISTICAL CAREER DEVELOPMENT PROGRAMS.....65 | |
| Fred S. Barrett, National Agricultural Statistics Service, Charles P. Pautler, Bureau of the Census, and David Williamson, Center for Disease Control | |
| PANEL DISCUSSION: STAFF STATISTICAL SKILLS REQUIREMENTS FOR THE 21ST CENTURY.....76 | |
| Rich Allen, National Agricultural Statistics Service, Cathryn Dippe, Bureau of Labor Statistics, Jeanne Griffith, National Science Foundation and Paula Schneider, Bureau of the Census | |

Session 4 - MEASURING CUSTOMER SATISFACTION

| | |
|--|-----|
| CONDUCTING CUSTOMER SATISFACTION SURVEYS IN THE FEDERAL GOVERNMENT..... | 93 |
| Sam Rives, National Agricultural Statistics Service | |
| MEASURING CUSTOMER SATISFACTION WITH CENSUS BUREAU PRODUCTS..... | 113 |
| John Kavaliunas, Bureau of the Census | |
| COMMENTS: WHEN TO LISTEN AND WHEN TO MEASURE..... | 117 |
| Jerry Coffey, Office of Management and Budget | |

Session 5 - ONE-STOP SHOPPING FOR FEDERAL STATISTICS

| | |
|---|-----|
| ONE-STOP SHOPPING FOR FEDERAL STATISTICS..... | 119 |
| Alan Tupek, National Science Foundation | |
| DISCUSSION..... | 126 |
| Ann Gray, Cornell University | |
| DISCUSSION..... | 128 |
| Lauris Olson, University of Pennsylvania | |

Session 6 - REVIEWING AND REPORTING QUALITY IN SURVEY DATA

| | |
|--|-----|
| QUALITY DECLARATION AT STATISTICS SWEDEN: PRINCIPLES AND PRACTICES..... | 131 |
| Claes Andersson, Haken Lindstrom and Lyrs Lyberg, Statistics Sweden | |
| DATA QUALITY AT THE ENERGY INFORMATION AGENCY: THE QUEST FOR A SUMMARY MEASURE..... | 145 |
| Renee Miller, Energy Information Administration | |
| DISCUSSION: REVIEWING AND REPORTING QUALITY..... | 157 |
| Jay Waite, Bureau of the Census | |
| DISCUSSION: REVIEWING AND REPORTING QUALITY IN SURVEY DATA..... | 161 |
| Richard A. Kulka, Research Triangle Institute | |

Wednesday, November 13, 1996

Session 7 - PERFORMANCE MEASUREMENT IN STATISTICAL AGENCIES

| | |
|--|-----|
| PERFORMANCE BASED MANAGEMENT: USING THE MEASURES..... | 165 |
| Nancy J. Kirkendall and Paul Staller, Energy Information Administration | |
| LINKING RESULTS TO MEASUREMENT AND ACCOUNTABILITY..... | 178 |
| George Wollner, Bureau of Labor Statistics | |
| COMMENTS..... | 202 |
| Robert Wise, Corporate and Government Consulting, Inc. | |

Part 2

Session 8 - SURVEY INTEGRATION: INITIATIVES IN HEALTH DATA

| | |
|---|-----|
| INTRODUCTORY REMARKS..... | 207 |
| Edward J. Sondik, National Center for Health Statistics | |
| THE DESIGN OF THE MEDICAL EXPENDITURE PANEL SURVEY..... | 211 |
| Steven Cohen, Agency for Health Care Policy and Research | |
| IMPLICATIONS FOR NCHS DATA SYSTEMS..... | 250 |
| Jennifer H. Madans and Edward L. Hunter, National Center for Health Statistics | |
| DISCUSSION..... | 257 |
| Steven G. Heeringa, University of Michigan | |
| DISCUSSION..... | 261 |
| William D. Kalsbeek, University of North Carolina | |

Session 9 - SHARING DATA FOR STATISTICAL PURPOSES

| | |
|--|-----|
| DATA SHARING: A CENSUS BUREAU PERSPECTIVE..... | 265 |
| Thomas L. Mesenbourg, Bureau of the Census | |
| DATA SHARING: A NASS/USDA PERSPECTIVE..... | 274 |
| Douglas Kleweno and Robin Roark, National Agricultural Statistics Service | |
| COMMENTS..... | 289 |
| James R. Nunns, Department of the Treasury | |

Session 10 - REINVENTING ECONOMIC CLASSIFICATION

| | |
|---|-----|
| REINVENTING OCCUPATIONAL CLASSIFICATIONS..... | 293 |
| Thomas J. Plewes, Bureau of Labor Statistics | |
| DISCUSSION..... | 302 |
| David W. Stevens, University of Baltimore | |

Session 11 - USING ADMINISTRATIVE RECORDS FOR STATISTICAL PURPOSES

| | |
|---|-----|
| CENSUS BUREAU USES OF ADMINISTRATIVE DATA..... | 305 |
| Ron Prevost, Bureau of the Census | |
| THE IMPORTANCE OF LINKED ADMINISTRATIVE-SURVEY DATA FOR UNDERSTANDING THE SOCIAL SECURITY PROGRAM..... | 316 |
| Bernard Wixon, Social Security Administration | |
| DISCUSSION..... | 347 |
| Michael V. Leonesio, Social Security Administration | |

Session 12 - VALIDATION OF COGNITIVE QUESTIONNAIRES PRETESTING METHODS

| | |
|---|-----|
| THE VALIDITY OF COGNITIVE INTERVIEW FINDINGS: | |
| SOME COMMENTS AND CAUTIONS..... | 353 |
| Paul Beatty, National Center for Health Statistics | |
| INSIGHTS FROM BEHAVIOR CODING AND COGNITIVE INTERVIEWS ON | |
| QUESTION DEVELOPMENT AND QUESTIONNAIRE DESIGN..... | 363 |
| Jennifer Hess, Bureau of the Census and Ellen Singer, University of Michigan | |
| DISCUSSION..... | 379 |
| Robert F. Belli, University of Michigan | |
| DISCUSSION..... | 383 |
| Theresa J. DeMaio, Bureau of the Census | |

TOWARD AN AGENDA OR THE FUTURE

| | |
|---|-----|
| Graham Kalton..... | 387 |
| Westat, Inc. and Joint Program in Survey Methodology, University of Maryland | |
| Nancy J. Kirkendall..... | 392 |
| Office of Management and Budget | |

Part 2

Session 8

SURVEY INTEGRATION: INITIATIVES IN HEALTH DATA

COUNCIL OF PROFESSIONAL ASSOCIATIONS ON FEDERAL STATISTICS

**November 13, 1996
Bethesda, Maryland**

Introductory Remarks

Session on Survey Integration: Initiatives in Health Data

**Dr. Edward J. Sondik
Director
National Center for Health Statistics**

I. INTRODUCTION

Welcome. Thank you for coming. I am pleased to see so many in attendance for a very important session. I think this session is important because it deals with a topic of enormous potential for the Federal statistical community as a whole--survey integration. And this is also an important session because we will hear from two people with much to tell us about survey integration in operation. They will be focusing on the specific plans and achievements in integrating health surveys at the Department of Health and Human Services, with a specific focus on the role and activities at the National Center for Health Statistics and at the Agency for Health Care Policy and Research.

But first, I would like to provide some background on the impetus for survey integration at HHS.

I would also suggest that we consider the potential for that type of effort beyond an individual Department, since many of the reasons that HHS turned to survey integration are at play *within* organizations and *among* the varied and many statistical activities of the Federal government.

II. IMPETUS FOR INTEGRATION AT HHS

- ◆ **Overlap and Duplication.** HHS annually conducts multiple, decentralized general purpose and program-relevant health surveys. Many of these surveys overlap in terms of populations, topics of interest, and collection methods. This overlap sometimes places undue burden on survey respondents.
- ◆ **Major Gaps in Data.** Despite well-designed individual surveys and the resources applied to data gathering, there still exists major gaps in the kinds of data needed to effectively assess the health status of the population, the access to and quality of health care, and the impact of changes in the health care system. In particular data to evaluate the economic aspects of health care are inadequate.
- ◆ **Inability to Analytically Link Data from Various Sources.** Despite the overlap in populations and content, data from various HHS surveys usually could not be analyzed in concert or linked to increase the analytical power of the data. With varying definitions, data standards, independent sampling frames and survey methodology, the surveys are not compatible or comparable.
- ◆ **Achieve Efficiencies.** HHS spends a considerable amount on health surveys and the integration of surveys offers a way to reduce costs or at least to achieve efficiencies that allow us to fill data gaps without increased funding.

II. EVOLUTION OF SURVEY INTEGRATION

A comprehensive survey integration plan does not spring forth fully-developed and ready for implementation. Needless to say there was much negotiation, consultation, one step forward/two steps back, in putting together a plan of this scope and impact. From an initial

concept of consolidation we realized that we were integrating and linking and that we were not eliminating surveys, per se, but making the structure more rational and streamlined. When we looked at the HHS data collection activities in their entirety we were able to develop a structure which met the data needs of specific programs while providing the array of data needed for public policy and public health management. The new framework provided the justification for efficient investment in data.

From an initial push to collect and expand data on health expenditures to meet a critical need for information to better manage health care resources and services, we moved to a balanced approach where economic statistics were integrated and became an important component of the overall system.

IV. DATA SHARING

We believe that we're making good progress in HHS, but the road is long and there are many paths and a few pitfalls. Looking at the potential for integration or collaboration on a more global perspective, brings us to the issue of data sharing within the Federal statistical community. I am sure that many of you know that legislation was submitted to Congress last year to permit limited sharing of statistical information by agencies within and among the 8 data centers created by the legislation. NCHS was one of the data centers, along with Census, Bureau of Labor Statistics, Bureau of Economic Analysis, National Center for Education Statistics, DOE's statistics division, National Agricultural Statistics Service and the National Science Foundation.

A primary objective of the Act was to reduce duplication of Federal data collection efforts and the reporting burden on the public. The Act envisioned agencies working collaboratively to reduce costs and improve data products. In effect, data collected by one Federal statistical agency

could be shared with another. Plans to share data on a prospective basis could lead to some standardization in data policy and methodology which would enhance analytical capacity. Agencies would also have to deal with issues of confidentiality and ensure that individual agency requirements were met as well as any new regulations. It's too soon to know if the legislation will be reintroduced in this Congress, or its fate if that should happen, but the legislation does offer real potential to reap the benefits of data sharing and collaboration across a much wider range of programs and a much broader spectrum of issues. Of course the challenges of inter-agency collaboration magnify the ones inherent in intra-agency efforts. Those are not negligible, however, and our next two speakers will tell us about some of the creativity and innovation which had to be applied to move ahead on the HHS Survey Integration Plan.

The Redesign of the Medical Expenditure Panel Survey
A Component of the DHHS Survey Integration Plan

Steven B. Cohen
Center for Cost and Financing Studies

Agency for Health Care Policy and Research
2101 E. Jefferson St., Suite 500
Rockville, Maryland 20852

Presented at the COPAFS sponsored seminar on
"Statistical Methodology in the Public Service"
Bethesda, Maryland, November 12-13, 1996.

The views expressed in this paper are those of the author and no official endorsement by the Department of Health and Human Services or the Agency for Health Care Policy and Research is intended or should be inferred. The author wishes to thank Joel Cohen, Amy Taylor and Mamatha Pancholi of the Center for Cost and Financing Studies, AHCPR, for their helpful comments in a review of sections of this paper, in addition to Steve Botman and Marcie Cynnamon of the National Center for Health Statistics for their provision of background information on the design of the National Health Interview Survey.

KEY WORDS: MEPS, NHIS, survey integration copafs.116

The Redesign of the Medical Expenditure Panel Survey
A Component of the DHHS Survey Integration Plan

Steven B. Cohen, AHCPR

1. Introduction

The Household Component of the 1996 Medical Expenditure Panel Survey (MEPS) was designed to produce national and regional estimates of the health care utilization, expenditures, sources of payment and insurance coverage of the U.S. civilian non-institutionalized population. The MEPS includes surveys with medical providers (MPS), employers and other health insurance providers (HIPS) to supplement the data provided by household respondents. The design of the MEPS survey permits both person based and family level estimates. The scope and depth of this data collection effort reflects the needs of government agencies, legislative bodies, and health professionals for the comprehensive national estimates needed in the formulation and analysis of national health policies.

More specifically, the MEPS collects data on the specific health services that Americans use, how frequently they use them, the cost of these services and how they are paid, as well as data on the cost, scope, and breadth of private health insurance held by and available to the U.S. population. MEPS is unparalleled for the degree of detail in its data, as well as its ability to link health service medical expenditures and health insurance data to the demographic, employment, economic, health status, and other characteristics of survey respondents. Moreover, MEPS is the only national survey that provides a foundation for estimating the impact of changes in sources of payment and insurance coverage on different economic groups or special populations of interest, such as the poor, elderly families, veterans, the uninsured, and racial and ethnic minorities.

In this paper, the sample design of the Medical Expenditure Panel Survey (MEPS, also referred to as the National Medical Expenditure Survey (NMES-3)) is described, with particular attention to the use of the 1995 National Health Interview Survey (NHIS) as the sample frame for the survey. The redesigned Medical Expenditure Panel Survey reflects the first stage of implementation of the Department of Health and Human Services' (DHHS) Survey Integration Plan, which provides directives targeted to the improvement in the analytic capacity of programs, the filling of major data gaps, and the establishment of a framework in which DHHS data activities are streamlined and rationalized. Through this effort, specifically through a linkage to the National Health Interview Survey, the MEPS has achieved a number of significant design improvements and analytic enhancements.

Attention is given to the resultant design efficiencies and enhancements in analytical capacity that have been and will be realized through MEPS sample design integration with the NHIS. The report includes a summary of sample size specifications and precision targets for national population estimates and health care expenditure estimates for policy relevant population

subgroups. A discussion is also provided regarding the modification of the Medical Expenditure Panel Survey from a periodic annual survey to an on-going continuous data collection effort with each expenditure panel of households followed for two years. Finally, the paper provides a summary of the respective survey designs that characterize the linked surveys of medical providers (MPS), employers and other health insurance providers (HIPS) associated with MEPS household participants, and their design enhancements attributable to the DHHS Survey Integration Plan.

2. Background

The original analytical goals for the 1996 National Medical Expenditure Panel Survey and budget constraints required that the sample design meet the following requirements:

- o The full series of interviews should be completed in approximately 14,600 households.
- o The sample should be spread over at least 120 separate areas to represent the civilian non-institutionalized population of the 50 states and the District of Columbia.
- o The sample should produce approximately unbiased national estimates and estimates at the census region level.
- o The sample should meet predesignated precision requirements for the following population subgroups of analytical interest: adults with functional impairments, children with limitations, individuals between the ages of 18-64 predicted to have high levels of medical expenditures, individuals with family incomes less than 200 percent of the poverty level, and the elderly (aged 65+).

The specification of at least 120 separate areas was intended to insure sufficient geographic dispersion of the sample and allow for separate regional estimates. The precision specifications were provided to insure that the design would meet analytical objectives and to facilitate stage specific sample size determination. Furthermore, it was recognized that in order to achieve these requirements for the overall population and for specified domains of interest, an initial sample substantially larger than the 14,600 households would need to be screened to permit oversampling of the policy relevant population subgroups. Consequently, the original design of the 1996 National Medical Expenditure Survey called for an independent screening round of data collection in the fall of 1995, with a self-weighting sample design that required the completion of screener interviews in a nationally representative sample of 30,000 addresses.

The original sample design for the NMES-3 household survey (HS) consisted of a multi-stage stratified national area probability sample of households (and noninstitutional group quarters) developed to represent the U.S. civilian non-institutionalized population in all 50 states and the District of Columbia. The design considered the following stages of sample selection: 1) selection of Primary Sampling Units (PSUs); 2) selection of segments within PSUs; 3)

selection and screening of households within segments; and 4) selection of households based on socio-demographic characteristics (both households and individual) from the set of screened households.

The Primary Sampling Units (PSUs) were to consist of counties or groups of counties. The sample reflected a union of national samples independently selected by Westat and NORC (the data collection organizations) for general use, and consisted of 162 PSUs (100 from NORC, 62 from Westat), located in 125 separate geographic sites (reflecting overlap between some PSUs selected by both Westat and NORC). Within PSUs, a sample of 2,585 segments were to be selected, with segments consisting of one or more blocks as defined by the Census Bureau (DiGaetano, 1994).

Within sample segments, all residential addresses were to be listed, from a which subsample would be selected for screening in the fall of 1995. Sample selection of segments and addresses were to be specified as self-weighting, resulting in an equal-probability sample of occupied dwelling units (DUs) across the nation. Several subpopulations were targeted for oversampling to improve the precision of the estimates for those specific portions of the population. Within each sampled DU, screening information was to be obtained for a single reporting unit (RU). An RU is defined to be a person, a group of related persons, or two or more persons living together as a family unit. Based on the screening information obtained and any imputations for missing data, an RU would have been assigned to one of six sampling domains, representing the different subpopulations of analytic interest. A sampling algorithm was to determine whether that RU (and therefore any other RUs in the same DU) was to be included in the sample for the year-long survey. The screener interview was to be conducted as a computer assisted personal interview (CAPI), with sample selection for the core expenditure survey done concurrently while the interviewer was conducting the interview. All RUs sampled into the main household survey through the CAPI algorithm specified to achieve sample size targets for the policy relevant groups were to be administered a baseline interview that obtained information of health insurance coverage and consumer satisfaction measures, which was also to be conducted as a CAPI interview.

2.1 Original Precision Requirements

The sample was designed to produce unbiased national estimates and unbiased estimates for the four Census regions. Further, the sample was designed to meet fixed precision requirements for the nation and for the following policy relevant population subgroups:

1. individuals aged 18 or older with functional impairments (at least 1 ADL);
2. children aged 17 or younger with physical limitations;
3. individuals aged 18-64 predicted to have high medical expenditures in 1996 (top 15%);
4. individuals predicted to have family incomes less than 200 percent of the poverty level in 1996;
5. individuals 65 years of age or older.

An overall precision requirement for the survey was the achievement of an average design effect of 1.7 for the survey estimates of health care expenditure and utilization measures that characterized the policy relevant population subgroups.

Precision requirements for the original NMES3 Household Survey were stated in terms of national estimates at the person level (presented in Table 3). To meet these requirements, the survey had to include a minimum number of persons in each subdomain of interest. However, the unit of interviewing and subsampling was specified as the household. Thus, a subset of the 30,000 screened households were to be selected for the full panel household survey on the basis of the characteristics of the persons they include. There were originally six sample domains of interest to which a screened DU could be assigned. These six domains and their corresponding sampling rates necessary to satisfy survey precision requirements appear in Table 1. The domains are listed in priority order in the sense that if an RU contains persons who fall in different domains, the DU containing the RU was to be assigned to the domain of highest priority for sampling purposes. For analysis purposes sampled persons can be assigned to any analytic group to which they belong.

Table 1. Original sample domains and sampling rates

| Domain | Rate |
|--|------|
| 1. Functionally impaired adults | 1.00 |
| 2. Functionally impaired children | 1.00 |
| 3. Individuals 18-64 years old with predicted high medical expenditures | 1.00 |
| 4. Individuals with family incomes predicted to be below 200% of poverty level | .70 |
| 5. Persons 65 years or older | .65 |
| 6. All others | .35 |

2.2 Using Predictive Models for Domain Assignments

Since a reporting unit's poverty status classification in 1996 would have been unknown at the time of the administration of the HS Screener interview (fall 1995), a prediction model was to be used to determine whether a household is to be oversampled. More specifically, a logistic regression model has been developed that estimates the probability that a reporting unit will have a family income less than 1.25 times the poverty level in a subsequent year based on the poverty status classification and other predictive measures obtained during the screening interview. Households with predicted probabilities above a certain threshold value were to be oversampled. In addition to facilitating an oversample of individuals with family incomes less than 125 percent of the poverty level, use of this prediction model will facilitate an oversample of individuals with family incomes less than 200 percent of the poverty level (Moeller and Mathiowetz, 1994).

The results listed below were observed based on an evaluation of the model's performance at the Reporting Unit level, using data from NMES2, and using a predicted probability of .3 or greater (derived from the logistic regression prediction model) as the criterion to target reporting units most likely to have members with family income less than 200 percent of the poverty level in 1996:

1. Based on the NMES2 experience, the expected prediction rate for true positive (family income less than 200 percent of the poverty level) is 83.1 percent among the 19.5 percent of reporting units predicted to have members with family income less than 200 percent of the poverty level.
2. The expected prediction rate for false negatives is 17.1 percent among the 80.5 percent of reporting units predicted to be other income and with members under the age of 65.

The logistic regression model under consideration was specified at the reporting unit level and requires data on the following measures obtained in the screening interview:

1. Age of reference person;
2. Home ownership;
3. Reporting Unit size;
4. Whether children of specific ages (<6, 6-15) are present in the RU;
5. Whether someone in the RU other than the reference person is at least 65 years of age;
6. Health status of reference person;
7. Race/ethnicity of reference person;
8. Census Division;
9. MSA status of PSU;
10. Education of reference person;
11. Martial status and gender of reference person;
12. Whether reference person or spouse was employed in the previous 3 months;
13. Whether the family income of the reporting unit as less than 1.25 times the poverty level; and

14. Whether anyone in the RU was covered by Medicaid.

Among the groups in Table 1 to be oversampled in the main survey are non-minority individuals between the ages 18-64 who are predicted as likely to incur high medical expenditures in the subsequent year. An individual's medical care expenditures in a future year was unknown at the time of the administration of the HS Screener interview (fall 1995); therefore, a prediction model based on NMES2 data was to be used to determine whether a household was to be oversampled as part of the high medical expenditures group because one or more of the family members are expected to incur high medical expenditures in the subsequent year. More specifically, a logistic regression model has been developed that estimates the expected probability an individual who is between the ages of 18-64 will incur high medical expenditures (top 15 percent of the health expenditure distribution) in a subsequent year based on predictive measures obtained during the screening interview. Households with at least one such person with a predicted probability above a certain threshold value were to be oversampled. The group was restricted to individuals who are between the ages 18-64, since the persons 65 or older were separately targeted for oversampling in the original design specifications (Mathiowetz and Moeller, 1994).

The logistic regression model under consideration is specified at the person level and requires data on the following measures obtained in the screening interview:

1. Gender;
2. Health status;
3. Marital status;
4. Poverty status;
5. Whether the person lives alone;
6. Age;
7. Whether the person's health keeps him/her from working at a job, doing work around the house or going to school;
8. Whether the person is unable to do certain kinds or amounts of work, housework, or schoolwork because of his/her health;
9. The number of visits to a medical doctor or other medical care provider the person has had during the last 6 months;
10. The number of times prescribed medicines were purchased or obtained for the person's use in the last 6 months;
11. Census Division; and
12. MSA status of PSU.

The results listed below were observed based on an evaluation of the model's performance at the individual level, using data from NMES2, and using a predicted probability of .4 or greater (derived from the logistic regression prediction model) as the criterion to target individuals who are between the ages 18-64 and considered likely to incur high medical expenditures in the subsequent year:

Based on the NMES2 experience, the expected prediction rate for true positive was 65.3 percent among the subset of individuals that are predicted to incur high medical expenditures.

3. Redesign of the Medical Expenditure Survey as a Component of the DHHS Survey Integration Plan

As part of the Reinventing Government Part II (REGO II) activities, the Department of Health and Human Services (DHHS) has targeted the improvement of the analytical capacity of HHS programs, the filling of major data gaps, and the establishment of a survey consolidation framework in which HHS data activities are streamlined and rationalized. A Survey Consolidation Working Group was charged with developing a consensus plan for meeting these objectives (HHS Survey Integration Plan, June, 1995). A major concentration of the Survey Integration Plan was focused on the redesign of the health care expenditure and insurance studies conducted by the Department, which includes the National Medical Expenditure Survey, the National Medicare Current Beneficiary Survey, the National Employer Health Insurance Survey and the National Health Interview Survey. The proposed integrated survey design was specified to achieve significant cost efficiencies by eliminating duplicative efforts and reducing overall respondent burden. Furthermore, by virtue of integrating the design features of the component surveys, their respective analytical capacities are enhanced. A number of survey design enhancements were also proposed to improve upon current survey design capabilities. This includes consideration of an on-going longitudinal survey effort, in addition to allowing for a future capacity to derive state specific health care estimates. Consideration was also given to the inclusion of a periodic institutional component to the survey which provides national use and expenditure estimates for the population resident in nursing homes (Hunter et al., 1995).

3.1 Design Enhancements and Efficiencies to be Achieved Through Survey Integration

One of the attractions of the Department of Health and Human Services Survey Integration Plan was the enhanced analytical capacity that would be achieved by the distinct surveys that would be linked through design integration. This could be realized by sample size expansions that would occur through survey mergers such as the planned integration between the Medical Expenditure Panel Survey and the Medicare Current Beneficiary Survey (MCBS) and the consolidation of employer surveys conducted by the Department. In a complementary fashion, use of the NHIS as a sample frame for the MEPS would increase the analytical content of the resultant linked surveys. Through design integration of the respective surveys sponsored by DHHS, inefficiencies associated with duplicative survey efforts would be significantly reduced. Another goal was to achieve reductions in survey design costs attributable to the implementation of a uniform framework for DHHS sponsored surveys with overlapping analytical focus with respect to questionnaire content, data editing, imputation, estimation, database structure and development of analytic files. Additional efficiencies in survey operations

were anticipated in future years as a consequence of the conduct of an annual medical expenditure survey in contrast to a survey that was conducted once every decade.

By moving to this integrated, annual household data collection effort, the Department expands and enhances its analytic capabilities as described below:

- Retains the design of the core NHIS household interview. This core will provide cross-sectional population statistics on health status and health care utilization with sufficient sample size to allow for analyses based on breakdown of detailed age, race, sex, income and other socio-demographic characteristics and will also allow for data on a broad range of topics currently provided by the NHIS.
- Retains the analytical capacity to obtain both annual and quarterly population estimates of health care utilization and the prevalence of health conditions for the nation and for policy relevant population subgroups.
- Provides the ability to model individual (and family-level) health status, access to care and use, expenditure, and insurance behavior over the year and examine the distribution of these measures across individuals. The longitudinal feature of the MEP survey to collect data over multiple years further enhances the capacity to model behavior over time.
- Provides the ability to relate data from a detailed survey sample (e.g., MEP) to a larger population sample (e.g., NHIS) to enhance the utility of the MEP for national health account estimation and microsimulation modeling, including disaggregation by age group or geographic area.
- Provide the potential to expand to State-level estimates for marginal costs using the enhanced 358 PSU sample design of the NHIS.
- The longitudinal (over several years) aspect of the MEPS integrated data collection effort provides the following:
 - An increase in statistical power to examine change or make comparisons over time;
 - The capacity to examine changes over time as well as changes in the relationship among measures of health status, access to care, health care use, expenditures, health insurance coverage, employment, functional limitations and disabilities, and demographic characteristics.
- Provide the potential to expand to State-level estimates for marginal costs.

3.2 MEPS Household Survey

The original NMES sample design called for an independent screening interview to identify a nationally representative sample and facilitate oversampling of policy relevant population subgroups. Associated data collection and training costs associated with this independent screening interview were projected to exceed \$8 million dollars. As part of the DHHS Survey Integration Plan, the separate screening interview to identify the expenditure survey sample was eliminated. As an alternative, the National Health Interview Survey (NHIS) was specified as the sampling frame for the medical expenditure survey, which is referred to as the Medical Expenditure Panel Survey. The NHIS is an on-going annual household survey of approximately 42,000 households (109,000 individuals) conducted by the National Center for Health Statistics to obtain nation estimates for the U.S. civilian non-institutionalized population on health care utilization, health conditions, health status, insurance coverage and access. In addition to the cost savings achieved by the substitution of the NHIS as the MEPS sample frame, the design modification will result in an enhancement in analytical capacity of the resultant survey data. Use of the 1995 NHIS data in concert with the data collected for the 1996 MEPS provides an additional capacity for longitudinal analyses not available in the original design. Furthermore, the greater number and dispersion of the sample PSUs that comprise the MEPS national sample should result in improvements in precision over the original design specifications.

To fill major data gaps identified by the Department of Health and Human Services, the MEPS is specified as a continuous survey with sample peaks at five year intervals. The initial sample of 10,800 NHIS households selected for the 1996 MEPS, is reduced from the original 1996 plan. A rotating panel design will be adopted for the MEPS, where the 1996 panel will be followed for data collection through 1997. A new nationally representative sample of approximately 5,600 households will be selected from the 1996 NHIS to supplement the 1996 panel in order to meet the original precision specifications for the specified policy relevant population subgroups, with the exception of the elderly. A preliminary contact with the NHIS responding households selected for the MEPS study was made prior to the start of the MEPS Survey, to announce the survey and introduce record-keeping activities. The revised study design of the MEPS survey includes several components: the Household Survey (HS) consisting of a rotating panel design in which any given sample panel is interviewed a total of 6 times over three consecutive years to yield annual data for two calendar years; the Medical Provider Survey (MPS) with a sample of medical providers that treated HS persons; and the Health Insurance Provider Survey (HIPS) with a sample of employers and other sources of health insurance of HS persons. The survey is co-sponsored by the Agency for Health Care Policy and Research and the National Center for Health Statistics. Westat and the National Opinion Research Center (NORC) are the data collection organizations for the 1996 MEPS Household Survey.

3.3 MEPS Household Survey Sample Design

The 1996 MEPS Household Survey sample was that was selected from households that responded to the 1995 National Health Interview Survey (NHIS). More specifically, the 1996

MEPS Household sample linked to the 1995 NHIS was selected from a nationally representative NHIS sub-sample from 2 NHIS panels out of 4 to represent the nation, and encompassed half of the households in the NHIS sample during the second and third quarters of 1995. It should be noted that the NHIS has been designed to permit nationally representative subsamples to be selected by restricting the sample to one of four distinct panels. Any combination of 1 to 4 panels will provide a nationally representative sample of households. Furthermore, each NHIS panel subsample for a given quarter of a calendar year is nationally representative.

The complete 1995 NHIS sample consists of 358 primary sampling units (e.g. counties or groups of contiguous counties) with a targeted sample of approximately 42,000 responding households. The sample PSUs selected for the NHIS were stratified by geographic (Census region and state), metropolitan status, and socio-demographic measures (Judkins, Marker and Waksberg, 1994). Within sample PSUs, a sample of blocks (segments) were selected after being stratified by measures of minority population density, which allowed for an oversample of blacks and Hispanics with high minority population concentrations. A nationally representative sample of approximately 71,000 addresses within sampled blocks was selected and targeted for further screening as part of the 1995 NHIS interview.

The nationally representative 1995 NHIS subsample reserved for the 1996 MEPS consists of 195 PSUs, and in the two targeted quarters of 1995 these PSUs include approximately 1,372 sample segments (second stage sampling units) and 10,799 responding NHIS households. This NHIS sample reflects an over-sample of Hispanics and blacks at the following approximate ratios of representation relative to the remaining households (Hispanics 2.0:1, blacks 1.5:1). The MEPS Household Survey sample for 1996 used this nationally representative sub-sample of NHIS households and individuals. Furthermore, this 1996 MEPS panel will be surveyed to collect annual data for two consecutive years.

A new 1997 MEPS panel sample will be selected as a nationally representative subsample from households that respond to the 1996 NHIS. More specifically, the 1997 MEPS sample linked to the 1996 NHIS will be selected from a nationally representative NHIS sub-sample from 2 NHIS panels out of 4 to represent the nation, and will reflect additional subsampling from half of the households in the NHIS sample during 1996 necessary to satisfy the precision requirements specified for the 1997 MEPS Household survey, which generally coincide with the original plan for the 1996 survey. As in 1995, the complete 1996 NHIS sample will consist of 358 primary sampling units (e.g. counties or groups of contiguous counties) with a targeted sample of approximately 42,000 responding households. The nationally representative 1996 NHIS subsample reserved for the 1997 MEPS prior to additional subsampling, will be obtained from the same 195 PSUs selected for the 1996 MEPS Household sample, and include approximately 3,400 sample segments (second stage sampling units) and approximately 21,000 responding NHIS households as eligible for sample selection. Once again, this NHIS sample reflects an over-sample of hispanics and blacks at the following approximate ratios of representation relative to the remaining households (Hispanics 2.0:1, Blacks 1.5:1). A nationally representative subsample of approximately 5,600 NHIS responding households will be selected for the new 1997 MEPS panel. This sample will consist of an oversample of the following policy

relevant population subgroups:

- a. adults (18+) with functional impairments;
- b. children with limitations of activity;
- c. individuals predicted to incur high medical expenditures;
- d. individuals predicted to have incomes LT 200% of poverty level.

An oversample of non-functionally impaired elderly individuals was not planned for in the 1997 survey, given the availability of the 1997 Medicare Current Beneficiary Survey (MCBS), and the planned future survey consolidation of the MCBS and the MEPS. The MCBS is an annual person based survey to obtain the same types of estimates derivable from the MEPS household Survey, on the health care utilization, expenditures, sources of payment and health insurance coverage for Medicare beneficiaries. In addition, the new 1997 MEPS panel will be surveyed to collect annual data for two consecutive years.

As part of the redesign, the 1997 MEPS Household Survey sample will consist of the new nationally representative 1997 MEPS panel in combination with the second year of the 1996 MEPS sample. Overall, the 1997 MEPS Household sample will consist of approximately 13,300 originally sampled NHIS households (adjusted for MEPS Round 1 "split-offs") completing the full series of MEPS interviews to obtain calendar year use and expenditure data for calendar year 1997. Sample selection procedures for the 1997 MEPS sample will be implemented in-house by AHCPR staff, based on data keyed from the 1996 NHIS interviews.

In 1998, a new MEP sample of approximately 5,200 households would be selected as a nationally representative subsample of households that responded to the 1997 National Health Interview Survey. In addition, the entire 1997 panel of 4,808 households would be continued to obtain calendar year 1998 data on health care use and expenditures (with a targeted round specific response rate of 97 percent). Consequently, the MEP sample for 1998 would consist of approximately 9,000 original NHIS households (adjusted for splits in Round 1) completing three core rounds of data collection to obtain calendar year data (4,465 households from the new sample, 4,524 from the 1997 MEP sample). In 1998, the 1996 MEPS Panel would be retired.

For years 1998-2001, the survey will scale back to an overall sample of approximately 9,000 completing three core rounds of data collection to obtain calendar year data on health care utilization and expenditures, with approximately 4,500 continuing from the previous year for each of the years. In 2002, the survey would begin the five year cycle again with increase to 13,300 households (adjusted for Round 1 splits) completing three core rounds of data collection to obtain calendar year data on health care utilization and expenditures. Coupled with data from the MCBS, this would provide the department with the analytic capabilities first proposed for the 1996 NMES-3 with respect to sample size.

3.4 Dwelling Units, Reporting Units and Other Definitions

The definitions for Dwelling Units and Group Quarters in the MEPS Household Survey are generally consistent with the definitions employed for the National Health Interview Survey. Reporting Units consist of individuals in the sampled dwelling unit that are related by blood, marriage, adoption or other family associations. College students under 24 years of age who usually live in the sampled household, but are currently living away from home and going to school, will be treated as separate Reporting Units for the purpose of data collection.

The 1996 MEPS sample consisted of households that responded to the 1995 NHIS in the two panels reserved for the MEPS, with the basic analysis unit defined as the person. Analysis is planned at both the individual and the household as units of analysis. Through the reenumeration section of the Round 1 questionnaire, the status of each individual sampled at the time of the NHIS interview is classified as "key or non-key" and "in-scope or out-of-scope". For an individual to be in-scope for person level estimates derived from the MEPS household Survey, the person needs to be a member of the civilian non-institutionalized population for some period of time in the calendar year of analytical interest. Because a person's eligibility for the survey may have changed since the NHIS interview, sampling reenumeration takes place in each subsequent reinterview for persons in all households selected into the core survey. The "keyness" and "scope" indicators, together, define the target sample to be used for person level national estimates.

Key Persons: Key survey participants are defined as all civilian non-institutionalized individuals who resided in households that responded to the nationally representative NHIS subsample reserved for the MEPS (e.g. approximately 10,800 households from the 1995 NHIS), with the exception of college students interviewed at dormitories. Members of the armed forces that are on full time active duty and reside in responding NHIS households which include other family members who are civilian non-institutionalized individuals are also to be defined as key persons, but will be considered out of scope for person level estimates derived for the survey.

All other individuals who join the NHIS reporting units that define the 1996 MEPS household sample (in Round 1 or later MEPS rounds) and did not have an opportunity for selection during the time of the NHIS interview will also be considered key persons. These include newborn babies, individuals who were in an institution or outside the country moving to the United States, and military personnel previously residing on military bases who join MEPS reporting units to live in the community.

College students under 24 years of age interviewed at dormitories in the 1995 NHIS will be considered ineligible for the 1996 MEPS sample and not included in that sample. Furthermore, any unmarried college students under 24 years of age that responded to the 1995 NHIS interview while living away at school (not in a dormitory) will be excluded from the sample if it is determined in the MEPS Round 1 interview that the person is unmarried, under 24 years of age, and a student with parents living elsewhere who resides at his/her current housing only during the school year. If, on the other hand, the person's status at the time of

the MEPS Round 1 interview is no longer that of an unmarried student under 24 years of age living away from home, then the person will be retained in the 1996 MEPS sample as a key person.

Alternatively, at the time of the MEPS Round 1 interview with NHIS sample respondents, a determination will be made if there are any related college students under 24 years of age who usually live in the sampled household, but are currently living away from home and going to school. These college students are considered key persons and will be identified and interviewed at their college address, but linked to the sampled household for family analyses. Some of these college students living away from home at the time of the Round 1 interview will have been identified as living in sampled household at the time of the 1995 NHIS interview. The remainder will be identified at the time of the MEPS Round 1 interview with the NHIS sampled households.

Non-key Persons: Persons who were not living in the original sampled dwelling unit at the time of the 1995 NHIS interview and who had a non-zero probability of selection for that survey will be considered non-key. If such persons happen to be living in sampled households (in Round 1 or later rounds) MEPS data, (e.g., utilization and income) will be collected for the period of time they are part of the sampled unit to permit family analyses. Non-key persons who leave any sample household will not be recontacted for subsequent interviews. Non-key individuals are not part of the target sample used to obtain person level national estimates.

In situations where key persons from the NHIS sampled household selected for MEPS move out (in Round 1 or later rounds) and join or create another household, data on all members of this new household who are related by blood, marriage, adoption or foster care to the persons from the NHIS sampled household will be obtained from the point in time that the NHIS sampled person joined that new household. Similarly, data will be collected (in Round 1 and later rounds) on all related persons who join NHIS sampled households selected into the MEPS.

Persons in NHIS sampled households selected in MEPS who subsequently enter an institution and leave the civilian, noninstitutionalized population of the United States will require data collection during their stay in institutions that are nursing homes. Alternatively, persons in NHIS sampled households selected in the MEPS who subsequently enter institutions that are not nursing homes and leave the civilian, noninstitutionalized population of the United States do not require any data collected in these institutions that are not nursing homes (this also applies for military service or moving out of the U.S.), but their whereabouts must be monitored during the field period. Upon their return to the U.S. civilian noninstitutional population, these persons shall once again be subject to HS data collection.

3.5 Sample Size and Yield

The 1996 MEPS sample size targets require approximately 9,000 originally sampled NHIS households yielding the complete series of core interviews (i.e., Rounds 1-3) to obtain use

and expenditure data for calendar year 1996. The expected yield at each of the stages of data collection for each new MEPS sample linked to the NHIS is: (1) a NHIS response rate of 94 percent at the household level; (2) a response rate of 86 percent (83 percent for the 1996 MEPS) among reporting units at Round 1 (conditioned on a completed NHIS interview): a round-specific response rate of 97.5 percent among reporting units at Rounds 2 and 3; a round-specific response rate of 97 percent among reporting units at Rounds 4 and 5; and a round specific response rate among reporting units of 98 percent at Round 6. The minimum acceptable response rate target for the core MEPS household survey for obtaining calendar year 1997 data on health care utilization and expenditures from the new 1997 MEPS sample is 81.75 percent conditioned on response to the NHIS (interviews for Rounds 1-3). Furthermore, the minimum acceptable response rate target for the core MEPS household survey within a PSU is 70 percent for calendar year 1997 data from the new MEPS panel, conditioned on NHIS response (interviews for Rounds 1-3), and is 65 percent for calendar years 1996 and 1997 for the 1996 MEPS panel (interviews for Rounds 1-5, conditioned on response to the NHIS).

Table 2. Expected number of responding households and associated response rate for each round of data collection of the 1996 and the 1997 MEPS Household Survey.

| | 1995 NHIS Linked Sample | Calendar Year 1996 | | Calendar Year 1997 | | Calendar Year 1998 | |
|---|-------------------------------|------------------------|-------------|-----------------------|-------------|--------------------|-------------|
| | | Round 1A | Round 2A | Round 3A | Round 4A | Round 5A | Round 6A |
| 1996 MEPS Panel | | | | | | | |
| Responding Households (by Round) | 10,800 | <u>9,500</u> 11,445 | 9,263 | 9,032 | 8,761 | 8,498 | 8,328 |
| (Response rate by Round) | (94%) | (83%) | (97.5%) | (97.5%) | (97%) | (97%) | (98%) |
| | 1996 NHIS Linked Sample | | | Calendar Year 1997 | | Calendar Year 1998 | |
| | | | | Round 1B | Round 2B | Round 3B | Round 4B |
| 1997 MEPS Panel | | | | | | | |
| Responding Households (by Round) | 5,600 | | | <u>5,057</u> 5,880 | 4,931 | 4,808 | 4,664 |
| (Response rate by Round) | (94%) | | | (86%) | (97.5%) | (97.5%) | (97%) |

The estimates of response rates in Table 1 are for the original sample of NHIS responding households, with the inclusion of splits (family member(s) that move apart from the originally sampled household) in Round 1 of the 1996 and 1997 MEPS panels. The rates specified in the table are also expected to apply to "splits" in subsequent rounds, i.e., households that will be created in the course of the survey field period as a result of key persons moving away from originally sampled NHIS households.

The sample size specifications have been set to meet precision requirements developed for the MEPS. Given the major changes in the design of the survey that were required as a consequence of the DHHS Survey Integration Plan, the sample size constraints placed on the MEPS as a consequence of restricting the sample to the 195 PSU NHIS subsample, and use of the first quarter of the 1995 NHIS sample for inclusion in a Disability Survey sponsored by the Assistant Secretary of Planning and Evaluation, DHHS, the precision requirements for the first year of the MEPS were relaxed relative to the original design specifications of the NMES-3.

For the 1996 MEPS sample, the relative standard error for a population estimate of 20 percent for the overall population at the household level was specified to be no more than 2.7 percent; and the relative standard error for a population estimate of 20 percent for the overall population at the person level was specified to be no more than 1.7 percent. For example, if it was determined that the national population estimate of the percent of the population ever uninsured in 1996 was 20 percent, the standard error of the estimate should not exceed 0.34 percent. That would translate to a 95 percent confidence interval of (19.33%, 20.67%) for the insurance coverage estimate that characterized the nation at the person level. Under the original MEPS design specifications, sample design analyses indicated that a national probability sample design that consisted of 125 unique PSUs, 2585 segments, and 14,600 households, with disproportionate sampling rates that ranged from 1.0 to 0.35 on a relative scale, would yield an average design effect of 1.7 for survey estimates. Preliminary design work suggested that a 1996 MEPS sample that was selected from a nationally representative 1995 NHIS subsample characterized by 195 PSUs, 1,372 segments and 9,000 households, with disproportionate sampling rates that ranged from 1.0 to 0.5, should yield average design effects for MEPS survey estimates in the 1.5-1.6 range.

The 1996 MEPS sample linked to the NHIS was designed to produce unbiased estimates for the four Census Regions. This NHIS linked sample reflects an over-sample of hispanics and blacks at the following ratios of representation relative to the remaining households (Hispanics 2.0:1, blacks 1.5:1). The overall expected sample yield after three Rounds of data collection at the person level is approximately 23,000 overall, with 3,500 black individuals and 4,400 Hispanic individuals. The average design effect target for survey estimates for the 1996 MEPS is 1.6. The sample design should satisfy the following precision requirements for mean estimates of the following measures of health care utilization and expenditures at the person level: (total health expenditures; utilization and expenditure estimates for inpatient hospital stays; physician visits; dental visits and prescribed medicines).

| Demographic Group | Persons at the end of Round 3 | Average relative standard error |
|------------------------|-------------------------------|---------------------------------|
| 1. Black/Non-Hispanics | 3,500 | .065 |
| 2. Hispanics | 4,400 | .055 |
| 3. Overall Population | 23,000 | .025 |

The precision requirements for the 1997 MEPS Household sample that combines the 1996

and the 1997 MEPS panels are presented in Table 3 in terms of relative standard errors for the following survey estimates:

- 1) a 20 percent population estimate at the person level for each specified domain (e.g. 20 percent of the U.S. civilian non-institutionalized population was uninsured for some time in 1996); and
- 2) mean estimates of the following measures of health care utilization and expenditures at the person level (precision requirement specified as an average relative standard error):
 - a. total health expenditures;
 - b. utilization and expenditure estimates for inpatient hospital stays;
 - c. utilization and expenditure estimates for ambulatory physician visits;
 - d. utilization and expenditure estimates for dental visits;
 - e. utilization and expenditure estimates for prescribed medicines.

The 1997 MEPS person level precision requirements are based on estimates derived from individuals that are considered full year respondents (individuals with responses for their entire period of eligibility in 1997). Consequently, in the determination of sample sizes necessary to achieve the precision requirements, additional adjustments must be made for survey nonresponse to obtain the targeted number of full year respondents.

Preliminary design work suggests that a 1997 MEPS sample that was selected from two pooled nationally representative 1995 and 1996 NHIS subsamples characterized by 195 PSUs, 2,000 segments and 13,300 households, with disproportionate sampling rates that ranged from 1.0 to 0.35, should also yield average design effects for MEPS survey estimates in the 1.5-1.6 range. Based on these initial assumptions, approximately 34,000 persons completing the three core MEPS household interviews to cover calendar year 1997 (Rounds 1-3 for the new 1997 MEPS sample; Rounds 3-5 for the carry-over 1996 MEPS Sample) will need to be selected to meet the precision specifications for population estimates that characterize the nation. Assuming 2.55 persons per sampled reporting unit, approximately 13,300 households completing the three core rounds in 1997 will be required. Table 4 indicates the desired number of persons in the various subpopulations of interest for analysis necessary to satisfy the survey precision requirements for the pooled 1996 and 1997 MEPS samples to permit 1997 population estimates.

Table 3. Precision requirements at the end of 3 core rounds for the 1997 MEPS for subpopulations of analytic interest and corresponding relative standard errors (RSE's)

| Subpopulation | RSE for 20% estimate | Average RSE for use and expenditure estimates |
|---|-------------------------|---|
| Persons with family incomes less than 125% of poverty level | .027 | .040 |
| Persons with family incomes between 125-200% of poverty level | .033 | .050 |
| Persons predicted to incur high medical expenditures | .040 | .060 |
| Persons 65 years or older | .043 | .062 |
| Adults (18+) with functional impairments (1 or more ADLs) | .058 | .085 |
| Children with limitations (age 17 or younger) | .080 | .120 |
| Overall sample population | .014 | .021 |

Table 4. Required sample yields at the end of three core data collection rounds for 1997 for subpopulations of analytic interest (assumes average design effect = 1.6).

| Subpopulation | Required sample yield | |
|---|-----------------------|--|
| Persons under 125% of poverty level | 9,150 | |
| Persons between 125-200% of poverty level | 6,100 | |
| Persons with predicted high medical expenditures | 4,000 | |
| Persons 65 years or older | 3,700 | |
| Adults (18+) with functional impairments (1 or more ADLs) | 2,000 | |
| Children with limitations (age 17 or younger) | 1,100 | |
| Overall sample population | 34,000 | |

Precision requirements for the 1997 MEPS Household Survey are stated in terms of national estimates at the person level. To meet these requirements, the survey must include a minimum number of persons in each subdomain of interest. However, the unit of interviewing and subsampling is the household. Thus, the 1996 NHIS households will need to be selected for the full panel 1997 MEPS household survey on the basis of the characteristics of the persons they include. As in the original design for the 1996 medical expenditure survey, there are six sample domains of interest to which a NHIS DU can be assigned:

1. adults (age 18+) with functional impairments;
2. children with functional limitations (under age 17);
3. individuals 18-64 years old with predicted high medical expenditures;
4. individuals with family incomes predicted to be below 200% of poverty level;
5. elderly individuals (65+); and
6. all remaining individuals).

The corresponding sampling rates for the six domains necessary to satisfy survey precision requirements will be determined by further internal research conducted to discern the expected design effects on survey estimates as a consequence of the use of the NHIS as a sampling frame. AHCPR has currently acquired the 1996 NHIS data and sample identifiers for the first 2 quarters of 1996 that have been allocated to the MEPS. By early November, the NHIS data for the third quarter of 1996 will be provided to AHCPR. At that time, it will be possible to determine the precision of survey estimates of surrogate health care measures available from the NHIS (e.g., number of doctor visits in past 12 months, number of hospitalizations in last 12 months, number of hospital days in the past 12 months), based on alternative sampling strategies. More specifically, for population subgroups that are not certainty selections, it will be possible to determine the impact on precision of concentrating the sample in fewer segments, which is more efficient from a data collection cost perspective. For the population subgroups that have been targeted as certainty selections based on current assumptions regarding average design effects, additional research will also be conducted to determine the design effects of survey estimates of surrogate health care measures available from the NHIS. The results of this investigation will inform the final sample size specifications and resultant sample selection strategy. As part of the survey design research, it will be necessary to determine the expected sample yields from the 1996 MEPS sample for these domains in order to determine the necessary sample selection rates to employ for the 1997 MEPS sample to satisfy the specified precision levels. It is important to note that all of the precision specifications for the domains specified for the 1997 MEPS co-incide with or improve upon the original precision specifications with the exception of the elderly population, which reflects a reduction in sample size as a consequence of survey integration, given this population is also represented in the Medicare Current Beneficiary Survey (MCBS).

As planned for the original 1996 National Medical Expenditure Survey, persons in families with total incomes near or below the poverty level are among the groups to be oversampled in the 1997 study. Since a reporting unit's poverty status classification in 1997 will be unknown at the time of the administration of the 1996 NHIS interview, the prediction model

described earlier will be used to determine whether a household is to be oversampled. Households with predicted probabilities above a certain threshold value above .3 will be oversampled. In addition, the prediction model described earlier to oversample individuals for the 1997 survey between the ages 18-64 who are predicted as likely to incur high medical expenditures in the subsequent year, will be used to determine whether a household is to be oversampled. Households with predicted probabilities above a certain threshold value above .4 will be oversampled. It should also be noted that for eligible 1996 NHIS households not targeted for an oversample in the 1997 MEPS, efforts will be made to retain the inherent NHIS oversample of minority populations.

3.6 Procedures for Data Collection

Preliminary Contact

The Preliminary Contact with households responding to the NHIS and subsampled as part of a MEPS panel in 1996 or 1997 has several objectives: 1) enlist in the MEPS study the household that participated in NHIS, 2) deliver record-keeping materials and instructions to the family respondent prior to the Round 1 interview in order to enhance the quality of the information collected in the first MEPS round, and 3) allow the interviewer to build rapport with those households that he/she will visit for an interview.

A "Dear Friend" letter with an enclosed MEPS Brochure will be mailed to each NHIS household subsampled for MEPS, and followed up by an interviewer call to verify the identity of the family, obtain the name of the MEPS family respondent, and update NHIS location data as appropriate (mailing address, telephone number, etc.). The MEPS brochure will introduce the study. The Assurance of Confidentiality is covered in both the letter and the brochure, and the Reporting Burden statement appears in the brochure. Households that cannot be contacted by telephone will receive a postcard (to be returned to the Home Office) with the advance letter. The postcard will request a work or relative's telephone number where the person can be reached (AHCPR, 1995).

Following the initial telephone call, and early in January 1996, a calendar and record file (the Health Events Record and the Health Events File) will be mailed to the MEPS family respondent along with \$5 as prepayment for the time devoted to record-keeping in anticipation of the Round 1 interview. The interviewer will call a second time to verify the arrival of the materials, answer any questions the respondent may have, and obtain best times for the Round 1 interview.

HS Main Rounds 1-5

Five interviews will be conducted with each NHIS panel selected for the MEPS at three-to four-month intervals over an approximately 24-month field period. The first three of these rounds (Rounds 1A-3A) define the 1996 MEPS Household survey, and will collect the main body of annual utilization and expenditure data for calendar year 1996. Rounds 3A-5A of the

1996 MEPS panel will be combined with Rounds 1B-3B of the 1997 MEPS panel to yield the sample base for the 1997 MEPS Household survey and the source of annual estimates for that calendar year. All interviews will be conducted in person with CAPI as the principal data collection mode. Round 1 will ask about the period since January 1 of the MEPS year to the date of that interview; Round 2 will ask about the time since the Round 1 interview through the date of the Round 2 interview; and Round 3 will collect data since the date of the Round 2 interview through the date of the Round 3 interview in 1997.

Questionnaires for these field rounds will parallel those used in 1987 NMES with some modifications implemented for the 1992 Feasibility Study, and with further changes indicated by the latter experience and the FAMES pretest. The instruments contain items that are asked once in the life of the study, items that are asked repeatedly in each round, and items that are updated in later rounds. Questions asked only once include basic sociodemographic characteristics. Core questions asked repeatedly include health status, health insurance coverage, employment status, days of restricted activity due to health problems, medical utilization, hospital admissions, and purchase of medicines. For each health encounter identified, data will be obtained on the nature of health conditions, the characteristics of the provider, the services provided, the associated charges, and sources and amounts of payment.

Permission forms for medical providers and for sources of employment and private health insurance coverage will be collected in the field. In addition, anyone who reports being employed but not covered by private health insurance will be asked to sign a permission form that will allow contact with the employer. A sample of medical providers will be contacted in the Medical Provider Survey (MPS) to verify and supplement information provided by the family respondent in the household interview; employers and other health insurance providers will be contacted in the Health Insurance Provider Survey (HIPS) to verify analogous insurance information and to collect other information on insurance characteristics that household respondents would not typically know.

As a consequence of a successful test in the Feasibility Study, copies of policies providing private insurance coverage to sampled persons will be collected from household respondents. These requests will be initiated in Round 1 and will be followed up in later rounds. Sampled persons will be asked to provide the policies directly or to obtain them from their health insurance provider(s). A description of the type of documents to be collected, a list of the policies identified by the respondent, and request forms to be given to providers will be given to interviewing staff to assist in this effort.

HS Main Round 6

Round 6 is concerned with obtaining valuable ancillary information before this MEPS panel is retired. It will take place after April 15, 1998 and ask for tax filing information details. Comparable information would have been collected for the 1996 panel in Round 4. Administration of the majority of Round 6 interviews will be by telephone from the interviewers' homes; in-person interviews will be conducted for those respondents without access to a suitable telephone or for those for whom telephone administration is not feasible, e.g., respondents with

hearing or comprehension problems.

4.0 MEPS Employer-Based Surveys (MEPS-IC)

The 1994 National Employer Health Insurance Survey (NEHIS) was developed to obtain national and State level estimates of the number of employers offering health insurance, their costs, the coverage and characteristics of their respective health plans. In the MEPS Health Insurance Plans Survey, detailed information related to employer provided health insurance plans is also obtained, including details of plans held by household respondents. As originally designed, there is noticeable overlap in the focus of the two surveys. The MEPS redesign integrates the analytical capabilities of these distinct surveys as part of the MEPS Insurance Component (MEPS-IC). The overall survey design of the NEHIS survey has been modified to improve upon the limitations of the 1994 survey. The revised MEPS-IC Establishment Questionnaire will serve as the core questionnaires to be administered to all MEPS sample establishments in the component surveys. The employers associated with the MEPS survey respondents will receive a supplemental streamlined HIPS-type questionnaire to obtain person-based information on employer sponsored health insurance coverage (e.g., household members specific coverage and premium).

The 1997 Integrated MEPS-IC HIPS component will consist of interviews with approximately 9,200 employers, 300 union officials, and 400 insurers, to obtain detailed information on the health insurance held by respondents to the 1996 MEPS Household Survey. The survey also collects information about other health plans available to, but not chosen by respondents.

The MEPS-IC HIPS sample design will have two stages of identification. The first stage will identify HIPS-eligible persons in the Round 1 household sample, and the second stage will identify the sources of health insurance for those persons.

Wage earners at establishments with only one location and employing only one worker are not eligible for inclusion in the HIPS sample. Establishments consisting of one self-employed person and no other employees will not be included in the HIPS sample of employers. Aside from these, persons in the Round 1 sample of the household survey who are eligible for the HIPS pretest include:

- a) policyholders, who on the date of the Round 1 interview, have health coverage through a current or former employer, a union, an insurance company or any other private health insurance source, and
- b) persons (16 or older) who are employed at a main job on the date of the Round 1 interview, whether or not the job provides health insurance

Some household persons may qualify for membership in each of the two HIPS-eligible groups (e.g., persons with coverage from a past retirement job who are working at a main job at the Round 1 interview date).

The second stage of the HIPS sample definition will identify the employers, unions, and other insurers for the first stage units (HIPS-eligible household persons). In some instances, the company or business that sponsors the plan associated with a particular job will be a union or other organization instead of the employer of the policyholder, and it may be necessary to contact both the employer and the other group. Some HIPS organizations will be contacted with respect to more than one household policyholder.

The HIPS interview will collect data about the coverage of individual policyholders. In the case of businesses and employers, the HIPS will also collect information about the characteristics of the company providing health coverage to the household person. HIPS-eligible household members will be asked to sign permission forms authorizing contact with each appropriate HIPS organization. In summary, the HIPS sample is designed as a person based sample, whereby HIPS data is to be combined with the MEPS household data to analyze individual behavior and choices made with respect to health care use and expenditures and insurance coverage.

In a complementary manner, the 1997 MEPS-IC Independent Establishment Component will consist of interviews at more than 30,000 establishments to obtain national and regional estimates of the availability of health insurance at the workplace. The analytical objective is to derive estimates of the amount, types and costs of health insurance provided to Americans by their employers. The sample design will also permit state-level estimates for the larger states. The sample of establishments will be selected from a list sample of business establishments (individuals sites) and governments. The resulting MEPS Insurance Component survey design will reflect a consolidation of the questionnaire designs, data collection efforts, imputation techniques, estimation tasks and data base designs across the MEPS-HIPS and independent establishment surveys.

5.0 MEPS Medical Provider Survey

The Medical Provider Survey in the MEPS was primarily designed to collect data for use in reducing the bias associated with national medical expenditure estimates, derived from household reported data, that was a function of item nonresponse and poor quality data. In the design of the survey, it was recognized that the household respondent was not always the best source of information on medical expenditures, particularly with the growth of managed care.

By selectively targeting those individuals and services for which charges and payments were most likely to be unknown or misreported by household respondents, medical provider data can be used in an efficient manner to improve the accuracy of national medical expenditure survey estimates. Consequently, the Medical Provider Survey was designed to obtain provider reported charge and payment data for household reported medical care events, and to serve as a data replacement strategy to reduce the level of nonresponse bias in survey estimates due to missing charge data. For individuals enrolled in managed care plans or covered by Medicaid, the Medical Provider Survey was designed to serve as the primary source of expenditure and payment information.

The Medical Provider Survey data will also be used to enhance the imputation strategy to correct for the remaining item nonresponse in expenditure data. Furthermore, to supplement the data replacement strategy in MPS, and to allow for methodological comparisons on reporting differentials between household and provider reported data at the person level, the survey included all providers that were associated with MEPS sample respondents identified in a nationally representative sub-sample of the dwelling units that completed the Round 1 household interview. This component of the Medical Provider Survey would provide a nationally representative pool of provider reported charges for all classes of medical care events identified in the household survey, to enhance the estimation and imputation strategies employed in MEPS.

The definition of a medical provider for the purposes of the Medical Provider Survey includes (a) any Medical Doctor (M.D.) or Doctor of Osteopathy (D.O.) who provides direct patient care; (b) any other medical provider (including inpatient facilities) identified in the household survey providing care under the supervision of an M.D. or D.O.; and (c) any person paid (regardless of the source of payment) to provide home health services as identified in the core questionnaire of the household survey.

5.1 Analytical Objectives

The MPS Medical Provider Survey was specifically designed to satisfy the following analytical objectives:

- o Serve as data replacement strategy for household reported events with missing expenditure information.
- o Serve as an imputation source to reduce the level of bias in survey estimates of medical expenditures due to item nonresponse and the use of household data of questionable quality.

- o Allow for an examination of the level of agreement in expenditure reporting obtained between data obtained from household respondents and medical providers.
- o Serve as the primary data source for expenditure estimates of medical care provided by separate billing doctors in the following settings: inpatient stays, emergency room visits and outpatient visits.
- o Serve as a data replacement strategy for household reported events with missing source of payment information.
- o Serve as an imputation source to reduce the level of bias in survey estimates due to item nonresponse for source of payment data and household data of questionable quality.
- o Serve as analytical database to support data adjustments to household reported medical expenditure data.
- o For sampled patient provider pairs, the MPS will also permit evaluations of the level of agreement between household and provider reported health care utilization.

5.2 Evaluation of the 1987 National Medical Expenditure Survey

Based on the 1987 NMES experience, an evaluation of the household reported data revealed that facility events were characterized by high levels of item nonresponse with respect to the reporting of expenditure data, and that individuals with public health insurance, primarily Medicaid, were unlikely to provide information on their health expenditures. More specifically, only one third of all hospital inpatient stays (33.8 percent) reported in the NMES Household Survey had expenditure data on the facility expenses for the stay that were of acceptable quality (Cohen and Carlson, 1994). For emergency room events not associated with hospital admissions, only 46.6 percent of the events reported in the NMES Household Survey had expenditure data on the facility expenses for the visit. A similar characterization was noted for outpatient department visits, with less than a third of the events (32.6 percent) reported in the NMES Household Survey having facility level expenditure data. For each of the facility specific events, the medical expenditure data under consideration reflect the facility expense for the stay or visit, not including any separate charges for physicians, but including expenditures for X-rays, lab tests, and diagnostic procedures. The charges for separate billing doctors were obtained directly from the medical providers in the 1987 NMES as a consequence of the gross levels of under-reporting in the household survey with respect to the identification of these medical providers. This design strategy that acquired the expenditure data for separate billing doctors associated with facility specific medical events directly from the medical providers was also specified as a design feature of the 1996 Medical Provider Survey.

Study findings also revealed that 63 percent of the ambulatory office-based medical provider events identified in the household survey had household reported expenditure data (Cohen and Carlson, 1994). Furthermore, an additional analysis was conducted to determine the quality of the household reported medical expenditure data, based on linked expenditure data obtained in the 1987 NMES from the Medical Provider Survey for the same health care events. Conditioned on reported data for medical expenditures from both the household and the medical provider in the 1987 NMES, a high level of agreement was observed for office based physician visits as well as the facility based events.

5.3 Exploratory MPS Sample Allocation Analysis

An exploratory sample allocation analysis was conducted, based on the initial budget specification for the survey (considering the planned NMES-3 survey design), to determine the MPS sample allocation that would minimize the variance of national estimates of total medical expenditures based data obtained from medical providers. The sample design analysis considered an optimal allocation analysis that assumed that the MPS budget as fixed, to determine the MPS sample allocation that will minimize variance in survey estimates (Cochran, 1963). The sample allocation analysis considered an expanded MPS design that would potentially allow for the inclusion of dental visits and prescribed medicine purchases in the Medical Provider Survey. Variance estimates of health care expenditure estimates for the events under consideration were derived from the 1987 NMES. This analysis was implemented to help prioritize the relative importance of specific health care events types with respect to their impact on the variance of the survey estimates of total medical expenditures.

Variable costs for this analysis were based on cost estimates from the original 1996 NMES-3 contract, and included all costs associated with data collection in addition to costs associated with coding, data preparation and data processing tasks. The following types of health care events were considered in this investigation: inpatient stays, emergency room visits, outpatient visits, office based visits, home health care, dental visits and prescribed medicines. The analysis allowed for separate billing doctors associated with selected facility events to be included in the MPS sample. Variances estimates of health care expenditure estimates for the events under consideration were derived from the NMES-2.

The sample allocation across event types to minimize the variance of the estimated population mean was based on the following relationship:

$$\frac{n_h}{n} = \frac{N_h S_h / \sqrt{c_h}}{\Sigma (N_h S_h / \sqrt{c_h})}$$

where N_h is the population estimate for the number of events of type h;

S_h is the standard deviation of the expenditure estimates for events of type h which has been inflated by the square root of the survey design effect associated with the mean estimate of expenditures; and

c_h is the variable survey cost per event of type h.

Since cost is fixed in this analysis, the overall value of n based on the optimal values of n_h is:

$$n = \frac{(C - c_o) \Sigma N_h S_h / \sqrt{c_h}}{\Sigma (N_h S_h / \sqrt{c_h})}$$

where C is the overall costs associated with the MEPS Medical Provider Survey; and

c_o is the fixed costs associated with the MEPS Medical Provider Survey.

Table 5 provides a summary of the MPS sample allocation that will minimize the variance of national expenditure estimates based on MPS data, subject to a fixed cost assumption, in addition to the percent of total health care expenditures represented by the medical event type. As can be noted in the table, other medical expenditures including those associated with medical equipment, hearing aids, eyeglasses, diabetic items, etc, were not inscope for a medical provider survey.

Table 5
MPS Sample Allocation to Minimize Variance for Fixed Cost

| Event Type | % of Total Health Care Expenditures | MPS Sample Allocation (% of subgroup) |
|--------------------------------|-------------------------------------|---------------------------------------|
| 1. Inpatient stays | 42% | 100% |
| 2. Emergency room | 2% | 27% |
| 3. Out-patient visits | 9% | 87% |
| 4. Office Based visits | 14% | 24% |
| 5. Home Health | 3% | 65% |
| 6. Dental | 8% | 31% |
| 7. Prescribed Medicines | 6% | 12% |
| 8. Separate billing physicians | 13% | based on facility sample |
| 9. Other Medical Expenditures | 3% | not applicable |
| Total | 100% | n |

Source: Agency for Health Care Policy and Research, National Medical Expenditure Survey, 1987.

Based on the results of this exploratory analysis, the sample design emphasized the inclusion of inpatient hospital events with certainty, with outpatient visits also being characterized by a high sample allocation level. Relative to inpatient stays and outpatient visits, the emergency room visits were not identified at the same high level of sample representation. In terms of survey operations, however, a sample allocation rule and data collection plan that included all hospitals associated with in-patient stays, would result in contacts with the vast majority of hospitals in which the MEPS household participants received emergency room care. Consequently, inclusion of all emergency room visits in the Medical Provider Survey under this type of data collection plan could be handled more efficiently than would be evidenced by the initial cost per case design parameters that were specified for the optimal allocation analysis.

The optimal allocation analysis also identified home health events as an event type that should be included in a Medical Provider Survey at a high rate of selection, to help reduce the variance of national survey estimates of overall medical care expenditures subject to fixed cost constraints. All remaining events, which included office based visits, dental and prescribed medicines, were targeted at relatively low levels of sample representation. Based on the low sample allocation result for dental visits and budget limitations, dental visits were ultimately not included in the MPS.

In addition to concerns regarding the variances of survey estimates obtained from the medical expenditure survey, attention was also focused on allocation strategies that would reduce potential sources of bias in survey estimates associated with item nonresponse and data of poor quality. For individuals enrolled in managed care plans or covered by Medicaid, it was recognized at the outset of the MPS design that their knowledge of the payments and expenditures for the medical care they received would be quite limited. Furthermore, it was noted that the hospital specific medical events, which consisted of inpatient stays, outpatient visits and emergency room encounters, were characterized by high levels of item nonresponse in the 1987 National Medical Expenditure Survey. The same pattern was noted for home health events.

The 1987 NMES data was used to estimate the cost of including a benefit for outpatient prescribed medicine utilization for Medicare beneficiaries, as a component of the Medicare Catastrophic Coverage Act of 1988 (P.L. 100-360). In the 1987 survey, all of the health care utilization and expenditure estimates associated with prescribed medicines were based on household reported data (Moeller, Mathiowetz and Cohen, 1989). As a consequence of noted differences in the national utilization estimates of prescribed medicines derived from the NMES data relative to alternative data sources (Moeller, 1994), and its significant relative importance as a component of total medical expenditures (6 percent of total medical expenditures in 1987, Table 5), there was a particular concern regarding the quality of household reports of prescribed medicine purchases.

Overall, 1987 NMES household participants provided expenditure information for 63 percent of the office based medical provider visits. An analysis of the quality of household reports of medical expenditures, conditioned on the availability of linked medical provider data, indicated a high level of accuracy for household based expenditure reports associated with fee for service office based visits (Cohen and Carlson, 1994). When attention was directed to the distribution of values that measured the absolute difference in reported expenditures between the two sources based on the 1987 NMES, at least 50 percent of these medical provider contacts were characterized by difference of at most \$1.00 .

5.4 Adopted MEPS Medical Provider Survey Sample Design

The MEPS Medical Provider Survey sample design that was adopted reflects an integration of the minimum variance sample allocation analysis, sources of potential nonresponse bias based on the 1987 NMES survey, and a smaller household sample specification for the 1996 Medical Expenditure Panel Survey relative to the original plans for the 1996 NMES-3. As a consequence of the overall sample size reduction in the 1996 MEPS, higher sampling rates could be specified for distinct classes of medical events than suggested by the optimal allocation analysis, while still achieving a reduction in the overall costs associated with the Medical Provider Survey.

Both the sample allocation analysis and the concerns with low levels of household reports of expenditure data provided a strong justification for the inclusion of hospital based events and home health events at relatively higher levels than the remaining eligible event types. As in the 1987 survey, the specified MPS sample design required that charges for separate billing doctors were to be obtained directly from the medical providers as a consequence of the gross levels of under-reporting anticipated in the household survey with respect to the identification of these medical providers. As noted, for individuals enrolled in managed care plans or covered by Medicaid, it was recognized their knowledge of the payments and expenditures for the medical care they received would be quite limited. Consequently, the MEPS Medical Provider Survey was designed to serve as the primary source of expenditure and payment information for these selectively targeted household respondents.

Another competing MPS sample design objective was to provide a basis for methodological analysis of household reported charges for all types of events. It was recognized at the outset of the MEPS survey that the survey costs associated with interviewing all the medical providers associated with the household respondents would be prohibitive. Consequently, the complementary design components of the specified MEPS Medical Provider Survey reflect a judicious balance between survey costs attributable to a nationally representative subsample of event types for which household respondents have historically provided expenditure data of acceptable quality, while preserving the primary design objective to correct for poor quality household reported charge data.

The adopted MPS sample is specified by provider type to help distinguish the distinct groups for purposes of data collection.

1. Hospitals. All hospitals including psychiatric hospitals, reported as the site of care for inpatient stays, outpatient department visits and emergency room encounters. The MPS

sample shall include 100 percent of hospitals identified as such by household respondents during the MEPS year.

2. Hospital physicians. All physicians identified by hospitals and/or households as providing care to sampled persons during the course of inpatient, outpatient department or emergency room care will be included in the MPS sample.
3. Office-based physicians. As of the first round of data collection in the 1996 MEPS household survey, all households will be classified according to the following hierarchy:
 1. Households with Medicaid recipients;
 2. Remaining households with HMO or managed care plans; and
 3. All remaining households.

All office based physicians reported as providers of care in household with Medicaid (or Medical Assistance) recipients will be included with certainty; as will all physicians associated with a nationally representative 75 percent sample of remaining households enrolled in an HMO or managed care plan, and a nationally representative 25 percent sample of remaining households. The subsample of households will be stratified by Census region, MSA status and race of householder.

4. Home health providers. All agency home health providers of care to sampled persons will be included in the MPS sample.
5. Pharmacies. All pharmacies that have dispensed prescribed medicines to sample persons will be included in a separate Pharmacy Component Survey.

All hospitals and home health providers are "in scope" for the MPS. Other providers and sites of care are in scope if the provider is either a doctor of medicine or osteopathy, or if the provider practices under the direction or supervision of a MD or DO. For example, physician assistants and nurse practitioners working in clinics are medical providers considered in scope for MPS. Chiropractors and dentists are out of scope (unless practicing in hospitals).

Based on sample projections from the 1987 NMES and the dispersion of the MEPS household sample, it is estimated that the MPS sample to be fielded in 1997 and linked with the 1996 MEPS Household Survey (approximately 10,000 households) consists of:

2,700 Hospitals

12,400 Office-based physicians

7,000 Hospital identified physicians

300 Home health providers

The sample will be heavily concentrated in the 195 NHIS PSUs that define the Medical Expenditure Panel Survey Household Sample.

The MPS sample fielded in 1998 and linked with the 1997 MEPS Household Survey (approximately 13,000 households) consists of:

2,800 Hospitals

15,000 Office-based physicians

8,000 Hospital identified physicians

500 Home health providers

For each year of the MPS, all providers will be screened over the phone to check their eligibility, their association with the MEPS household respondent, and to acquire information to better facilitate the conduct of the core MPS interview. Data collection methods will include phone, fax transmission and self-administration. It is expected that the majority of all interviews will be conducted by telephone (80% minimum). A small number of hospitals with the largest number of linked MEPS Household Survey participants, will be contacted in-person (AHCPR, 1995).

5.5 MPS Data Replacement Strategy to Supplement Household Reported Expenditure Estimates

As indicated, the MPS is primarily designed to provide data to help reduce the bias associated with national medical expenditure estimates derived from household reported data. The estimation strategy that has been devised to support the data replacement strategy is comprehensive in nature, making full use of MPS data to correct for missing and poor quality household reported expenditure data. In addition, it will allow for an adjustment (recalibration) of household reported data, if significant reporting differentials are observed in expenditure data between households and medical providers.

The foundation on which this estimation strategy rests is the household reported utilization experience. It is clearly recognized that household reports of medical utilization will be affected by errors of omission and commission that are a consequence of length of recall, memory loss, salience and proxy response. However, the primary focus of this estimation task will be to correct household expenditure reports associated with a household reported medical

event. At this stage in the MPS estimation strategy, no adjustments to household reported utilization patterns will be made. Separate analyses will be conducted, however, using data on linked person-provider pairs, to assess the level of divergence between household and provider reports of health care utilization.

For the purposes of this estimation strategy, which combines the household reported and provider reported expenditure data, the unit of interest is the household reported utilization. A utilization may be a visit to a specific doctor or clinic, or it may be an event involving several providers, such as a hospitalization. Once the data collection phase of the MPS survey is completed, the first stage of this estimation strategy will attempt to match all the provider reported expenditure data to the household reported utilization.

For a sample person participating in the MPS, there are three distinct outcomes with respect to matching the MPS and the Household survey data. First, the household respondent may report a utilization that matches to the data reported in the MPS. The second possibility is that a utilization is reported in the MPS, but not by the person in the household survey. The third possibility is that a person may report a utilization that does not match any utilization in the MPS. This could happen if the permission form is not signed by the household respondent, if the provider does not respond to the MPS, if there is insufficient information to match their reports, if the provider did not give a complete response, or if the household respondent erroneously reported the event.

A computerized matching algorithm developed at Statistics, Canada (1985), referred to as CANLINK is being considered as the method to use in order to match household and provider reports of medical care utilization. The matching criteria will include characteristics of the date of the utilization, the type of event (hospitalization, clinic visit, medical provider visit), and the household reported condition and provider reported diagnosis that described the purpose of the utilization. The matching rules will be developed to maximize the correct matches while minimizing the false matches and non-matches.

A. For all household and provider reported utilizations that match, and for which MPS reported expenditure data exists, the MPS data will be used as the appropriate value of the expenditure:

Y_{ij} = MPS expenditure data for matched utilization j associated with person i .

B. For the subset of household and provider reported utilizations that match and for which both household and provider reported expenditure data exist, the relationship between these alternative sources of expenditure data will be modelled to determine whether it will be necessary to implement a recalibration procedure. More specifically, let Y_{ij} be estimated as

a model based function of X_{ij} , or

$Y_{ij} = f(X_{ij})$ where

X_{ij} = HHS reported expenditure data for matched utilization j associated with person i .

The purpose of the recalibration procedure is to rescale the person-reported data so that it is comparable to the provider reported data. The improvement from recalibration is based on the assumption that the provider's responses are more accurate than the person's expenditure responses. If it is determined that there are significant differentials in the reporting patterns of medical expenditures between household respondents and their associated medical providers, the recalibration strategy should serve to reduce some of the bias in NMES national expenditure estimates associated with person-level reporting.

Based on the resultant model, all remaining household reported utilizations not included in A for which a household reported expenditure is present, X_{ij} , will be recalibrated to a predicted provider reported response

$Y_{ij} = f(X_{ij})$.

If recalibration is not supportable, all remaining households not reported in A for which a household reported expenditure is present, will be specified as

$Y_{ij} = X_{ij}$.

C. The remaining household reported utilizations not characterized in A and B for which no household reported expenditure data is present will be corrected by an imputation strategy. Additional analyses would be conducted to determine whether the imputation strategy that is implemented to adjust for missing expenditure data, regardless of the techniques employed (e.g. whether it is model based or a "hot-deck" approach), should be based (1) wholly on the MPS data, or (2) should consider the combination of replacement MPS and recalibrated household data that characterize the household respondents identified in A and B.

It should be noted that for medical care provided in managed care settings where no expenditure data is available from either the provider or the household participant, but other relevant data is obtained in MPS about the procedures that characterize the event, a valuation of the expense for the event will be implemented. The MPS questionnaire will obtain information on both the medical and financial characteristics of the applicable medical events. This will include for office visits and hospital events, diagnoses (ICD-9s and DSM-IVs); procedure and inpatient stay codes (CPT-4s and DRGs); charges or charge equivalents (where available) before any contractual

adjustments or discounts, sources and amounts of all payments made, and the reasons for any difference between charges and payments. In the absence of information on the cost of a visit in a managed care setting, these additional measures of the intensity of the services provided will be used in an imputation strategy, that will allow comparable health care events that occur in managed care settings where cost data is available, to serve as donor records.

5.6 Redesign Plan for Surveys of Health Care Institutions and Providers

DHHS currently conducts multiple provider-based surveys, including components of NCHS' National Health Care Survey (hospitals, physicians, nursing homes, ambulatory surgery, and home and hospice care), and the provider followup components of the MEPS. In instances where multiple HHS surveys approach the same class of providers, efforts will be integrated operationally so that there is a common field staff, procedures, computer-assisted survey software, and post-processing capabilities. Common core questionnaires will be identified for use in surveys that would approach the same type of provider. In addition, common classification systems, standards, procedure coding, will be adopted that would maximize efficiency as well as enhance data comparability and analytic utility.

6.0 Surveys of Nursing Homes and Related Long Term Care Institutions

Three existing surveys of nursing homes are addressed by the Survey Integration Plan: the institutional portion of the MCBS; the National Nursing Home Survey (NNHS) conducted by NCHS; and the National Nursing Home Expenditure Survey (NNHES), conducted by AHCPR and part of the NMES-3 plan. The MCBS includes an annual institutional component; the NNHS was to have been conducted in 1995 and 1997; and the NNHES is being fielded in 1996 as part of the MEPS. To complement the 1996 MEPS Household Survey, the National Nursing Home Expenditure Survey collects data from a sample of 800 nursing homes and more than 5,000 residents nationwide on the characteristics of the facilities and services offered, expenditures and sources of payment on an individual resident level, and resident characteristics, including functional limitation, cognitive impairment, age, income, and insurance coverage for calendar year 1996. The survey also collects information on the availability and use of community-based care prior to admission to nursing homes.

Under the Survey Integration Plan, these three surveys will be changed or more closely coordinated, as follows:

- 1) The NNHES will be conducted every 5 years (an initial 6 year interval from the 1996

survey to the 2002 nursing home survey to coincide with the sample peak years in the MEPS). This survey will obtain calendar-year use and expenditure estimates, facility characteristics, and resident information. This combines the analytic objectives of both the original NMES-NNHES and the NNHS, and includes a sample of 800 facilities and 3,200 residents in facilities at the start of the survey year and 2,400 first admissions over the course of the survey year. With this broad scope and depth of data collection on a sufficiently large sample, this component of the integrated design will serve as the anchor for other related data collection efforts in the long term care sector.

The data collection in the long term care sector occurring between the MEPS peaks will be done as part of the coordinated provider data collection plan and will be integrated with other aspects of the Survey Integration Plan.

- 2) Data on the capacity, staffing, and services provided by the institutions will be collected as part of the nursing home survey conducted every 5 years. In addition, as part of the Integration Plan's efforts to develop ongoing measures of the capacity of the health care and public health systems, nursing home facility measures will also be collected in years between major nursing home surveys.
- 3) This sample of institutionalized residents would be coordinated with the institutional sample selected from the MCBS sample (approximately 1,000 residents), who would be followed longitudinally across multiple years according to the MCBS data collection plan.

Survey Design Enhancements

To obtain complete annual profiles of health care expenditures at the person level, individuals sampled from the household component of the MEPS who entered long-term care facilities would be followed and their institutional use and expenditure data collected. This is consistent with the current MCBS approach. Beginning in 1998, this annual sample of institutional users selected from the MEP would be combined with the MCBS institutional sample to increase the precision of survey estimates that characterize the institutional population over levels currently attained through the MCBS. The current MEPS survey restricts coverage of the institutional population to individuals in nursing homes. The scope of the survey will eventually be enhanced to attempt to represent individuals who reside in board and care homes.

The integrated design provides the analytic capability to:

- examine the health status, medical care use and associated expenditures for nursing home residents over the course of a year, paralleling the data available for the noninstitutionalized population;

- assess the size of the Medicare-population institutionalized in personal care homes and explore the feasibility of using the Medicare beneficiary sample to identify personal care homes for estimating personal care home use by non-Medicare beneficiaries;
- examine acute care use (e.g., hospitalizations) for institutionalized individuals; and
- examine nursing home use for the non-Medicare population (a growing sector of the nursing home population) and changes in utilization by this population over time.

Through an integrated survey design, the redesigned surveys of nursing homes are expected to achieve efficiencies with respect to questionnaire design and implementation, and efficiencies with respect to post-data processing (editing, imputation, weighting, production of analytic data files)., similar to those noted for the medical provider survey integration effort.

7.0 Summary

The benefits of the redesigned National Medical Expenditure Panel Survey include significant cost savings, enhanced analytical capacities, increased opportunities for longitudinal analyses, reduction of major data gaps and major improvements in providing timely data access to the research community at large. The MEPS will provide information to help understand how the dramatic growth of managed care, changes in private health insurance, and other dynamics of today's market-driven health care delivery system have affected, and are likely to affect, the kinds, amounts, and costs of health care that Americans use. The survey will also provide necessary data for projecting who benefits from, and who bears the cost of changes to existing health policy and the creation of new policies.

The MEPS data will serve as the primary source to inform research efforts which examine how health care use and expenditures vary among different sectors of the population, such as the elderly, veterans, children, disabled persons, minorities, the poor, and the uninsured; and how the health insurance of households varies by demographic characteristics, employment status and characteristics, geographic locale, and other factors. The MEPS data will provide answers to questions about private health insurance costs and coverage, such as how employers' costs vary by region, and help evaluate the growing impact of managed care and of enrollment in different types of managed care plans.

The first MEPS data will be available on public use data tapes starting as early as spring 1997. MEPS data also will be used in a series of studies to be published by AHCPR, and by Agency and other researchers publishing in the scientific literature. As a consequence of the shift to a continuous ongoing annual survey, additional efficiencies in survey data collection, data

editing and imputation tasks will be realized, as well as further improvements in the timely release of MEPS data products to the research community.

8.0 References

- AHCPR (1995). MEPS Supporting Statement for OMB Clearance. Internal CHEIS Report.
- Cohen, S. and B. Carlson (1994). A Comparison of Household and Medical Provider Reported Expenditures in the 1987 NMES. The Journal of Official Statistics. vol 10:3-29.
- Cohen, S. (1995). Sample Design of the Medical Expenditure Panel Survey. Internal CCFS Report.
- DiGaetano, R (1994). Draft: Sample Design of the Household Component of the National Medical Expenditure Survey (NMES-3). Westat Report.
- Hunter, E., R. Arnett, N. Mathiowetz, S. Cohen, J. Madans, and J. Feldman (1995). HHS Survey Integration Plan: Background Materials.
- Judkins, D., D. Marker and J. Waksberg (1994). National Health Interview Survey: Research for the 1995 Redesign, Draft Report. Westat Report prepared for the National Center for Health Statistics.
- Mathiowetz, N. and J. Moeller (1994). Specifications for NMES High Expenditure Prediction Model. Internal CHEIS Report.
- Moeller, J, N. Mathiowetz, and S. Cohen (1989). Report to Congress: Expenses Incurred by Medicare Beneficiaries for Prescription Drugs: Appendix 5. National Center for Health Services Research and Health Care Technology Assessment.
- Moeller, J. and N. Mathiowetz (1994). Specifications for NMES Poverty Status Prediction Model. Internal CHEIS Report.
- Statistics Canada (1985). Generalized Iterative Record Linkage System. Research and General Systems, Informatic Services and Development Division. Ottawa, Ontario.

Survey Integration: Implications for NCHS

Jennifer H. Madans, Ph.D.

And

Edward L. Hunter, MA.

The National Center for Health Statistics (NCHS) is the nation's principal health statistics agency, with a primary mission to collect, disseminate, and analyze health data. NCHS, along with other Department of Health and Human Services agencies, has embarked on a major effort to improve the quality, efficiency and timeliness of data by integrating what have been discrete and isolated systems. The challenge is to develop the new systems so that analytic potential is improved and to do so within an environment of diminishing resources. The effect of the Department's commitment to integrate survey systems will be far reaching. All of NCHS's data systems will be affected in some way, as all aspects of data collection including sample design, survey content, data linkage and data editing and processing are being re-evaluated.

Integration at the sample level

A major part of the integration plan is the designation of the National Health Interview survey (NHIS) as the sampling nucleus for a number of DHHS' household surveys, including the National Medical Expenditures Panel Survey (MEPS), the National Health and Nutrition Examination Survey (NHANES) and the National Survey of Family Growth (NSFG). The NHIS collects information yearly from approximately 40,000 households and 110,000 people on health status, access to care and insurance, health services utilization, health behaviors and other topics. The sample is of sufficient size and scope to cover many sub-population groups. Using this large scale, broadly focused population survey as the "sampling frame" for other population surveys not only results in significant reductions in sampling and screening costs, but increases the breadth of data available for any given respondent: the MEPS, about which you will hear more, provides in-depth information on utilization, access, insurance and expenditures; the NHANES provides in-depth information on objective measures of health status and risk factors; and the NSFG provides in-depth information on issues related to family formation. As has been the case for the NHIS since 1957, both the MEPS and the NHANES will now be conducted continuously, thus providing essential data for monitoring changes in the health care system and the health status of the population.

Coordination and redesign of questionnaires

An essential part of the integration plan is the redesign of the NHIS questionnaire. The NHIS has been composed of a core set of data items that are repeated every year, and a set of supplements which can change each year to address current health topics. As the need for data has changed, NCHS have had to increase the amount of time allocated to the supplements. The burden - on respondents, interviewers, and NCHS staff - had become unreasonable, and is threatening data quality. The survey has been redesigned so that the data will be more useful for disease and risk factor surveillance, and will be better able to address emerging health issues. Concurrent with the redesign of the questionnaire, the mode of data collection for the NHIS will move from traditional paper and pencil approaches to use Computer

Assisted Personal Interviewing. This will significantly reduce the amount of time it takes to collect and disseminate the data, making this data system even more useful. While the plans for the redesign actually predated the development of the survey integration plan, the streamlining of the questionnaire and the change to a computed assisted mode of data collection are not only consistent with the objectives of the plan but are essential for the plan's success.

Coordination of national, state and local data collection

While the majority of NCHS's survey systems have been designed to produce national data, there is a growing need for data at the state and local level. As health care markets respond to new incentives and States gain increasing responsibility for administering health and welfare programs, high quality State level data are recognized as increasingly important to the public health and health policy community. A number of major Federal programs, such as the Childhood Immunization Initiative, are implemented by the States, which require data to target specific programs efforts, as well as by the Federal government to evaluate programs and award incentive payments. In other areas of historical Federal responsibility, States are gaining increasing flexibility for administering health and welfare programs through waivers and legislated reforms, and market reforms are further changing the nature of the health care system. While considerable health related data are available at the national level, there is a variable amount at the State level to track and monitor alternative strategies adopted by the States. NCHS is moving toward the collection of more data at the State level.

Data needed for monitoring State-level changes in the health care system include basic information on health status, access to care, health insurance coverage, and utilization of health services. In addition to basic health data, information on income and program participation is important to examine the interrelationship between health and social services programs. Basic demographic information, including employment status, is needed to interpret the impact of change on individuals and families. Given the rapidity of change and the prospect for further changes through waivers and legislation, the establishment of current baseline data at the State level is of paramount importance. Furthermore, as changes occur in the future, mechanisms are needed to estimate their impact.

A primary consideration in the design of a mechanism to track and monitor changes in the health care system at the State level is that it needs to be designed in an integrated, coordinated framework in order to maximize analytic potential, minimize cost, provide data for sub-national and national comparisons, and avoid unnecessary respondent burden. In recognition of the increasing need for State level data, DHHS is considering a new integrated survey activity to monitor the impact of changes in the health care system at the State level. CDC, working with the HHS data

Council, ASPE and collaborating agencies, is developing a national capacity to generate high quality broad-based State level data for tracking and monitoring current and emerging health related issues which is responsive to State needs for data.

The study design uses mechanisms and questionnaires from two existing national surveys, the National Immunization Survey (NIS) and the NHIS. In the NIS, interviews are conducted on a random sample of telephone households to produce vaccination coverage estimates for children 19 to 35 months for all 50 states, the District of Columbia, and 28 urban areas. The NIS Computer Assisted Telephone Interviewing (CATI) system offers a mechanism for rapid data collection to assess the impact of various changes in factors that affect and define health status. In addition, since the design for the NIS requires screening 20 households to identify a single household with an age eligible child, a potential cost-effective opportunity exists to make use of the large probability sample of telephone numbers for other emerging health care issues. Use of an abbreviated set of questions from the NHIS for the proposed integrated telephone survey will allow for standardization of the questionnaire across States and for comparisons with national data. Questions to be selected will include measures of insurance coverage, access to care, health status, and utilization of services. This will allow broad monitoring of health and health care at the State-level. Quality of the data collected by telephone can also be improved with adjustments for nontelephone households using information from the NHIS.

This proposed strategy of building on two established systems has several advantages. It uses a data collection mechanism that already exists; the questions have been developed with a wide range of input from both within and outside DHHS and have been thoroughly tested; and implementation can occur rapidly since the NIS contract includes an option for additional questionnaire items. In addition to providing State level data, the initial study will help determine whether an ongoing national capability is feasible, and if so, how it might be best achieved.

The system is being developed so that the collection of State level data serves the needs of not only the Federal government but of the States and local areas as well.

Development of the capability to conduct population-based integrated systems at the state level will be an important complement to NCHS's long standing ability to monitor birth and death data at the state and local level. The National Vital Statistics System is an excellent example of an integrated approach--where data are collected once in a standardized manner but analyzed and used at multiple levels for multiple purposes. NCHS is pleased that we have made major improvements in the most pressing problem facing this system-- the time delay in the dissemination of the data. Over the last decade, CDC/NCHS and

its partners in the states have taken significant incremental steps to improve the vital records system. The vital statistics system is now undergoing a more basic restructuring to allow it to respond to growing demands for current data. By the year 2000, birth and death certificates will be created, edited, coded, queried, and corrected at the source point in electronic form; transmitted electronically to a central location in each state for processing and management; forwarded electronically to CDC/NCHS on a frequent and regular basis; and released on a current flow basis for analysis and surveillance. Changes and updates to the coded record would be transmitted to NCHS and entered in the data file on a continual basis. These changes would shift the focus from an annual data release to a current flow release as the data are received from the states. As this system is being developed, several intermediary steps are being taken to improve the timeliness of vital statistics data reporting. Beginning with data year 1995, data will be released in two waves: a "preliminary" file which will be approximately 80-90 percent complete and a final, complete file. Data will also be released quarterly and will include 12 month moving averages. Preliminary data for 1995 were released in October, a full year earlier than final data would be released.

Integration through data linkage

An efficient and cost effective way to improve data availability is to link data from various sources. In particular, in some instances, administrative files provide data of superior quality to that which can be obtained from the respondent. For example, methodological research has found that respondents are poor reporters of their use of health care services. This information is also expensive and burdensome to collect from respondents. Two important sources of administrative data are NCHS's National Death Index (NDI) and HCFA's Medicare records. NCHS surveys obtain from survey participants consent and the information needed to link to these data bases. Such linkages are routine part of the survey design process.

NCHS is also exploring the possibility of expanded statistical matches or modeling for those instances where direct linkages is not possible. DHHS is concerned about the integration of health, social well being and human service issues, and the interrelationship of these domains; e.g. the impact of changes in eligibility for Welfare payments on access to medical care and rehabilitation for disabled children. In modeling the impacts of transformations in any of these areas, we would ideally start from databases that contain measurements over time at the micro-level for all the relevant variables. Unfortunately, such databases do not currently exist. While the Survey of Income and Program Participation (SIPP), for instance, closely follows the economic fortunes of families over time and collects comprehensive data on program participation, the data it provides in areas such as health, disability, and medical care is not sufficiently detailed for purposes of policy analysis. On the

other hand, the NHIS does not collect the detailed, in-depth data on public program participation and employment found in SIPP.

The Department's survey integration plan may eventually result in a set of detailed surveys with linkages such that exact matching will provide sufficient breadth to serve our needs. However, the availability of such a database is a way off. Our policy analytic needs have heightened urgency given the rate of institutional transformation underway. Thus, to meet the needs of the present and the immediate future, a project is currently underway to develop a linked data file based on the statistical matching of files from the National Health Interview Survey (NHIS) and the Survey on Income and Program Participation (SIPP), and to evaluate its utility for analysis, policy research and micro simulation modeling.

ASPE, NCHS and the Urban Institute, the Contractor, will be: 1) conducting a review of past attempts in DHHS and SSA to develop similar linked data bases, particularly statistical matches and concatenations conducted subsequent to the publication of the report of the National Academy of Science's Panel on the Uses of Micro simulation Modeling; 2) developing an approach for statistical matching of the data bases; 3) developing a statistically matched file of the NHIS and the SIPP, and 4) evaluating its utility for analysis, policy research and micro simulation modeling.

To develop a *statistically* matched database, we begin with two or more surveys that have some data items in common but that have other batteries of data items that differ and are complementary to each other in terms of the issues to be analyzed. *Statistical* matching involves combining the sets of complementary data items for families or individuals which resemble each other on the common set of items. The set of common variables between the NHIS and SIPP is extensive and powerful. This situation should permit sufficiently good matches to give credibility to simulations and analysis of interrelationships. In addition to providing insight into the potential impact of institutional change, this activity will inform the survey development process, making data gaps and methodological problems apparent to our survey planners.

An Integrated Approach to Data on the Health Care System

Rapid changes are occurring in health care financing and organization, including how health care providers are affiliated, how they respond to market and regulatory incentives, and who bears risk. Many of these changes have profound impacts on the delivery of, and access to, health care.

As part of process that led to the HHS Survey Integration Plan, it was clear that the health care industry, and particularly the provider/supply side, is evolving so rapidly that existing measurement tools are no longer sufficient. Rather than beginning to tinker with the design of existing data collection mechanisms, HHS concluded that a more fundamental reappraisal was needed, beginning with the development of a conceptual framework of the health care delivery system. It was clear that a new look at underlying policy questions is required, as well as a rethinking of the rationale behind both public and private data systems that address health care organization and delivery. From a more fundamental conceptual framework, we can explore new ways in which HHS and non-governmental organizations can collaborate to better characterize the provider/supply side of the health care system, and to the redesign of our data collection mechanisms.

NCHS and HHS' Office of the Assistant Secretary for Planning and Evaluation (ASPE) have initiated a long-range process to address these issues, in collaboration with our colleagues in other parts of HHS, the academic community, and at other governmental levels. The first part of this process is the identification of what data will be required in the future to address policy issues dealing with the supply side of the health system (i.e., health facilities, personnel, and other resources), the organization of these resources into health systems and plans, and the utilization and outcomes that result from the application of these resources and systems.

We have recently contracted with Mathematica Policy Research, Inc. to help us with the beginning phase of this effort. MPRI will gather information about prior and existing efforts to identify major policy questions related to health care organization and delivery, address information gaps created by the evolution of the health care industry, and identify the groups participating in efforts to evaluate these changes. HHS will use this information as the starting point for a series of workshops and seminars, in which policy makers, researchers, and data specialists can begin to develop a consensus on new approaches.

Session 8 Discussion

Papers presented by Steven Cohen (AHCPR) and Jennifer Madans (NCHS)

Steven G. Heeringa, Institute for Social Research, University of Michigan, Ann Arbor, MI 48104

The two papers presented in this session describe a major program to integrate the design, data collection and analysis activities of national survey programs that are conducted by the National Center for Health Statistics (NCHS). The integration effort represents a major rethinking of all aspects of the component survey programs. Consolidated and coordinated systems for sample development, questionnaire design, data processing and administrative record linkage are expected to yield gains in quality of the data products, efficiency of operations and timeliness of data delivery to the research community. Improved analytic potential of survey program data is expected through coordination of core survey elements and expanded linkages to administrative data sources. The integration of the major survey programs also prepares NCHS to better adapt to future changes in analytic and data reporting requirements for its survey programs. Last but not least, the coordination of sample designs and household screening activities and the elimination of redundant activities in core data collection and data processing is expected to minimize future costs of the survey programs.

In her overview paper, Jennifer Madans outlines five major steps to the integration of NCHS survey programs. The first and possibly the most recognizable step is the integration of the sample designs for the National Health Interview Survey (NHIS), the Medical Expenditure Panel Survey (MEPS), the National Health and Nutrition Examination Survey (NHANES), the National Survey of Family Growth (NSFG) and the National Household Survey of Drug Abuse (NHSDA). The large size and monthly periodicity of the NHIS make it an ideal vehicle for identifying stratified probability samples of households and individuals to be recontacted and interviewed for the MEPS, NHANES and NSFG. Throughout our careers, those of who work in the field of sampling and research design search for opportunities to share the costs of large and complex national samples across two or more survey programs. As logical as the idea may seem, such opportunities present themselves only on very rare occasions. The NCHS and other federal statisticians who have guided the design and development of the integrated program are to be commended for their insight into its possibility and their perseverance in seeing it over its many hurdles to successful implementation.

The second step in the integration of NCHS's major survey programs is the transfer of data collection from paper and pencil to computer assisted (CAPI/CATI) interviewing modes. With proper systems design and procedures, the transition to CAPI/CATI enables the survey practitioners to achieve flexibility and dependability in the questionnaire design and accuracy and timeliness in data output and data delivery to research users and dependent survey programs (e.g., transfer of detailed NHIS data to the designers of the MEPS or NSFG). The integrated survey program will also take steps to enhance the coordination of national, state and local survey and administrative data collection systems. Improvements in automated collection and transfer of vital statistics data will improve the timeliness of national data on births and deaths. Investigation of the expanded use of question modules on the NHIS combined with modest expansion of the question sequences on state-level data collection programs such as the National Immunization

Survey (NIS) should lead to better state-level estimates of immunization rates, health insurance coverage and other important topics. The larger plan for the integrated survey program also foresees expanded use of exact matches of the NCHS survey data to the wealth of health and health care expenditure data that exist in administrative systems such as the National Death Index (NDI) and the Medicare system. Investigations are also planned into the feasibility of statistically matching NCHS data to other detailed survey data sources such as the Survey of Income and Program Participation (SIPP).

My comments here will focus primarily on the challenges inherent in the integration of the sample design, specifically the dual use of NHS sample households for detailed and demanding longitudinal follow-up studies. Specific attention will be given to the combined design for the NHIS and the MEPS which is described in detail in Steve Cohen's paper.

The first major challenge to the successful integration of the sample designs is the potential for gridlock in the flow of NHIS sampling operations. The integrated sample design calls for very careful timing of sample extracts for MEPS, NSFG and NHANES. Careful coordination of staff support and systems for NHIS sample management and post-survey processing of NHIS data is essential. The demands for special samples of subpopulations will need to be carefully coordinated across the programs that will draw all or part of their samples from the NHIS. Optimal integration of the NHIS and the other survey programs will place demands on the NHIS itself in the form of added questionnaire content for supplements, two-phase sample stratification data, and information needed for nonresponse adjustment and the proposed matching to other data bases.

NCHS must use the field experience and cost data from the first years of the integrated program to evaluate the cost/error trade-offs of the two-phase approach to developing samples for MEPS, NHANES and NSFG. Several important questions that must be asked include: Do the cost savings and analytic benefits of the two-phase sample/household screening approach truly offset the costs of tracing and relocating subsampled NHIS respondents? How does the added unit nonresponse of the two-phase approach affect the survey error of the component programs? Can households that are highly mobile or move between the NHIS interview and the recontact for the MEPS or, NSFG be relocated and reinterviewed? Can movers be cost effectively included in the physical measurement studies of the NHANES?

Respondent burden on households and individuals who participate in the NHIS baseline and multiple longitudinal follow-ups may lead to larger than desired panel attrition over time or higher costs to employ counter measures to ensure that sample households are retained as panel participants. Issues of confidentiality and disclosure avoidance present another set of challenges to the integrated survey program. Wide reaching linkages and inter-survey sharing of data elements compound the task of protecting the confidentiality of the respondent or disclosure of protected data. My personal view is that the confidentiality concerns virtually preclude incorporating the NHSDA in the integrated survey program. While these issues are commonly addressed late in the survey process, the integrated survey program would clearly benefit from early planning in this area.

Special Issues for the MEPS/MPS

Steven Cohen's paper provides a detailed review of the integrated redesign of the MEPS and the associated Medical Provider Survey (MPS). Special issues related to the redesigned MEPS/MPS include the following. The MEPS begins as a stratified probability of households that completed a baseline NHIS interview. The stratification employed in this subsampling involves multivariate models (logit, multiple logit) of the propensity that a household will be low income or that household members will require costly medical treatment in the months covered by the reference periods for the MEPS sequence of longitudinal interviews. Since poverty and health status can be transitory states it will be interesting to learn just how efficient these models are at predicting the states of greatest interest to the MEPS data analysts. Will models that predict future expenditures on medical treatment be equally effective for capturing oversamples of individuals that will require future treatment for chronic and acute health conditions? On a technical note, these prediction models require special procedures to quickly impute item missing data for NHIS variables that are needed to carry out the MEPS model fitting and stratum assignments for NHIS sample households.

The MEPS utilizes an overlapping panel design. Each year's sample of households and individuals will include subsamples of observational units from both a current and the previous years' panels. The overlapping panel design will be an important asset in analyzing the characteristics of panel attrition and performing adjustments for nonresponse in the longitudinal data collection.

The proposed longitudinal design for the MEPS employs a dynamic procedure for tracing and following split-offs from the original NHIS sample households. Barring attrition due to nonresponse, the MEPS panel "following" rules guarantee that when properly weighted, the panel will retain its cross-sectional representativeness over its two-year longitudinal data collection span. My own experience with similar sample following rules in the Panel Study of Income Dynamics (PSID) suggests that the MEPS will benefit greatly from early efforts to build streamlined sample control and weight development protocols for the split-off households in the sample.

Estimation and Imputation for the MEPS/MPS Data

The provider reports of medical treatment costs collected in the MPS are extremely important in addressing the problem of estimating household medical expenditures from the MEPS. Steven Cohen's paper outlines a composite procedure for imputation and estimation of household medical expenditures that is dependent on the pattern of missing data and auxiliary provider information for the household. The procedure begins at the design stage where decisions concerning which medical events and providers to select for the MPS are based on expected rates of missing cost data and the size of expenditures for distinct classes of provider visits and medical events. Final measures of expenditures will be a mixture of actual household reports, MPS reports of expenditures associated with reported household visits and imputation of expenditures for provider visits that lack both an MEPS or MPS report of costs. Regression models based on available comparisons of household and MPS reports may also be used to calibrate MEPS-only

reports of expenditures.

Direct substitution of MPS cost data is a nonstochastic imputation procedure and as such does not contribute to the variance of the final estimates for the completed data set. Stochastic imputation of expenditure amounts in cases of complete item missing data does contribute an additional component to the total variance of estimates that are derived from the completed data. Multiple imputation (Rubin, 1987) is one recommended procedure to obtain valid inferences from the completed data set of observed, calibrated and imputed values. Alternative methods for obtaining correct inferences from imputed data are described by Rao and Shao (1992).

Researchers involved in the design of the MPS may also want to look at a paper by Raghunathan and Grizzle (1995) that examines the use of multiple imputation in combination with modularized sample designs to yield efficient estimation of multivariate relationships. This latter procedure is particularly applicable in cases where the burden or cost of collecting all data elements from each respondent is prohibitive.

Statistical Matching

Jennifer Madans' overview paper describes a current NCHS investigation into the potential for statistical matching of NHIS and SIPP data. Rodgers (1984) conducted an early investigation into statistical matching of SIPP data to other federal data bases. Successful statistical matches for bivariate pairs (X,Z) require matching on a covariate vector, Y, such that the partial correlation, $r(x,z/y) \sim 0$. This is equivalent to saying that given Y, X and Z are missing at random. It should be noted that statistical matching is a form of imputation in which the subvectors of variables, X and Z are completely missing. Therefore, simulations and analysis based on statistically matched data should reflect the imputation variance associated with the matching process.

Rubin, D.B. (1987). *Multiple Imputations for Nonresponse in Surveys*. New York: Wiley.

Rao, J.N.K., and Shao, J. (1992), "Jackknife variance estimation with survey data under hot deck imputation," *Biometrika*, Vol. 79, pp. 811-822.

Raghunathan, T.E., Grizzle, S.E. (1995), "A split questionnaire survey design," *Journal of the American Statistical Association*, Vol. 90, No. 429, pp. 54-63.

Rodgers, W.L. (1984). "An evaluation of statistical matching," *Journal of Business and Economic Statistics*, Vol. 2, pp. 91-102.

DISCUSSANT COMMENTS

COPAFS Meeting Session on *Survey Integration: Initiatives in Health Data*

William D. Kalsbeek

November 13, 1996

1. Introduction and Summary

In this session we have heard about efforts by two federal agencies to integrate the designs and operations of several health surveys. One cannot help but heartily applaud these efforts, for surely they will benefit the Nation's health care data system. However, as I consider the effects of these changes, I am convinced that they must be viewed as our first steps towards an even greater consolidation of efforts to gather health survey data. My remarks will consider what I see as some of the major advantages of the survey integration plans we have heard. Having done that, I will suggest some extensions to these plans.

A few definitions may help. First, I draw the distinction between *data items* (i.e., answers to specific questions in a survey questionnaire), and what I will call *information products* (i.e., useful things one learns from available data items). The distinction is needed if, as I believe, we should evaluate each data systems on the basis of its *information efficiency rate* (i.e., the number of information products it can yield, divided by the number of data items that comprise it).

Finally, *survey integration*, as we see it in these two papers, might be defined as a merger (at some level) of the designs of two or more surveys, in the name of reducing costs and increasing the number of information products. Integration can occur both within agency, as with the inter-relationship of various component surveys of the Medical Expenditure Panel Survey (MEPS), or between agency, as demonstrated by using the National Health Interview Survey (NHIS) household sample as the subsampling frame for the MEPS Household Survey. The hoped-for result of survey integration is a new, more cost-efficient whole that is greater somehow than the sum of its individual parts, with combined survey efforts yielding the highest possible information efficiency rate.

2. Survey Design Integration --- Its Implications

But what of the effects of survey integration? How do things change as the result of these design modifications? Several positive implications came out in the two papers:

- (1) Sample Improvements --- One is sample improvements, as seen in the use of the NHIS household sample as a screening mechanism to oversample important population subgroups (e.g., the elderly, those in poverty, etc.), although (as noted in the Cohen paper) there can be losses in the precision of estimates for non-targeted subgroups because of planned sample disproportionality due to oversampling.
- (2) More Information Products --- Another benefit is more information products, leading to the expansion of one's ability to explain important descriptive findings by linking data from multiple surveys at the individual level, as in the case of tying expenditure data from the MEPS Household Survey to insurance provisions available from the MEPS Health Insurance Plans Survey). Design linkage at the aggregate level is also useful though less powerful (e.g., for PSUs in the NHIS and NHANES).
- (3) Better Information Products --- A third advantage is better information products, resulting, for instance, from the use of correlated NHIS data to improve the quality of MEPS estimates through ratio adjustment.

- (4) Improved Operational Efficiency --- A fourth is improved operational efficiency, due to reductions in the cost of recruiting and training interviewers, resulting, for example, from the use of NHIS PSUs for other NCHS surveys (e.g., NHANES and NSFG).
- (5) Improved Control of Nonsampling Errors --- A fifth benefit is improved control of nonsampling errors, through linkages (by respondent consent) to hopefully more accurate administrative record systems, as in matching data from Medicare files with respondent data from NCHS surveys, or in the use of health provider data from the Medical Provider Survey of MEPS to impute missing data items or otherwise correct for corresponding respondent measurements obtained in the MEPS Household Survey. The utility of these linkage possibilities, however, is highly dependent on how "link-able" the two data sources are. (High matching rate may sometimes be difficult to achieve.)
- (6) Flexible Data Content --- Finally, one realizes benefit in the form of flexible data content, arising, for example, from the use of supplementary modules to the NHIS questionnaire, thus allowing greater responsiveness to changing information needs.

While each of these six advantages clearly contributes to improved health care data, other stated virtues of the presented design modifications were less clear to me. For instance, I did not see how changes in MEPS will lower respondent burden. While (as previously noted) there will be time and resource savings in collecting MEPS data, it seems that if anything, individual survey participants will be contributing more of their time as part of the newly created two-year panels.

Also, I was not convinced that MEPS sample design modifications (even with several hundred PSUs) will improve our ability to produce estimates for states and local areas, other than possibly the very largest ones. While the key to unlocking the secret to defensible small area estimates from nationally designed surveys may yet be found in some modeling strategy, I suspect that any hope of being able to mass-produce direct small area estimates will have to come through less costly state-level statistical monitoring systems similar to the National Immunization Survey described in the Madans paper. Accompanying these new state-oriented systems, however, will be the need to improve the way we conduct surveys with more modest per-respondent budgets. This must be done if we are to create credible survey data systems in a new world of higher information priorities and limited resources to gather needed data.

3. Survey Design Integration — A Ways Yet to Go?

The present configuration of health surveys generates varying amounts of the following data components: preventive behavior, knowledge, insurance coverage, access to care, health status, provider utilization, limitation in activity, expenditures, and sources of payment. Some of these surveys gather more than one component of data, and some components are collected in more than one survey. What survey integration work then remains? My view in a word is --- MORE --- more of at least the following seven things, which I believe would lead to a more process-efficient and useful health survey data system:

- (1) Data Components --- One thing we need is more data components, specifically, by recognizing *episode of illness* (i.e., a person's experiences from onset to resolution of a particular medical condition) as a unit of observation and by adding data items on *outcome* related to treatment, episode, and peoples' satisfaction with the health care process. I realize that these types of data are being gathered on an ad hoc basis by some hospitals and the managed care industry, but it is

essential that they be folded into major health surveys to fully accommodate the increasingly important evaluative side of health care delivery.

- (2) Survey Integration at the Person Level --- I believe that we also need more survey integration at the person level, meaning at least partially overlapping samples of individuals (not just aggregations of individuals, like PSUs) in the merger of survey designs. Because health is a personal experience, health surveys must capture many facets of survey respondents' lives. Only then do we have a chance to understand the dynamic of health and our increasingly complex health care system. This means that a wide variety of linked person-level data must be collected, from describing respondents and how they promote their health, to details on their experiences with the health care system.
- (3) Emphasis on Longitudinal Data Collection --- A third related need is for more emphasis on longitudinal data collection, because health is such a total life experience and because retrospective methods often fail to accurately capture it. We must rely instead on longitudinal methods of data collection, where life events are recorded closer to when they occur. Although much of a person's health experience can be gathered from secondary sources (e.g., providers' and insurers' records), longitudinal data gathering can become excessively burdensome to respondents, not to mention costly to do well. To deal with the burden issue we may need to rely on panel approaches with followup of limited duration that is long enough to span most episodes of illness (e.g., two years as in MEPS). Rotation in these panels (similar to the old 4-8-4 scheme used in the Current Population Survey) might also be considered.
- (4) Priority Setting --- More priority setting may also be needed. Accepting the relatively costly notion of gathering a broader range of data by following persons through time in an era of budget limitations implies the need to carefully weigh the utility of many features of a more fully integrated health survey data system, from the set of data items one collects to the sample sizes one hopes to achieve. When planned design features make the survey too expensive, less important features must be scaled back or eliminated altogether. In deciding which data items to retain, one must consider the information utility of the item in relationship to other retained items, as well as the importance of the information products the data items will yield in relationship to the demand for this information. As regards priorities related to sample size and precision, the resolution may be to more carefully examine the relative plausibility of less costly modes of data collection (e.g., as in the use of the telephone for immunization and health risk monitoring by NCHS and CDC).
- (5) Flexibility --- With almost daily change of the Nation's health delivery system there also exists the need for more flexibility in the health survey data system. This means that the data system must have features which allow it to change with evolving information needs. Several features can facilitate adaptability, including: (i) the use of questionnaire modules (as in the NHIS) to alter the set of data items one collects at any given time, (ii) the use of screening and disproportionate subsampling to increase the sample sizes of policy-relevant population subgroups (as in MEPS), and (iii) continual updating of provider listings (as in the NCHS health provider inventory) to maintain coverage of emergent health delivery sources.
- (6) Inter-Agency Cooperation --- A sixth necessity is for more inter-agency cooperation. The greatest potential for benefit through design integration exists in an atmosphere of greater cooperation among the agencies responsible for collecting health data. Starting with a sense of current and future health information needs, and led by a common vision for how to create the

needed data, this cooperation can evolve from existing interagency working groups, such as the one that led to the papers we have heard today.

- (7) Inter-Organization Collaboration --- And finally, greater collaboration among those who design and collect survey data may be needed. Given its likely size and scope, a fully integrated set of health surveys would probably require technical and organizational skills that exceed the capacity of a single data collection operation. This would imply the need for the integrated surveys to be conducted by some combination of commercial survey organizations, academic institutions, and the Census Bureau. Would this kind of organizational collaboration work? I think so, given the growing number of consortia and cooperative agreements that have successfully developed large data systems.

In closing, I would once again express my thanks to the two authors for their inspiring efforts, and add my sincere hope that the integration we have heard about today is the beginning of a broader and continuing union of health-related population surveys by all of those who produce these data. The future of the Nation's health and health care system is at stake.

Session 9

SHARING DATA FOR STATISTICAL PURPOSES

REALIZING THE PROMISE OF DATA SHARING

THOMAS L. MESENBURG
ASSISTANT DIRECTOR FOR ECONOMIC PROGRAMS
BUREAU OF THE CENSUS
WASHINGTON, DC

COUNCIL OF PROFESSIONAL ASSOCIATIONS
ON FEDERAL STATISTICS
BETHESDA, MARYLAND
NOVEMBER 13, 1996

REALIZING THE PROMISE OF DATA SHARING

I have three purposes today -- to (1) outline my perspectives on new data sharing authority, (2) engage you in the challenges of implementing new authority, and (3) suggest an available enhancement opportunity.

I. PERSPECTIVES ON DATA SHARING

Both the promise of new Federal data sharing authority, and the barriers to obtaining it, are real and worth some reflection.

Data Sharing Has Promise

The promise of data sharing is to improve official statistics, and strengthen supporting agency operations, without having to change our decentralized statistical "system."

- It offers a "win-win" opportunity for users of statistics, providers of business and household information, and participating Federal agencies.
- It won't eliminate Federal debt, or assure world order, but it does present an exceptionally favorable balance of likely benefits and modest risks.

Improved Statistics. Improving official statistics is a goal that has widespread and long-standing support. Data sharing would help achieve this goal for many economic, and demographic data series by:

- Making single reports more complete or reliable where issuing agencies can use similar or related content information shared by other agencies.
- Making related reports more comparable where separate issuing agencies can receive shared technical detail and applied methods.
- Encouraging wider interagency use of "best series" content and "best methods" statistical practices.

Strengthened Operations. Strengthening statistical operations across agencies is the means by which data completeness, reliability and comparability can be increased. It is also the means by which additional benefits can be realized. For example:

- Completeness of Census information on foreign-ownership of U.S. companies would be increased if it could include confidential BEA data.
- Reliability of BEA national income and product estimates would be higher if BEA's sources could include confidential Census and BLS data.
- Comparability of Census and BLS industry statistics would be greater if they could use the same 4-digit SIC classifications for each establishment.
- Statistical agency costs would be lower if separate data collections could be avoided, reduced in size, coordinated, or streamlined in process.

Despite the real promise of increased statistical data sharing, it cannot begin and will not be possible unless significant barriers are recognized and overcome.

- Data sharing authority is not on "cruise control;" enacting it will require navigation and direction by the interested parties!

Barriers To New Authority

The barriers to new Federal statistical agency data sharing authority are substantial, and include both legal considerations and practical concerns.

Legal Considerations. Legally, data sharing cannot greatly expand without changes in existing law. And to be fully effective, data sharing amendments must be adequate in scope and tailored to statistical agency differences.

Amendments proposed in 1996 would be adequate to remove existing barriers. Taken together, H.R. 3924 (introduced by Cong. Horn) and Internal Revenue Code amendments (forwarded by Treasury Assistant Secretary Samuels) would:

- Authorize eight major statistical agencies to receive shared data.
- Allow sharing only by written agreement of all participating agencies.
- Assure effective actual use of shared data and full data confidentiality.
- Provide conforming amendments for six statistical agencies.

It is critical that data sharing amendments allow tax information to be shared with the designated statistical agencies. Without including the Treasury-proposed Internal Revenue Code amendments, data sharing's promise cannot be realized. To illustrate:

- Since the early 1950's, the Census Bureau has received and used extensive tax information from the Internal Revenue Service (IRS).
- We use it primarily to help maintain a complete and current business register, and augment the coverage of five-year census statistics.
- The 1996 register includes tax information for some 5 million smaller businesses that is essential for efficient and reliable survey samples.
- For other data sharing agencies, use of this complete register is indispensable to permit smaller survey samples, reduced burden, and lower costs.

Practical Concerns. We expect the 1996 data sharing and companion bills to be repropounded in 1997. However, I don't need to remind you that proposing and enacting legislation can be two different things. Timely action on data sharing amendments will require :

- Congressional leadership, and broad committee support.
- Interagency cooperation on action strategies and plans
- Demonstration of significant public and program benefits.
- Sensitivity to individual and information confidentiality concerns.

One likely legislative issue will be -- what should be the priority for different kinds of statistical changes? That is, can organizational and data sharing changes both be considered; if not, which should come first, and why? I:

- Assume there may be diverse views around the room on this issue.
- See data-sharing as clearly worth pursuing, with or without other changes.
- Urge that agencies and the Congress make it an early action priority.

A different kind of practical concern is our deep-seated tradition of decentralized statistical operations. It has strengths that we are familiar with and value. However, one of them is not natural and sustained cooperation on specific initiatives.

Obtaining and using adequate data sharing legislation is an initiative that will require our highly decentralized statistical system to get together and stay together until the “deed is done.”

- It will involve eight major agencies plus the IRS, several Congressional committees, and numerous data users and policy participants.
- It means that to enact legislation, and achieve “hard” benefits, affected agencies must institute and sustain new levels of specific cooperation.
- Instituting and sustaining the needed levels of interagency cooperation will be easier said than done, and require arrangements yet to be developed.

But let’s move on. Let’s assume adequate data sharing authority becomes law, and consider some of the implementation challenges we would face.

II. CHALLENGES TO SHARING DATA

Data sharing legislation would not be self-executing. It would “authorize” but not “require” implementing actions. Affected agencies would include the Office of Management and Budget (OMB), eight designated statistical agencies, the Internal Revenue Service, and perhaps other agencies. What actions might be needed, what might those actions require, and what risks might they present?

What should be the priorities for implementation?

As proposed in 1996, data sharing authority would have broad statistical program coverage. The eight statistical data centers are responsible for most general purpose statistics; potentially any Federal executive agency could participate in a data sharing project; and data sharing will require some resource commitments.

Whether by choice or by consequence, priorities will become part of the data sharing implementation process. I would prefer that we prioritize by choice, and seek projects that (a) have the biggest impact in terms of our statutory purposes, and (b) present the lowest risks in terms of threats to data security and confidentiality. For example, we might initially emphasize:

- Projects that seek to greatly improve sample frames, and can reduce burden and costs without compromising information quality.
- Data collections that involve business respondents, who seem to be less concerned than individuals about data confidentiality and security risks.

Should “third parties” have access to shared data?

The 1996 proposal required data centers to use shared data “exclusively for statistical purposes,” but did not require them to limit data access to their own employees. Centers could allow access by “agents” who work under the agency’s supervision, and accept the legal conditions under which data were shared.

But statistical agency operating practices vary widely; “agents” can hold quite different degrees of discretion; and our “information age” demands that agencies be sensitive to real and perceived concerns about individual information security. Specific issues regarding “third party” access may include:

- Whether use of an agent or contractor is new; should long-standing or existing uses be different from proposed new uses?
- Whether the basis for accountability is important; should agency-sworn individuals be different than contractor employees?
- Whether the functions purchased are important; should hired central computer services be different than local field interviewers?
- Whether other contractor activities are important; should small or specialized firms be different from major information companies?

Because data sharing will be new, carefully scrutinized, and easily misinterpreted, I would urge initial caution in allowing third parties access to shared data. For example, I would think that third party access to confidential data files:

- For production purposes as part of an ongoing contracted-out data processing service would be warranted, as clearly cost-effective.
- For sample selection purposes by a national household survey company would be unwise, as subject to real and perceived misuse.

Should historical data files be shared?

Substantial benefits can come from sharing data that have been collected and are now held by Federal statistical agencies. For example, Census Bureau business register information could help social program agencies that survey for-profit service providers, such as hospitals, medical clinics, and professional schools. However, virtually no existing data were collected with notice to respondents that their information might be shared among Federal agencies.

- At best, this presents a fairness issue to be resolved by data sharing participants. If historical data files are to be shared, how can this be done without unacceptably violated past assurances of limited use?
- At worst, this presents a potential legal barrier to sharing historical data. If data were collected for use by one agency, how can it be provided for the use of another agency?

The 1996 data sharing bill proposed a public notice-and-comment procedure to identify and resolve concerns regarding the use of historical data files. I think this offers a useful way to begin to address specific concerns and hopefully to avoid major objections. However:

- Not everyone reads the *Federal Register* from cover to cover each day, and a public notice that yields no objections should not be taken as sufficient.
- As data sharing arrangements become more widely used and known, newly interested parties can be expected to raise new policy and legal concerns.

In addition to a public notice process, I would urge that each data sharing arrangement include explicit (a) assessments of anticipated sharing benefits and fairness risks, and (b) determinations that the benefits clearly outweigh the risks.

What notices should future respondents receive?

Virtually all existing notices to census and survey respondents provide assurance that individual responses will be used only for statistical purposes; most provide this assurance only on behalf of the collecting agency; and none that I know of suggest that responses may be shared with another Federal agency. Future data collections that may be subject to data sharing will likely require changes in these existing respondent notices.

The issues we face in this area involve to what extent, and how notices should be changed. Here, I urge that we be inclusive, appropriate, and reasonably consistent.

- Inclusive in that a "sharing notice" should be included wherever future sharing might result (i.e., in virtually all substantive collections).
- Appropriate in that the notices should briefly outline the prospect of statistical sharing, without detailing expected recipients or arrangements.
- Consistent in that notices used for separate data collections should be identical if possible, or quite similar if variations are needed.

The reasons for consistency are that (a) the intended substantive message should be the same for all collections, (b) individual business and household respondents may receive notices from different programs and agencies, and (c) we should avoid creating real or suggesting false differences in sharing prospects.

III. ENHANCEMENT OPPORTUNITY -- NAICS

Let me conclude by commenting on data sharing and the implementation of NAICS -- the North American Industrial Classification System. Convergence of these two initiatives offers an important and immediate opportunity to enhance data sharing benefits.

- Data sharing will encourage more open and complete statistical use of information obtained and used by Federal agencies.
- NAICS will establish a greatly updated and more consistent system of industry classification for first-ever use throughout Canada, Mexico, and the U.S.

NAICS is scheduled for implementation beginning in 1998. NAICS will fundamentally redesign industry classifications for business establishments, and implementing NAICS will require new information to assure accurate reclassifications. For this purpose:

- In December 1997, the Census Bureau will request reclassification information from 5 million business locations as part of the five-year economic census.
- In fiscal years 1997 and 98, the Bureau of Labor Statistics will request similar information from the 3.5 million business locations most affected by NAICS.

If revised data sharing authority were available by mid-1997, (a) both of these overlapping data collections would not be necessary, and (b) a reduced collection effort could obtain better and more up-to-date business information. That is:

- The Census Bureau's collection of reclassification information could proceed for larger businesses operating in the census year.
- BLS' subsequent collections could be reduced in scope and focused on smaller businesses not covered in the census, and businesses formed after the census.
- Both agencies could receive and use the considerably expanded reclassifications resulting from the redesigned collection effort.

I noted that "revised" authority would be needed because the provisions of H.R. 3924 would need to be enhanced to permit this change. This is because under the 1996 proposal, data could be shared only for "exclusively statistical" uses, and BLS' information program uses are not exclusively statistical.

- BLS' industry reclassifications are intended for use by cooperating agencies in each state, and their uses are inherently regulatory as well as statistical.
- This limitation would prevent Census Bureau reclassifications obtained during the 1997 economic census from being made available for full use by the BLS.

Statistics Canada has followed a practice for 20 years that could resolve this problem. It is to legislate that industry classifications assigned to individual establishments by the statistical agency are not confidential, they are available as public information.

- This policy has two primary rationales -- classifications are (a) derived by statistical agencies, and (b) widely available in commercial business directories.
- Statistics Canada's practice is well-accepted, it has caused no public controversy, and it has facilitated both public statistics and private economic opportunities.

A provision that declares statistically derived information, such as industry and geographic classifications, as not confidential could be a very useful enhancement to 1997-proposed data sharing legislation.

In conclusion, let me confirm that I believe data sharing promises clear benefits for information providers, data users, and participating agencies. However, this promise will not be realized by relying on abstract policies, outside events, or the efforts of others. Realizing the promise of data sharing will require the "real time" attention, action, and cooperation of most of us in this room. And it will require an agenda that includes adequate legal authority, useful sharing projects, and new program relationships. Let us be aggressive in insuring that data sharing becomes a reality in 1998.

Data Sharing- A USDA/NASS Perspective

Douglas G. Kleweno
National Agricultural Statistics Service (NASS)

Council of Professional Associations on Federal Statistics
Bethesda, Maryland
November 13, 1996

Data Sharing- A USDA/NASS Perspective

I. Introduction

The principle mission of the National Agricultural Statistics Service (NASS) is "to serve the United States, its agriculture, and its rural communities by providing meaningful, accurate, and objective statistical information and services." This charge was established with the founding of the agency in 1863 within the Department of Agriculture (USDA) and continues to be the focus as described in NASS'S Strategic Plan for 1996 [1]. The task includes official USDA forecasts and estimates prepared by the NASS Agricultural Statistics Board (ASB). NASS conducts reimbursable survey work for other Federal, State, and producer organizations and provides technical assistance for other countries developing agricultural statistics programs. Most recently, the 1996 Congress appropriated funds for NASS to conduct the 1997 Census of Agriculture.

The statistical data series on U.S. agriculture are the result of data collection survey activities requiring cooperation from thousands of respondents each year surveyed on production and marketing decisions. Field observations and administrative data compliment this effort in the annual publication of approximately 350 reports prepared by the ASB. Additional survey data are compiled on agricultural resources and their management as related to economics, the environment, and farming practices.

NASS State Statistical Offices (SSO's) in 45 locations provide the infrastructure to efficiently service all 50 States with statistical activities from data gathering to data sharing. Each office operates under a cooperative funding arrangement with State Departments of Agriculture and/or land-grant universities. These arrangements facilitate collaboration of State and Federal agricultural survey data management resulting in efficiencies which reduce duplication of functions and increase services to data users. The challenge to continue improving the cooperative effort among State and Federal Agencies as it relates to this data process is vital to the role of a statistical agency.

As a federal statistical organization, NASS is charged with the dissemination of reliable, timely, and relevant information to data users. The agricultural data series are used for production and marketing decisions by farmers, ranchers, and other agribusinesses. These statistics also are critical to the agricultural industry and policy makers in responding to the dynamics of the sector and the Nations' economy, environment, and natural resources. The importance of data sharing must remain a priority; however, the preserving of data confidentiality cannot be compromised.

The following sections describe NASS's data sharing policy and

management, define current data sharing between NASS and the Census, illustrate other NASS external data sharing arrangements, and discuss future possibilities and challenges of data sharing by NASS.

II. NASS Data Sharing Policy and Its Management

Authority

As a Federal statistical organization, NASS has the authority and public financial support to collect and publish official U.S. agricultural statistics. This authority carries a major responsibility of maintaining data confidentiality. The authority is established by the United States Code, Title 18, Section 1905 and Title 7, Section 2276. Key provisions of this law that protect the confidentiality of reported data follow:

- Neither the Secretary of Agriculture, or any employee of USDA, or any other person may disclose such information to the public, unless such information has been transformed into a statistical or aggregate form that does not allow the identification of the person who supplied particular information.
- Such information shall be immune from mandatory disclosure of any type, including legal process without consent of such person.
- Any person who shall publish or publicly release such information collected shall be fined up to \$10,000 or imprisoned for not more than one year, or both.

A confidentiality pledge is required on all NASS questionnaires requiring data from a respondent and is a part of the introductory statement read to telephone respondents. Any request for an exception must be submitted in writing to the NASS Associate Administrator who is also referred to as the ASB Chairperson.

Other legislation also defines data sharing procedures as established for Federal statistical organizations including NASS. The Privacy Act of 1974 was enacted by Congress to protect individuals' interest in controlling identifiable information maintained by federal agencies. The act covers record systems which contain personal information and from which agencies retrieve information by personal identifiers such as name or social security number. Only a limited number of NASS systems of records qualify and then just a few items on those survey instruments are applicable to these provisions. Most NASS surveys request items of information on the farm business organization which are exempt from this act. Agency guidelines on the Privacy Act are documented in a Policy and Standards Memorandum Number 5-95 [2]. The Privacy Act does not halt research that can be accomplished with nonidentifiable data or dissemination of anonymous information.

The 1966 Freedom of Information Act regulates the disclosure for research and statistical records and applies to data for individuals and organizations. Public access is permitted to records from federal agencies except for specific exemptions. The exemption most frequently employed by NASS is disallowing access to identifiable records which would disclose sensitive information about individual units.

Management Plan

Data sharing is a significant and integral part of NASS's mission as a Federal statistical agency. The importance of this function is clearly stated in the 1996 NASS Strategic Plan which outlines specific goals relevant to the data sharing issue. Goal 1 states that NASS will be recognized as the Federal agency prepared to provide all essential statistics related to U.S. agriculture and the mission of the USDA. Goal 2 states NASS will be the first choice for customers seeking value and excellence in new agricultural and rural statistical services. Goal 3 states NASS will preserve its integrity as an impartial and highly credible source of statistics and as a protector of confidentiality for individual data sources. Associated with each goal are performance targets, strategies, and measures used to evaluate progress as required by the 1993 Government Performance and Results Act.

All NASS data sharing requests are reviewed and either approved or denied by the Associate Administrator. NASS organizational structure also includes a Standards Officer staff position whose responsibilities include coordination of Agency data sharing policy and other standards documented in NASS Policy and Standard Memoranda (PSM). These PSM's establish Agency program policies, standards, and guidelines relating to agricultural statistics and address in detail the topics of confidentiality of information, Privacy Act of 1974, access to lists and individual reports, release of unpublished summary data and estimates, and standards for suppressing data due to confidentiality.

Requests for and sharing of NASS data occurs in various ways. First, the most direct method of data sharing is official NASS data published by the Agricultural Statistics Board. Annually, estimates for about 120 crops and 45 livestock items in addition to environmental and economic data are published. These data must adhere to the NASS standard for suppressing data due to confidentiality [3]. This policy applies to all estimates and summary data published in either releases or research reports. To avoid disclosure of individual operations for a given item of interest, summary data and estimates must NOT be published or released if either: (1) the nonzero value for the item of interest is based on information from fewer than three respondents; or (2) the data for one respondent represents more than 60 percent of the value to be published. Any exception to this rule is granted only with written and signed permission of the respondent concerned.

Suppressed data may be aggregated to a higher level for publication. Care is required, however, to ensure that the suppressed data cannot be reconstructed within the published document or when the same data are republished at various time intervals.

USDA Regulation 1042-42, issued January 16, 1996 defines the functions and authority of the NASS Agricultural Statistics Board. Authority given the ASB Chairperson stipulates that each employee, before being given access to confidential reports and unpublished records, is required to read Departmental Regulation 1042-42; U.S. Code Title 18, Section 1905; and U.S. Code Title 7, Section 2276. A recertification of all employees occurs every 3 years. Individuals other than NASS employees must request access to unpublished data through the ASB Chairperson.

Second, unpublished NASS data sharing follows rigid guidelines outlined in an agency PSM titled, Release of Unpublished Summary Data and Estimates [4]. Unpublished summary data and estimates are made available only for appropriate research and statistical purposes which are considered beneficial to USDA, NASS, and public interest and not detrimental to the survey respondents. These statistics cannot be released while the estimates to which they apply are still current. Requests for release must be submitted on a specific form. Release of survey indications to external data users is permitted only when approved by the Associate Administrator. NASS research or staff reports meeting the above criteria may publish survey indications to support research conclusions and recommendations with the approval of the Associate Administrator.

When there is a perceived general demand and recurring requests are likely for certain unpublished information, the Associate Administrator will likely recommend publishing the summary statistics, given they are statistically reliable and confidentiality rules are maintained. This action ensures all interested data users are equitably served.

Third, the NASS policy in response to data requests for access to NASS lists and individual reports by external data users is described in PSM 6-90 [5] as follows. "List sampling frame names, area sampling frame names, and individual reports will not be made available outside of NASS if it is determined to be detrimental to NASS; the Department of Agriculture; or the farmers, ranchers, and agribusinesses included in the sampling frames. Such information can only be made available for approved statistical activities. In all activities, the identity of respondents or data included in individual responses must not be made available to unauthorized persons. All requests based on Freedom of Information Act or other legal aspects must be referred to the Associate Administrator."

Data Sharing Rules for NASS List Sampling Frames:

NASS list sampling frame rules ensure that respondents' status as

an agricultural enterprise and their reported survey data are kept confidential. The list sampling frames can only be used by external data users outside the Agency for limited uses following specific procedures. Notable examples follow:

1. The Bureau of the Census has historically had access to lists of farm operators on NASS list and area sampling frames for list building and coverage evaluation activities associated with the Census of Agriculture. The U.S. Code, Title 13, provided confidentiality protection.
2. General Accounting Office and Office of Inspector General auditors in program evaluation efforts have had access to reports from samples of NASS list frame names under tightly controlled arrangements. The requesting organization must provide sufficient information to allow NASS to evaluate how the sampled data will be used, and be assured that the review is of a statistical nature with no punitive or enforcement purposes directed at individual respondents. Reported data are released only with the written approval of selected respondents. Auditors must be certified with NASS confidentiality forms and any release of sampled data must be approved by the Associate Administrator.
3. NASS will conduct surveys for local cooperators using the NASS list sampling frame if a formal cooperative agreement or memoranda of understanding exists. Respondent burden and benefits to the total Federal/State statistical program are evaluated.
4. Other governmental officials are asked to review list frame names and addresses to assist NASS list building and maintenance efforts. Individual identifiers such as social security numbers and employer identification numbers are removed from the document before the review begins.
5. NASS will make mailings for State and federal government agencies of unbiased informational items or referendum ballots to portions of the list sampling frame if the purpose contributes significantly to the agricultural sector.
6. Other data user requests for NASS lists not established by the above procedures are addressed with a justification statement to the Associate Administrator.

Data Sharing Rules for NASS Individual Reports:

Information collected from individual or organizations is protected from disclosure by U.S. Code Title 18 and Title 7. However,

restricted on-site analytical use of microdata is permitted when certain procedures are followed. The site location is restricted to NASS offices or Economic Research Service offices with specific arrangements to ensure data security. Non-NASS users are limited to academia, private nonprofit organizations, and governmental analysts conducting research that serves the general public and contributes to the understanding of agriculture or statistical procedures used by NASS. These analysts must complete a data request form, subject to the approval of the Associate Administrator, which documents the data requirements, project schedule, statistical methodology, summary output, intended use and distribution of the report, and nondisclosure assurances.

Analysts must become familiar with concepts underlying the NASS data to ensure appropriate statistical interpretation and use of the data. Knowledge and understanding of the survey design, sampling procedures, response rates, data editing and imputation, outlier adjustments, and other factors impacting summarization and data usage are important. Most Agency data are obtained using complex sample survey designs and the Agricultural Statistics Board ensures that defensible statistical standards are met for all published data. Because sample surveys are designed to produce estimates for the NASS statistics program, it is important that the Board evaluate and provide guidance to ensure non-NASS analysts also adhere to appropriate statistical methods.

All direct identifiers such as name, address, telephone number, and social security numbers are deleted from the individual identifier data files before access is granted.

Another data sharing alternative used by NASS is special tabulations and data aggregations. The non-NASS analysts specifications are used to create the desired summary instead of direct access being granted to individual identifier NASS data files. The agency is reimbursed for expenses incurred.

NASS also conducts surveys and shares data with other federal, State, and producer organizations using cooperative agreements, memoranda of understanding, and contracts. NASS has cooperative arrangements with State departments of agriculture and/or land grant universities in each field office which service all 50 States. These reimbursable surveys offer the same protection to respondents as other NASS surveys with the same U.S. Code Title 18 and 7 rules for data access. The national level rules apply to the cooperative statistical programs in each State Statistical Office. All NASS employees and sworn agents are fully responsible for the various statutory confidentiality provisions and subject to the penalties for disclosure. Another benefit of this cooperative agreement to the public is the amortization of the cost of frame development and maintenance across more sample surveys.

Two methods of data sharing are acceptable to NASS for statistical research projects. First, special research can be conducted through the Statistical Survey Institute Program. Fellowships in

this program are solicited and supported by the American Statistical Association (ASA) and NASS. Fellows become agents of NASS and usually take a one year sabbatical. Research is conducted on-site at NASS with access to appropriate microdata files. For example, Bargmann [6] conducted research to develop a technique to generate artificial data sets with similar statistical characteristics for annually collected economic data.

Second, research agreements can be made for individual data analysis. A research proposal must be submitted for approval by the NASS Associate Administrator documenting individual data access requirements, research to be conducted, uses and benefit of the research, and data security procedures to be employed. The researcher(s) must read and sign the statement which documents the rules of confidentiality and disclosure compliance. Once the document is completed and approved, the cooperating institution researchers can have access to individual data at their site. To protect respondent identity, it may be necessary for NASS to use data masking techniques or remove unusual observations from the file. Microdata sharing is permitted outside NASS office facilities under these conditions but only if other alternatives such as the use of summary data or access within NASS State offices are not feasible for the study being conducted.

III. Current Data Sharing Between NASS and Census

The Agriculture Division in the Bureau of the Census and the National Agricultural Statistics Service in USDA have a long history of cooperation. NASS agricultural data collection began in 1842 as part of the U.S. Patent Office and the Census of Agriculture began in 1840 as part of the decennial Census. NASS collects and reports the current happenings in agriculture, i.e., acres planted, yields, effect of weather on crop production, livestock production, and prices, while the Census provides very detailed demographic and agricultural information by county every 5 years. Up through the 1920 decennial census, the agricultural information was collected with the population census. For the first quinquennial Census of Agriculture taken in 1925, NASS handled the field enumeration through its State offices. It is obvious that the Agencies managed to share data at that time.

However, later confidentiality legislation prohibited the sharing of data and a serious problem arose in 1948 between the two agencies due to the inability to share data collected from cotton gins. The Census was required by law to report the number of bales ginned by certain dates during the harvest season, and NASS was responsible for forecasting total cotton production each month and relied on cotton ginners as an excellent source of information to help forecast production. The Cotton Ginners Association successfully lobbied Congress to eliminate the duplicate monthly surveys and the Appropriations Committee threatened to reduce each agency's budget unless some means could be worked out to share data. Subsequently, an agreement was reached in which the Census agreed to obtain a "waiver" from each gin that provided them

authority to share the data with NASS.

The Reports Act of 1942 prohibits duplication of data collection by the Federal Government. However, over 50 years after its enactment, considerable duplication continues because NASS and the Census are unable to share data. Title 13 of the U.S. Code allows only Census Bureau employees (including special sworn employees) access to individually identifiable information. Also, the purpose of use must be within the scope of the Bureau's statutorily defined mission.

The National Agricultural Statistics Service has shared its individual farm list with the Census Bureau in preparation for each census of agriculture. NASS SSO staff, as sworn Census agents, also have participated in data edit processes at the Ag Census offices. Title 13 prohibits any reciprocal flow of information to NASS for use in its development and maintenance of a list of farm operations. The provisions of the Tax Reform Act of 1976 (P.L. 94-455) do not permit the Internal Revenue Service and Census, who uses farm tax return lists as a primary name source, to share the same lists with NASS. The result has been increased costs to NASS to construct and maintain its list sampling frame.

Funding for the 1997 Census of Agriculture was shifted by Congress from the Bureau of the Census to NASS starting with fiscal year 1997. However, the proposed legislation to transfer full authority of the Ag Census to NASS has not been approved yet. Because of that outcome, the Census Bureau has not permitted NASS offices to have access to previously reported data and names and addresses from the 1992 Ag Census. If NASS SSO's could access this data, it might be possible for those offices to identify further duplication in the mailing list and reduce expenses for the 1997 Census.

NASS also has not been permitted any access to the Internal Revenue Service (IRS) farm tax returns which will be used as input to the 1997 Census mailing list development. NASS hopes to work with the IRS to obtain access based on the shift of the Ag Census responsibility to NASS and the NASS confidentiality procedures. Legislation will be resubmitted to obtain the full Ag Census authority.

Proposed action for efficiencies and for improving the quality of agricultural and economic statistics is not new. For example, former Chairman Michael Boskin of the Council of Economic Advisers [7] recommended legislation to allow limited sharing of confidential statistical information for statistical purposes only among statistical agencies with rigorous safeguards. The Data Sharing Task Force, convened by the Office of Management and Budget (OMB) to assist in the formulation of legislation by OMB, stated in their December 1994 draft report [8], "...we believe that data sharing presents an opportunity for tangible benefits in our decentralized statistical system, and 1995 is the time to act on this opportunity." In the past year, however, the effort to pass interagency data sharing legislation was unsuccessful. Even with

projected data and cost efficiencies, less respondent burden, and data quality improvements, congressional approval will be difficult. The concern about loss of privacy with increased data sharing, regardless of any legislated protection, is a growing public concern.

IV. Other NASS External Data Sharing Arrangements

The NASS data sharing policies and their management, discussed earlier, establish a process to ensure respondent confidentiality while providing data access for legitimate statistical purposes. Described below are cases of data sharing which illustrate these agency procedures in use. The examples are not intended to be an exhaustive list but instead represent recent successful NASS data sharing programs.

1. Farm Injury Survey - The survey data were collected for all 50 States over a three year period under a cooperative agreement. The sponsor for the three year project was the National Institute of Occupational Safety & Health (NIOSH) of the Center for Disease Control (CDC). The client requested that NASS be as open with sharing of the data as possible within NASS policy. When requested, the agency has provided outside data users with not only the summary data, but also masked individual record data. Several steps were taken to avoid disclosure of the identity of individual records.
2. ARMS - From the initial development and design of the Agricultural Resource Management Study (ARMS), a cooperative data collection and sharing agreement between NASS, the Economic Research Service (ERS), and the Animal Plant Health Inspection Service was a critical decision. This new survey program integrates multiple survey activities and survey data requirements for an annual farm finance survey, cost of production surveys, cropping practice surveys, chemical use survey, and a national animal health management survey. Data sharing details are still being finalized. However, the basic premise is that by sharing in the survey design, data collection, and summary process, respondent burden will decrease, survey costs will decline, and each agency will be able to make better use of a richer data set. Each of the federal agencies will abide by the NASS confidential and data disclosure policies. Individual identifiers will not be included in any sharable files.
3. NAHMS/ERS - As part of the continuing effort to reduce burden and lower costs, the National Animal Health Monitoring System (NAHMS) and ERS are making more use of data already being collected as part of the regular NASS survey program. Several data sharing examples include the use of the Agricultural Survey death loss data for cattle, hogs, and sheep for NAHMS and the land use data from the June Area Survey for ERS. These data are provided to NAHMS and ERS at virtually no cost other than extra summary processing. In most cases, NASS is

providing these organizations with data aggregations. NAHMS and ERS then perform additional data analyses and set their needed estimates.

4. Sheep Herd Management Survey - During 1996, NAHMS and the American Sheep Industries (ASI) requested NASS to provide a list of producers to contact for a mail survey on sheep herd management practices. NASS could not give a list of names to NAHMS or ASI because of list confidentiality but was able to provide the sample. A sample was drawn from the NASS list sample frame and sent to the Colorado State Statistical Office. The questionnaires were provided to the State office by NAHMS in pre-stuffed and sealed envelopes. The envelopes were then labeled by the office and mailed out to the respondents. By not labeling the questionnaire, when the respondent returned the questionnaire to NAHMS, the individual or organization identity was not available unless the respondent provided it. Only the State and strata size were known by NAHMS for each response.
5. Farm Labor Survey - The California Employment Development Division (EDD) and NASS are finalizing a cooperative agreement to consolidate each organization's survey program into one joint activity to begin January 1997. NASS will maintain the sampling frames and select a California sample to meet both agencies data requirements. EDD will collect and edit the data and the NASS California State Statistical Office will process the survey data. NASS and EDD will coordinate data analysis, data summary review and estimation, and data publication. EDD plans to implement an estimation procedure similar to NASS. A joint review of survey data is intended to eliminate program differences. EDD will adhere to NASS data confidentiality and disclosure policies.
6. Cooperative Research - The ERS and NASS, through a cooperative agreement [9], share microdata from the jointly funded annual farm finance survey. NASS collects data with the survey instrument identifying the collaborative effort of the two federal agencies. NASS and ERS share in data analysis and summary review. National production expense estimates are set in an Agricultural Statistics Board session where both organizations are represented. ERS abides by NASS Title 18 and 7 policies for data publication and research activities. An extension of this agreement is the joint approval by both agencies to research requests, typically from land grant universities. A memorandum of understanding is required between the researcher(s) and ERS which follow NASS established data confidentiality and unpublished data access policies. The research is conducted at ERS or NASS State Statistical Offices where microdata are used with individual identifiers suppressed. The number of these arrangements continues to increase as the federal agencies attempt to be responsive to a 1992 General Accounting Office audit report [10] recommending an expansion of this data sharing program.

7. EMAP - The Environmental Monitoring Assessment Program (EMAP) was a joint survey project with a memorandum of understanding (MOU) between NASS, North Carolina State University, the Agricultural Research Service-USDA, and the Environmental Protection Agency (EPA). On-site access to microdata was granted to this agriculture group of researchers and analysts after signing appropriate security and disclosure statements. Also, NASS removed individual identifiers, such as the latitude/longitude coordinates, from the first microdata file and created a second public database for other EPA research groups associated with forestry, arid lands, and surface water areas. These later EPA groups were not part of the original MOU and were not bound by NASS confidentiality and disclosure rules. NASS verified the public data base met disclosure requirements and approved all data requests. This second data base had a hydrological unit as a first level record identifier to ensure confidentiality. The unit is a large area of land usually greater than a county in size. This complex data sharing arrangement with multiple data bases showed creative efforts were used to follow disclosure rules while providing data user access to microdata. A loss of EMAP funding by EPA, however, ended this project.
8. Nematode Study - A data sharing agreement between NASS and the Smithsonian for a Nematode Study, conducted as an extension of the EMAP project, is still pending. When funding of EMAP was lost, there was interest in preserving the research and analysis of nematodes removed from the soil samples. The Smithsonian initially asked for the latitude/longitude identifier. To ensure confidentiality of reported data, NASS agreed to accept this request with the condition the coordinates of the identifier on individual data be aggregated to a county level.

Numerous other requests for unpublished NASS data are recorded by the agency. For example, in 1995 the Associate Administrator approved 31 data requests. So far in 1996, 24 new requests for unpublished data have been filed. This does not include NASS responses to data sharing needs that extends over several years. For example, unpublished State level farm wage rate data and production input price data are provided annually to ERS for internal analytical uses, again subject to disclosure and publication constraints.

V. Future Data Sharing Possibilities and Challenges

For over 125 years, NASS data sharing procedures have been rigorously enforced by adhering to legislation and Agency policy that ensures individual respondent protection from disclosure, facilitates proper interpretation of data and sound statistical procedures, and responds to customer requests for access to survey data for statistical purposes beneficial to the public interest. The Agency's reputation and credibility have and will continue to

depend on keeping this commitment to excellence in service to both the survey respondent and data user.

Significant data sharing challenges face the future of NASS and other federal statistical organizations. Some broad issues include the changing U.S. public sentiment towards an individual's privacy, the downsizing and reinventing of the federal government for more efficient use of public resources, the rapid advancement of technology in the areas of data storage, data transmission, and data processing, and the increased domestic and international demand for information. Also, new legislation to permit restricted access to data by federal statistical agencies for statistical purposes must be a priority.

More specially, there are several issues which impact current and future NASS data sharing policy. First, the new role of conducting and being the caretaker of the Agricultural Census will impact how NASS collects survey data as well as disseminates that data. Current statutes require that data sharing will continue with safeguards to ensure data confidentially and individual disclosure protection. Still unknown is how legislative policy, new statistical procedures, and data processing and systems development will change how business is conducted.

Second, NASS has expanded the scope of survey activities to a more diverse subject matter and thus the type of data and customers requesting data are more diverse. For example, the NASS Associate Administrator in a 1995 memorandum to Agency management [11] discussed recommended guidelines for public use geographic data. Two of the principles addressed were:

NASS will work with others to create data files with maximum allowable geographic identifiers, and

NASS will consider computer classified remote sensing map products as publicly releasable.

Third, NASS must continue to explore innovative methods to increase data access within policy guidelines utilizing advancements in technology and statistical procedures for data capture, storage, masking, dissemination, and security. Off-site processing and research activities are becoming a greater likelihood with improved data systems security and high speed telecommunications capability. Data warehousing from centralized storage of microdata files will increase the efficiency and accessibility of longitudinal and time series data relationships but will require suppressing individual identifiers for any data sharing to occur within confidentiality rules.

The NASS research agenda, as presented by Bosecker [12] in the 1996 NASS Research Division Missions and Projects Executive Summary, describes projects related to data sharing efforts. For example, a cooperative agreement is being arranged with George Mason University to develop and evaluate data smoothing and display

techniques that have meaningful information content to data users but do not reveal any individual data. Also, a project is underway to evaluate the quality of California Environmental Protection Agency data used by NASS in lieu of collecting similar data from sampled growers during the Fruit and Vegetable Chemical Use Surveys.

To conclude, NASS considers data sharing an integral part of its mission as a federal statistical agency. Each data request is responded to in a manner that ensures data confidentiality is enforced; data integrity is achieved; and data accessibility as a service to the public interest is accomplished. As stated by Allen [13], "It will continue to be NASS policy that data sharing will occur on a case by case basis as needed to address an approved, specified USDA or public need." It is imperative that data users understand, respect, and protect the confidential use of data with the same zeal and commitment as NASS to maintain and potentially improve future data sharing policy.

References

- Allen, Rich, (1995), "Memorandum on Guidelines and Principles for Public Use Geographic Data." [11]
- _____, (1993), "Memorandum Summarizing the Office of General Counsel's Interpretation of Statutes Governing NASS Data Sharing." [13]
- Bosecker, Ron, and Research Division Staff (1996), "Research Division Missions and Projects Executive Summary." [12]
- Caudill, Charles E. (1990), "A Federal Agency Looks for Answers to Data Sharing/Confidentiality Issues." [6]
- Kibler, W. E., and John E. Lee (1986), "Memorandum Stating Conditions on Use of Farm Costs and Returns Data." [9]
- OMB Task Force (1995 draft), "Data Sharing Task Force Report." presented to Office of Management & Budget, December 1994. [8]
- Peach, J. Dexter (1992), "Data Collection, Opportunities to Improve USDA's Farm Costs and Returns Survey." A report by the U.S. General Accounting Office to the Chairman, Committee on Agriculture, Nutrition, and Forestry, U.S. Senate, July 1992. [10]
- Panel on Confidentiality and Data Access (1993), "Private Lives and Public Policies." National Academy Press, Washington, D.C. [7]
- USDA, National Agricultural Statistics Service (1996), "NASS Service to Agriculture Through Statistics, Strategic Plan 1996." [1]
- _____, (1995), "PSM No. 5-95, Privacy Act of 1974." [2]
- _____, (1990), "PSM No. 6-90, Access to Lists and Individual Reports." [5]
- _____, (1990), "PSM No. 7-90, Release of Unpublished Summary Data and Estimates." [4]
- _____, (1996), "PSM No. 12-96, Standard for Suppressing Data Due to Confidentiality." [3]

Comments by James R. Nunns¹
on Papers Presented at the session on
“Sharing Data for Statistical Purposes”
Seminar on Statistical Methodology in the Public Sector
Council of Professional Associations on Federal Statistics
November 13, 1996

Introduction

It may be useful to begin with some background information on the Office of Tax Analysis (OTA) and its role in the Federal statistical community before commenting on the papers. OTA is not a designated statistical agency, but relies heavily on many statistical sources in preparing the official Administration forecasts of budget receipts, revenue estimates for proposed legislation, distributional and other economic analyses of tax provisions, and related responsibilities. Our main source of statistical data is tax return information prepared by the Statistics of Income (SOI) Division of the Internal Revenue Service (IRS), but we also make extensive use of the Current Population Survey and Survey of Income and Program Participation from the Department of Commerce, the Consumer Expenditure Survey from the Department of Labor, the National Medical Expenditure Survey from the Department of Health and Human Services, the Survey of Consumer Finances from the Federal Reserve Board, and many others.

OTA's heavy reliance on non-Treasury statistical data gives us a stake in improving these data. One potential source of improvement could be through expanded sharing of IRS data. However, along with IRS we have very real concerns about the impact of sharing IRS data on the level of tax compliance, on the public's perception of IRS, on the IRS's ability to monitor the use of shared data, and on the right of taxpayers to privacy with respect to the sensitive personal information contained in tax returns. For these reasons, OTA (and our legal counterparts in the Office of Tax Legislative Counsel) worked closely with IRS and Katherine Wallman and her staff in OMB (and through them, the statistical agencies) on drafting the amendments to Section 6103(j) of the Internal Revenue Code that are the companion to the proposed “Statistical Confidentiality Act.” Section 6103(j) governs the disclosure of tax return information for statistical use. We also prepared a “Commentary” on the proposed 6103(j) amendments, which is meant to spell out more fully the intent of the legislation.

The proposed amendments to Section 6103(j) would allow access to certain tax return information to the six “Statistical Data Centers” designated in the proposed “Statistical Confidentiality Act” that currently have no access (the only Centers that currently have access are the Bureau of the Census and the Bureau of Economic Analysis in the Department of Commerce). The Division of Research and Statistics of the Federal Reserve Board would also be given access.

¹ Director for Individual Taxation, Office of Tax Analysis, Department of the Treasury. The views expressed are those of the author, and do not necessarily reflect Treasury policy.

However, the current access to individual income tax return information by the Bureau of the Census would be conformed to that of other agencies', and thereby limited relative to current law.

From my perspective, the guiding principle behind the proposed amendments to Section 6103(j) was to balance the potential benefits of lower agency costs, improved quality and timeliness of data, and reduced respondent burden against the potential costs of lower tax compliance, adverse public perception of IRS, IRS's monitoring of shared data use, and loss of taxpayers' right to privacy. This principle lead to a "rifle shot" rather than a broad approach -- to give limited and closely controlled access to a limited number of agencies and only for constructing sample frames for censuses and surveys and related statistical purposes authorized by law. Only the minimum amount of information needed will be releasable -- minimum in items included, years covered, and number of taxpayers covered, and the data will be in categorical form with the largest possible categories. Further, sharing beyond the receiving agency can only be among the other eight designated agencies, only for purposes authorized by law, and only with the approval of the IRS.

Comments on Papers

Let me turn now to the two papers presented at this Session. I found both papers informative and useful. Thomas Mesenbourg's paper, "Realizing the Promise of Data Sharing," makes the very important point that the proposed legislation, both the Statistical Confidentiality Act and the amendments to Section 6103(j), won't be self executing. The Section 6103(j) amendments will require IRS to develop regulations (as under current law) that describe precisely which tax return information items, for which returns, and on what schedule, each agency will have access to. This will require complete documentation from agencies covering the legal basis for their requests, alternatives to tax return information, security arrangements for the data, and related issues. In addition, data sharing agreements will have to be approved by the IRS. To complete all of this regulatory apparatus will take time and effort on the part of agencies as well as the IRS.

Douglas Kleweno's paper, "Data Sharing -- A USDA/NASS Perspective," notes the transfer of responsibility for the agricultural census from the Bureau of the Census to NASS, but without implementing legislation (including the proposed amendments to Section 6103(j)) which would allow NASS access to tax return information. The IRS, Bureau of the Census, NASS, and OMB are working together to see whether we can solve the problem for now administratively. But it is worth noting that the proposed amendments to Section 6103(j) were developed to handle such problems, and this situation gives us a concrete example of why the legislation is necessary.

I also found Doug's paper very useful in explaining in some detail the mission of NASS, and in particular the extensive data sharing arrangements that NASS has entered into. As a cautionary note, NASS and other agencies may have to rethink some of these arrangements under the proposed Section 6103(j) amendments if they begin to receive tax return information and integrate it with data from their own surveys and other sources. However, there clearly are ways to work out many arrangements for data sharing, and I was impressed with the effort NASS has put into finding ways of sharing data while maintaining confidentiality.

The related point I would like to make is that the "Commentary" on the proposed Section 6103(j) amendments suggests that agencies and the IRS should search carefully for ways to limit, or even eliminate, the need for disclosure of confidential tax return information, while still achieving the agencies' statistical purposes. One approach, which I think might give us the best balance, would be to work out arrangements whereby agency personnel would be detailed to IRS and be given access to certain tax return information while under the direct control of the IRS. This would allow the agency personnel to do preliminary statistical analyses on the data, and to determine the minimum amount of data that would be necessary to leave IRS in confidential form under 6103(j). In some cases, perhaps no confidential tax return information would need to be disclosed to the agency. In this way, the agency would effectively have greater access to tax return information than they would otherwise have, while IRS would maintain maximum physical control of all confidential tax return information.

Concluding Remarks

I would like to conclude by urging all statistical agencies to look very carefully once more at the proposed amendments to Section 6103(j), with an eye to making sure the legislation is workable for them. While no one involved (I think I can speak for all) wants to reopen this proposal, we would much rather do that than find in a year or two (assuming the legislation is introduced and passed) that it doesn't work.

Session 10
REINVENTING ECONOMIC CLASSIFICATION

Revising the Standard Occupational Classification System
Thomas J. Plewes
Bureau of Labor Statistics

I. Introduction

Current occupational data and their underlying classification structures have come under criticism for being fragmented, incompatible, outdated, and lacking in skills information. In response to these criticisms, the Office of Budget and Management (OMB) decided to revise the U.S. Standard Occupational Classification system (SOC) to develop a unified classification structure that maximizes the usefulness occupational information collected by the Federal government.

To conduct the revision, OMB founded the SOC Revision Policy Committee (SOCRPC) in 1994. The Bureau of Labor Statistics chairs the SOCRPC and the Bureau of the Census, the Employment and Training Administration (ETA), the Office of Personnel Management (OPM), and the Defense Manpower Data Center serve as Committee members. OMB, the National Occupational Information Coordinating Committee and the National Science Foundation participate as *ex officio* members. Since its founding, the SOCRPC has operated under the following OMB guidelines:

The Policy Committee is charged with the examination of the Federal Government's various occupational classification systems for statistical and administrative uses, and with providing recommendations to OMB on the structure and implementation of a new SOC. The charge to the Committee includes: (1) identifying the major statistical uses of occupational classifications; (2) identifying and developing new concepts, structures, and methodologies to determine what constitutes an occupation; (3) developing and empirically testing a standard occupational system based on these concepts; (4) planning and the implementation of the new classification system; and (5) ensuring that there is ample opportunity for widespread public participation in the revision process.¹

This paper examines the past history, current process, and expected future results of the SOC revision. The "History" section traces some of the past difficulties of earlier classification systems, identifies the issues that a comprehensive classification system must address, and describes two classification innovations that have influenced greatly the current revision effort. The "Process" section describes the SOCRPC's mission and classification principles, outlines the Committee's research efforts, and describes the work groups formed by the Committee to produce the building blocks of the revised SOC. The "Result" section describes the Committee's progress to date, outlines the schedule for completing the revision, and concludes by discussing the future of the revised SOC.

¹ *Standard Occupational Classification Revision Policy Committee Charter*. Office of Management and Budget (October 1994).

II. The History of the SOC

The need for a classification standard was recognized with the development of a *Convertibility List of Occupations with Conversion Tables and Industrial Classification for Reports from Individuals*. These publications served as a bridge between the occupational classification system of the 1940 Census and the system used by the U.S. Employment Service to classify its operating statistics. Modifications to the Census classification system and publication of the third edition of the Dictionary of Occupational Titles (DOT) rendered the convertibility tables obsolete.

The development of an SOC began in December 1966 on the recommendation of the Interagency Committee on Occupational Classification. While work began in the mid-sixties, the first SOC was not published until 1977. The system was revised three years later. The members of the 1980 SOC policy committee agreed to a common SOC structure and to maintain "crosswalks" from their individual systems back to the SOC. The committee expected to update the classification system every five years. No subsequent maintenance was performed, and Federal agencies did not embrace the system. Subsequently, Federal occupational classification systems again drifted apart. As the systems drifted, their "crosswalks" became increasingly difficult to use, just as the convertibility tables of the forties became obsolete with the DOT's publication.

Essentially the same problem exists in 1996 that existed when common development began in 1940. There is a fundamental incomparability between the Federal government's two major occupational classification systems. Further complicating statistical comparability, other agencies have developed separate classification systems to meet their specific needs. The issue of reconciliation to achieve comparability has taken on greater significance since 1940, as Federal laws and mandates have increased demand for occupational data. In this sense the mission of the present revision is clear: Integrate the existing systems in a way that is responsive to data-users' needs.

The 1993 International Occupational Classification Conference served as a clearinghouse of new ideas and alternative approaches to occupational classification. The Conference included many individuals and agencies directly involved with the occupational classification user community, as well as international occupational experts from numerous countries. The papers, discussions, and ideas generated at the conference have informed the current SOC revision process.

A major area of discussion at the conference was the alternative classification concepts of "work-performed" versus "skills-based" classification. The 1980 SOC employed a work-performed model that grouped occupations into a socio-economic hierarchy.² The "work performed" criteria did not necessarily take into account the

² A socio-economic hierarchy refers to classifying workers into occupational groups such as managerial, professional, technical, sales, clerical, service, agricultural-forestry, or production.

education, training, and certification of individuals. An individual was classified in an occupation by a perceived level of "work performed." Many experts suggested a skills-based approach as a replacement to the previous work-performed standard. Proponents claimed a "skills" model for occupational classification better reflects the changing structure of the economy and is more responsive to the needs of data-users.

The Canadian system received a lot of attention at the international classification conference for its use of a skills-based model. The Canadian system employs a two-dimensional matrix approach to skills classification. The system defines a skill type combined with a skill level to classify a given occupation.³

The Ohio Bureau of Employment Services presented their use of the skills-based Canadian system to aid job placement. The Ohio Bureau needed a system for matching job openings to candidates. The State agency previously had used the DOT, but found the dictionary had too many titles (making it difficult to find matches). Moreover, agency staff claimed the DOT was out-of-date, and the work-performed criteria made classifying occupations difficult. By comparison, staff and customers found the "skills-based" system easier to use in matching job seekers with employers. The system also cost less to maintain in terms of data entry and computer processing time. From their experience, Ohio concluded that the "skills-based" system better reflected current and changing job requirements and furthered their ability to conduct labor market analysis.

ETA presented the findings of the 1993 Advisory Panel for the Users of the Dictionary of Occupational Titles (APDOT). The panel is credited with recognizing the DOT's "identity crisis" and acknowledging the need for covering new occupations, for developing a representative occupational structure, and for identifying skill and skill transferability.⁴ APDOT also outlined a list of classification issues that should be addressed in any comprehensive occupational system. A comprehensive system should define skill and worker distinctions, establish a common language for occupational information, and distinguish occupations that are measurable and collectible.

The success experienced by Canada and Ohio led to skills-based classification efforts by Federal producers and users of occupational information. Information on skills transferability is critical to understanding our labor market, which increasingly requires

³ The Canadian system bases skill type upon the skills required to perform the tasks and duties of an occupation. Skills type can include work that is specific to an industry when that industry is the sole employer of those skills. The Canadian system developed ten skill type categories such as Health, Sales and Service, Manufacturing, Trade, Transportation, Equipment-operation, etc. The skill or preparation level is defined by the length of education, training, or experience that is required for employment. The Canadian system defined four skill levels. The highest skill level requires at least a Bachelors, Masters, or Doctorate degree; the lowest requires high school education with a small amount of on-the-job training. *Proceedings of the International Occupational Classification Conference*, Report 883, Bureau of Labor Statistics (1993).

⁴ Pearlman, Kenneth. *Advisory Panel for the Users of the Dictionary Occupational Titles* (1993).

workers to move from occupation to occupation and from industry to industry. Two particularly innovative Federal skills-based classification schemes include the BLS Prototype Skills-Based Matrix and ETA's O*NET.

To make existing labor market information more useful to customers, BLS crafted a Prototype Skills-Based Matrix. The Matrix arranges Occupational Employment Statistics (OES) occupations by distinguishing work area and preparation level. From the Matrix, BLS developed job-search software called LASER. The LASER system provides labor market information on occupations requiring a skills mix similar to the customer's current occupation. By focusing on skills, the system shows customers the occupations to which they can most easily move.

In response to APDOT's call for skills information and common language, ETA began developing O*NET. O*NET has joined the descriptive language of the DOT to the labor market information developed by the OES system in a way that highlights skill and skill transferability. To create O*NET occupational units, analysts first mapped all DOT occupations to the OES structure. After evaluating the degree of "homogeneity,"⁵ "belongingness,"⁶ and "retraining time"⁷ of the DOTs assigned to each OES, analysts subdivided OES occupations into subclusters to best fit assigned DOTs. This effort resulted in disaggregating the approximately 750 OES occupations into about 1100 O*NET occupational units. To date, the system has gone a long way towards meeting APDOT's call for establishing a common language and for distinguishing occupations that are measurable and collectible.

The BLS Prototype Matrix and ETA's O*NET process furthered the skills discussion presented at the International Conference. The BLS Prototype Matrix and Matrix-based LASER software showed that a single skills-based system could serve the analytical needs of the research community and the pragmatic needs of job counselors. O*NET has refined skill information existing in the DOT by linking it to statistical labor market information. Both applications gave the SOCRPC a base on which to build a unified, skills-based SOC system. What remained was a question of process. A process requiring inter-agency cooperation, broad consensus, and commitment to implementation.

⁵ Homogeneity: "A consistent level of skill transferability between occupations within each occupational unit." John Nottingham and Jane Golec. *Prototype Development of the O*NET: The Occupational Information Network* (1995).

⁶ Belongingness: "Work activities of the DOT occupation match the work activities described in the OES definition." *Id.*

⁷ Retraining time: "The amount of time required by a worker in one DOT occupation to acquire the additional occupation-specific knowledge and skill required to perform proficiently in another DOT occupation." *Id.*

III. The SOC Revision Process

Given the OMB charge, the committee set about to fulfill its mission to integrate both household and establishment surveys and, to the greatest extent possible, meet the needs of the broad spectrum of occupational data users. The SOCRPC would meet these needs by adopting a common language for occupational classification that was skills-based and by developing a mechanism to accommodate occupational changes in the economy. To guide the development of the new classification structure, the SOCRPC crafted ten classification principles:

- (1) The Classification should cover all occupations in which work is performed for pay or profit, including work performed in family-operated enterprises by family members who are not directly compensated. It should exclude occupations unique to volunteers.
- (2) The Classification should reflect the current occupational structure of the United States and have sufficient flexibility to assimilate new occupations into the structure as they become known.
- (3) While striving to reflect the current occupational structure, the Classification should maintain linkage with past systems. The importance of historical comparability should be weighed against the desire for incorporating substantive changes to occupations occurring in the work force.
- (4) Occupations should be classified based upon work performed, skills, education, training, licensing, and credentials.
- (5) Occupations should be classified in homogeneous groups that are defined so that the content of each group is clear.
- (6) Each occupation should be assigned to only one group at the lowest level of the Classification.
- (7) The employment size of an occupational group should not be the major reason for including or excluding it from separate identification.
- (8) Supervisors should be identified separately from the workers they supervise wherever possible in keeping with the real structure of the world of work. An exception should be made for professional and technical occupations where supervisors or lead workers should be classified in the appropriate group with the workers they supervise.
- (9) Apprentices and trainees should be classified with the occupations for which they are being trained, while helpers and aides should be classified separately since they are not in training for the occupation they are helping.

(10) Comparability with the International Standard Classification of Occupations (ISCO-88) should be considered in the structure, but should not be an overriding factor.

After reaching consensus on the classification criteria, the Committee initiated several actions to launch the revision process and to fulfill the OMB charge. To ensure ample opportunity for widespread public participation in the revision process, the SOCRPC invited outside comment through two Federal Register notices. The first notice invited comment regarding the classification criteria, and the second invited comment regarding the organizational model. The SOCRPC also sought input from the Federal Consultation Group -- a group of Federal agencies who use occupational classification systems. OPM headed-up this group, which met quarterly to discuss the SOCRPC's progress.

In researching the underpinnings of the revised SOC, the SOCRPC commissioned six papers, which were presented at a seminar on research findings to the SOCRPC, the Federal Consultation Group, and other interested parties. The SOCRPC also collaborated with the Joint Program in Survey Methodology⁸ (JPSM) to develop a better understanding of how people perceive skills and training. The JPSM designed and conducted two focus groups to gain a qualitative understanding of how people potentially would react to a survey device seeking skills information. The JPSM found that since participants viewed concepts of skills in very different ways, questions designed to determine "general characteristics" of a job would be most effective. General questions, such as the degree of autonomy and level of education, may serve as good proxies for skill level.

Based on input from outside groups, Committee-commissioned research, and collaboration with the JPSM, the SOCRPC knew the kind of classification system it wanted to craft. To develop the detailed occupational units, which will comprise the SOC, the Committee organized six work groups based on skills groupings used in the BLS' Prototype Skills-Based Matrix.

- Work Group 1--Administrative and Clerical Occupations;
- Work Group 2--Natural Science, Law, Health, Education and Arts Occupations;
- Work Group 3--Sales and Service Occupations;
- Work Group 4--Construction, Extractive, Agricultural, and Transportation Occupations;
- Work Group 5--Mechanical and Production Occupations; and
- Work Group 6--Military Occupations.

⁸ The JPSM is a survey practicum course taught at the University of Maryland.

The SOCRPC designated the current OES classification system as the starting point for the recommended SOC occupations.⁹ These new SOC occupations would form the building blocks of the new skills-based system. The work groups conformed to the classification principles in bringing together three occupational classification systems: OES, O*NET, and the Census. The work groups invited input from experts in the field and solicited comments from professional and vocational associations to arrive at the recommended SOC occupations.

The SOCRPC chartered the secretariat to coordinate work group and policy committee interaction. The secretariat developed a report format for the review of the work group recommendations by the policy committee. The report incorporates the OES, O*NET, and Census components of each new SOC occupation presented, and maintains a historical record of decisions and changes to the occupations.

The policy committee is now in the process of reviewing the work group recommendations and converting the SOC to a skills-based job family matrix. At present, the SOCRPC has created 21 job family categories to serve as the structure for the revised SOC. The next step is to map the revised SOC occupations into the matrix structure.

Proposed Job Families

- Administrative and Financial
- Computer Related
- Engineering, Science
- Health Service
- Behavioral Science
- Community Service/Sports
- Education and Training
- Communications and Art
- Sales and Marketing
- Legal/ Protective Services
- Hospitality
- Cleaning
- Personal Care
- Extractive
- Construction
- Transportation/Material Moving
- Farming/Forestry/Landscaping
- Mechanical and Repair
- Production
- Plant and System Operation

⁹ The OES structure was used by the new O*NET system as well.

- Military

The SOCRPC agreed to the following aggregation due to the foreseen difficulty for occupational surveys to produce publishable estimates of twenty-one job families.

Data Publication Aggregation

- Administrative and Financial
- Engineering, Science, and Computer Related
- Communications, Art, and Recreation
- Education and Training
- Sales and Marketing
- Service
- Construction, Extraction, and Crafts
- Transportation and Material Moving
- Farming, Forestry, and Landscaping
- Production, Repair, and Plant Operation
- Military

IV. The Result

The process is not complete but the policy committee can see the light at the end of the proverbial tunnel. The SOCRPC will observe the following revision schedule. The committee expects to publish a third Federal Register notice by the end of January, 1997 and respond to comments in a timely manner. OMB will publish the revised SOC hard bound version by the Autumn of 1997. The publication will include occupational definitions and a list of alternate titles organized into a job family skills matrix. The SOC will be made available on diskette and a version of the SOC will be posted on the World Wide Web. A complete list of Census index items will be developed for household collection of the year 2000 Census. OES expects to adopt the revised system for the 1998 survey round. The revised SOC will be incorporated into the Post-2000 Current Population Survey.

The revised SOC will integrate household and establishment surveys together in a skills-based system. This system will reflect a changing economy and respond to the needs of data-users. Among the accomplishments of the revised SOC, is the direct link the system will have with O*Net, Census, and OES. Another innovation of the new system will be a skills matrix incorporating military occupations and reinforcing the idea of public/private job skills transferability. Crucial to the success of the new system will be OMB's mandate of Federal compliance. This mandate ensures Federal agencies will adopt the revised SOC. The revised system will implement these changes while maintaining the ability to make historical comparisons.

A final note, once developed this system must be maintained to avoid becoming obsolete as have past systems. To this end, the SOCRPC will maintain a review and decision making tracking system. Efforts must be made to ensure this revised SOC will be kept current well into the next century. Because, to borrow from Robert Reich, "Good public policy [regarding the work force] depends on good data about the workforce."¹⁰

¹⁰ Reich, Robert. *Proceedings of the International Occupational Classification Conference*, Report 883, Bureau of Labor Statistics (1993).

Reinventing Occupational Classification

Discussion by

**David W. Stevens, Executive Director
The Jacob France Center, Merrick School of Business
University of Baltimore**

Tom Plewes has provided a succinct description of the process followed and issues covered since 1994 by the Standard Occupational Classification Revision Policy Committee. Their effort will soon culminate in the release of a new SOC taxonomy. My comments reflect 30 years of interaction with both the users and producers of occupational information. This historical perspective underlies a prediction about the value that might be expected to emerge from use of a new SOC.

The concept of an *occupation* has evolved as the organization of work responds to the varied forces that determine how and where the nation's goods and services are produced. It is more difficult to agree about a practical definition of occupation today than before. This difficulty coincides with rising stakes in the classification that is adopted. Value gained or lost because of a particular classification decision is not uniform across all uses of the taxonomy. Education choices, training assignments, vocational rehabilitation strategies, and alien worker certification practices illustrate the transaction uses of employment statistics and descriptors that are released in occupational "buckets".

The collectability of accurate information is important here. There is a continuum of accuracy. A different cost is associated with achieving each point on this continuum. Compromises are inevitable in the process of reinventing an occupational taxonomy. The aggregation standard that is chosen at the data collection stage predetermines what can then be done with this information.

The current Congress is unlikely to appropriate sufficient funds to reach a level of accuracy in occupational statistics that would truly respond to the needs described above. The loss-of-value resulting from this underinvestment will be diffused across people and through time. This will jeopardize the nation's productivity and prosperity. Members of Congress and selected interest groups are challenged to think again about the future consequences of such casual parsimony.

The Federal government has made an irrevocable commitment to let the *Dictionary of Occupational Titles* be relegated to extinction. A relational database of occupational descriptors, now known as O*NET, will replace the DOT. Substantial thought and effort has gone into the design and pilot phases of this new approach. Routine public access to this modern approach to disseminating occupational information looms on the horizon. The value to the nation that can be expected to flow from this access will depend upon the accuracy of the descriptors contained in the database, and on the public's ability to use the database in a responsible manner.

The accuracy standard should not be set by default based on Federal funds availability. This standard is too important to individual and collective future well-being. It is expensive to collect useful data. It will be even more expensive not to meet an appropriate standard of accuracy. Poorly informed decisions will have such real consequences as mistaken career paths, improper vocational rehabilitation plans, and wrong alien certification decisions. Each, and all, of these will affect the nation's ability to take full advantage of its human capital potential.

Similarly, the Federal government should not walk away from its responsibility to raise public awareness of the availability of occupational information and how to use it. A case can easily be made that profit-seeking vendors will eagerly enter a queue to market the basic O*NET database. Proprietary refinements can be expected to appear. This will create confusion among those who have long relied upon the *Dictionary of Occupational Titles*. Which substitute product and bundle of supporting services should be adopted? Some level of consumer protection and information will be needed. This will be a common need across the states. Partnerships among federal, state and local governments, and vendors, should be encouraged. Again, this should not be done as an afterthought with residual funds. A conscious decision should be made to move ahead aggressively and soon.

Under Tom Plewes' able leadership the SOC Revision Policy Committee has completed its work. Tom has moved on to Army Reserve leadership. Other committee members have turned to new challenges. The new SOC that is their collective legacy should trigger a renewed commitment by the Congress, affected Executive Departments, and such organizations as the National Occupational Information Coordinating Committee and its state affiliates, the National Skill Standards Board, and the Interstate Conference of Employment Security Agencies, to champion the interests of the users of their data and services. This is not a time in the nation's history to short-change those who seek, and should have an entitlement to receive, occupational information that can be understood and acted upon to better their own lives and the lives of others who would benefit from more informed education, training, rehabilitation, and alien certification decisions.

Session 11

USING ADMINISTRATIVE RECORDS FOR STATISTICAL PURPOSES

ADMINISTRATIVE RECORDS IN OFFICIAL STATISTICS --

***HOW CAN OUR RESEARCH ON ADMINISTRATIVE
RECORDS IN CENSUS 2000 LEAD TO
GREATER USE IN 2010?***

Presented by:

**Ronald C. Prevost
Administrative Records Research Staff
Bureau of the Census**

**Council of Professional Associations of Federal Statistics
November 15, 1996**

ADMINISTRATIVE RECORDS IN OFFICIAL STATISTICS --

HOW CAN OUR RESEARCH ON ADMINISTRATIVE RECORDS IN CENSUS 2000 LEAD TO GREATER USE IN 2010?

Presented by:

Ronald C. Prevost
Administrative Records Research Staff
Bureau of the Census

Abstract

The vision of the Administrative Records Staff is to develop a corporate administrative records infrastructure that will maximize the effectiveness of each survey, census, and estimation operation by 2010. This evolution requires immediate research support on the topics of administrative records, census and survey micro-data linkages, aggregate administrative records and survey data uses, content modeling, data processing and warehousing. Concurrent research efforts addressing data access and privacy issues relevant to use of administrative records in statistical programs are essential. This paper identifies the research required to address our current perception of the major technical questions for social and demographic statistics.

BACKGROUND:

The expanded use of administrative records to develop official statistics in the U.S. has been discussed for decades. The Census Bureau has limited its use of administrative records largely to aggregate data, such as vital statistics and elementary school enrollment, in the development of population estimates. In the past, Census Bureau use of individually identifiable records has been limited to:

- Demographic Programs -- the Population Estimates Program uses IRS income tax return information in the computation of implied net migration rates,
- Economic Programs -- the Economic Area uses IRS forms to constitute the universe of establishments and provide wage and salary information for a variety of purposes in its census and survey programs,
- Decennial Census Program -- the Decennial Census uses information from administrative records for institutions and other non-household sources.

Operations based on the use of administrative records have created highly effective, timely, and cost-efficient measurements for the past 50 years. Recent advances in automation and record linkage accuracy have significantly increased the potential cost benefits from expanded uses of administrative records micro-data in statistical programs. Additionally, census and survey recipients have become increasingly resistant to respond to data collection requests. The combination of these factors has increased the opportunity for significant benefits through using administrative records in measurement systems. The potential use of administrative records as a means to (1) reduce cost and respondent burden in the census and, (2) improve the consistency and quality of surveys and population estimates, has become extremely attractive in the current environment of reinventing and refocusing the way we do business. This possibility encourages us to seriously address content comparability and quality issues inherent in any use of administrative records.

The Plan for Census 2000, released in February 1996, includes a proposal to use administrative records as a data source for 5 percent of non-responding households through Integrated Coverage Measurement and to supplement item responses. This proposal is in addition to the use of administrative records from the U.S. Postal Service's Delivery Sequence File that will help the Census Bureau improve the list of residential addresses. This bold approach has led to heightened interest in research efforts, such as data linkage and database construction, as well as methodologies for incorporating administrative record information into new estimation applications.

The Census Bureau documented its initial experience with developing administrative records sources during the 1995 Census Test and is continuing research to determine how we can improve upon our approach. We have created the Administrative Records Research Staff to conduct research on the use of administrative records in demographic surveys, population estimates programs, and population and housing censuses. This approach is in concurrence with Recommendation 8.2 of the National Research Council (NRC) report, "Modernizing the U.S. Census," (see attachment). Based on our initial results and the many uncertainties associated with proposed changes in the Federal programs that generate many of the records, it appears that full use in Census 2000 may exceed current understanding and capabilities.

OUR VISION FOR 2010 AND BEYOND:

The vision of the Administrative Records Staff is to develop a corporate administrative records infrastructure that will maximize the effectiveness of each survey, census, and estimation operation by 2010.

We are in the initial stages of developing a plan to accomplish this vision. This evolution requires immediate research support on the topics of administrative records, census, and survey micro-data linkages, aggregate administrative records and survey data uses, content modeling, data processing and warehousing. Concurrent research efforts addressing data access and privacy issues relevant to the use of administrative records in statistical programs are essential.

The focus of our vision is on how the Census Bureau could reinvent and revolutionize data collection and processing operations using administrative records in preparation for the next millennium. As this system develops, we plan to blend the results from each survey, census, and estimation operation into a single, combined, final product that incorporates the best attributes of each individual measurement device. Ultimately in this type of system, the decennial census would become an extension of annual operations and would serve the purpose of benchmarking and quantifying the effectiveness of these operations on a national scale. This vision distinctly contrasts with our old view of a demographic system based primarily on the once-per-decade census, which was a "snapshot" of America.

This vision of a statistical measurement system is an aggressive approach designed to draw strength from integrating current survey or census operations with data from administrative record files. This vision was clearly endorsed by the National Academy of Sciences' recent reports on rethinking methods for decennial censuses, population estimates, and periodic surveys using administrative record sources. We anticipate this approach will:

- Reduce respondent burden by either eliminating individual contacts or reducing the information requested of individuals,

- Reduce the cost of data collection by using administrative record information in lieu of survey or census data,
- Increase data quality by employing individual measurement operations as evaluations of each other.

Our approach is driven by three premises:

1. Administrative records will have the greatest effect on Census 2000 through a variety of coverage improvement operations. Administrative records can be used to improve address lists, define sites that require special enumeration and marketing, target field follow-up operations, and enhance Integrated Coverage Measurement procedures. Data collection costs will be reduced as a result of these operations.
2. Administrative records have the potential to substitute for, or supplement data on, census and survey forms. The National Academy of Sciences has strongly recommended that the Census Bureau pursue increased use of administrative records for censuses as an alternative to direct enumeration. This recommendation was motivated by escalating costs of census enumeration, increased reluctance of the public to respond to census data requests, and budget uncertainties. This alternative methodology requires extensive development before it is ready for implementation in large-scale operations.
3. Administrative records in combination with direct measurement operations have the capability to create high quality, timely annual statistics of population, housing and their characteristics for census tracts and blocks.

We recognize that this reinvention process will require a substantial level of research to become a successful endeavor. We must also develop an approach that provides the necessary flexibility to circumvent inevitable roadblocks to our success. Therefore, we are approaching our vision from two stages--Research Toward 2000 and Toward 2010.

RESEARCH TOWARD 2000

A reinvention activity requires revisiting the premises underlying an operation, determining if current or proposed technologies can be implemented practically, and configuring ongoing operations in the most effective manner. Since measuring the resident population is one of our primary objectives, we will need to reexamine the current operations in light of expanded administrative record use.

Three options are available for statistically representing our nation's population, primary measurement, secondary measurement, or a combination. Primary measurements include operations that ask individuals specific questions through censuses and surveys. Secondary measurements employ data collected by other agencies such as administrative records, which the Census Bureau uses as a proxy for individual responses. Because administrative records are a secondary data source and are not collected for the purposes of enumerating the population and its characteristics, the responses provided by individuals on these records may not match the concepts measured through direct data collection. Using both primary and secondary record sources, we must design a system that provides the flexibility to produce a statistical representation of the United States under constraints of changing administrative records availability and funding for direct collection efforts. Initially, the best solution is the integration of primary and secondary measurements. Ultimately, the solution may rely more heavily on secondary data.

The research agenda must be structured to meet the short-term goal for research and evaluation in the context of the 2000 Census operation and the long-term goal of an administrative records database as a corporate resource for all statistical programs. The agenda should be developed to create a body of general knowledge applicable to our long-term goals. There are many questions and uncertainties about the process of creating population and housing databases from administrative records and their use by the Census Bureau's statistical programs to make counts and estimates. We have simultaneously begun research from two different approaches (1) micro-data modeling and (2) aggregate data modeling.

Micro-data Modeling

In the micro-data modeling approach we are exploring our ability to create population and housing databases with administrative records and to use these records as a proxy for direct measurements. (Note: numbers enclosed in parentheses relate each topic to the chief NRC Recommendation they address --see appendix.)

Issues identified for research:

- What is the geographic and population coverage, content, quality, and timeliness of information contained on Federal, state, local, and private vendor administrative records in relation to direct measurement instruments such as censuses and surveys? (8.2)
- Can we develop processing procedures that produce accurate matches of administrative records to the MAF and facilitate accurate geocoding? (8.2)
- Can we develop administrative record systems that provide accurate population and housing unit coverage at different levels of geography? (8.2)

- What matching variables such as name, address, date of birth, social security number, phone number, and record linkage procedures will provide accurate record matching and elimination of duplicate administrative records gleaned from multiple sources? (5.6)
- How can we best link telephone numbers to individuals at their residence or elsewhere? (5.6)
- How do we use administrative records to improve the sample design for various surveys (e.g., improved stratification of frame or identification of special interest populations)? (6.2)
- What is the magnitude of missing data items in administrative files? In the absence of key data items such as gender, race, and Hispanic origin, can procedures be developed for accurate imputation? (7.1)
- How do we use administrative records in estimation to enhance the quality of the estimates now made by demographic surveys or those anticipated from the American Community Survey? Are reductions in mean square error achieved? (6.2)
- How do we use administrative records in imputing for missing data from direct enumeration? (6.2)
- How do we construct households and families using administrative records? How do we estimate their characteristics using administrative records? (8.1)
- Do administrative records have the potential to meet emerging data needs of the type previously met through direct data collection? (6.2)

Benefits to Census 2000 -- An increased capacity to employ micro-data administrative records can be used to:

- Evaluate and improve the quality of MAF by enhancing the address list and geocoding addresses correctly, (8.2)
- Supplement Census 2000 operations for non-respondents, (5.6)
- Enhance implementation of Small Area Estimates Program statistics (see Aggregate Data Modeling section for discussion) in refining MAF evaluations and targeting areas requiring field follow-up procedures, (8.1)
- Reduce the cost of collection operations by correctly attributing a telephone number to a non-responding address, (5.6)

- Aid in decennial census coverage evaluation (ICM) and non-response operations. (5.6)

Aggregate Data Modeling

In the aggregate data modeling approach, we are exploring a limited expansion of the Small Area Estimates Program. The Small Area Estimates Program of the Census Bureau is a cost-effective solution to providing annual statistics for Federal resource distribution. Expansion of this program will provide local governments with the capacity to enhance their decision-making process to further direct resources to areas of need. This is particularly important in an environment of devolution.

Issues identified for research:

- How do we expand the geographic detail of our subcounty estimates to develop accurate population and poverty estimates for census tracts and blocks? (8.1, 8.2)
- How do we develop estimates of housing, housing vacancy, tenure, and structural characteristics for counties, census tracts, and blocks? (8.1, 8.2)
- Can we improve the quality of the current product through an increased understanding of administrative records data concepts? (8.1, 8.2)
- Can we develop/enhance products that work in cooperation with ongoing surveys and the new American Community Survey to improve the quality of the surveys and develop feedback loops that also provide quality assessments of the small-area estimates? (8.1, 8.2)
- Can aggregate administrative records information be used as initial cost-effective coverage and content assessments for micro-data administrative records? (8.1, 8.2)

Benefits to Census 2000 -- Expansion of the Small Area Estimates Program provides products that can be used to:

- Evaluate the MAF along with the MAF Quality Improvement Program, (5.5)
- Preselect/target resources to areas requiring special enumeration procedures, (5.3)
- Target Census 2000 promotion activities, (5.3)
- Assist Demographic Analysis in enhancing survey-based ICM estimates, (5.6)
- Review the final statistics produced from Census 2000. (7.1)

TOWARD 2010:

We do not yet know whether a 2010 administrative records census is a realistic goal. What we learn between now and the year 2000 will drive the research agenda. If the planned uses of administrative records prove feasible, their implementation will result in a wealth of data for analysis. Immediately after the 2000 Census, the census records become a very valuable research record set to use in comparisons with administrative record files. Research questions will focus on the feasibility of an administrative records census in 2010. The cumulative knowledge acquired through the research agenda and the implementation process for Census 2000 will produce necessary evaluations to help us determine how aggressively we pursue the incorporation of administrative records into an ongoing statistical measurement system.

Appendix: Selected Recommendations from the National Research Council
Source: Summary: Modernizing the U.S. Census., 1995

Recommendation 5.3 The panel recommends that the Census Bureau incorporate successfully-tested procedures to increase the initial response rate in the 2000 census, including the use of respondent-friendly questionnaires and expanded efforts to publicize the mandatory nature of the census.

Recommendation 5.5 The panel recommends that the U.S. Postal Service and the Census Bureau continue to work together to improve the decennial census. We endorse the expanded role for the U.S. Postal Service in the 2000 census in several areas: (a) development, maintenance, and improvement of an accurate address file for the nation's residential housing units, (b) checking the address list prior to the census to improve accuracy, (c) delivery of the mailed forms, and (d) ascertainment of the vacancy status of housing units during the census.

Recommendation 5.6 We recommend that the Census Bureau undertake a thorough reexamination of the basic structure, organization, and process by which the decennial census is conducted to obtain the full cost-saving potential of the proposed redesigned census. As one part of its reexamination, the Census Bureau should develop a plan for the 2000 census that eliminates a substantial fraction of the \$1.3 billion cost increase (in 1990 dollars) from 1970 to 1990 that is not accounted for by the growth in housing units and the decline in the mail response rate. The target for this plan should be much more than the \$300 to \$400 million we have already identified.

Recommendation 6.2 The panel recommends that the Census Bureau broaden its research on alternatives for more frequent small-area data to encompass a wider range than continuous measurement, as currently envisaged. In that context, the Census Bureau should examine the cost-effectiveness of alternatives, the ways in which they meet user needs, and the manner in which continuous measurement or other alternatives could be integrated into the nation's system of household surveys. The research program should be carried out in cooperation with the federal statistical agencies that sponsor household surveys and should include evaluation of the quality of important data elements, the frequency and modes of data collection, and the manner in which the results would be presented, as well as methods for introducing change over time.

Recommendation 7.1 The panel recommends that the Census Bureau expand its examination and testing of race and ethnicity questions to provide comprehensive information on: (1) public understanding of the concepts and acceptability of questions, (2) compatibility among the several census items and the utility of cross-tabulations, (3) comparability of census data to race and ethnicity data collected in other federal surveys or obtained from administrative records, and (4) the quality of data for small areas and specific groups. This research needs to

be given high priority so that the results may be incorporated into the review of Statistical Directive 15 currently being conducted by the Office of Management and Budget.

Recommendation 8.1 The panel recommends that the Census Bureau work to improve the amount, quantity, and frequency of small-area intercensal data:

- The Census Bureau should conduct experiments with federal administrative records for deriving more frequent small-area intercensal data estimates. At a minimum, the panel recommends that the Census Bureau geocode several large federal administrative record systems and use them to produce small-area estimates.
- The Census Bureau should work with state and local governments to enhance the quantity and frequency of small-area data.

Recommendation 8.2 The panel recommends that the Census Bureau give a single unit sole responsibility to exploit administrative records and produce small-area intercensal estimates on a frequent basis. Its work on administrative records should examine geographic consistency and quality. The unit should develop methods for increasing geographic content; establishing consistency of federal, state, and local administrative data; augmenting content on national records; augmenting usefulness of the resulting information through modeling; and computerizing approaches to database management to facilitate the use of administrative data in a census. If the content of administrative records can be improved for use in preparing small-area estimates, that is desirable, but the major purpose of the unit would be to produce small-area intercensal estimates.

Matched Data and Social Security Analysis: Nuts and Bolts

Bernard Wixon

(202) 282-7183

Bernard.J.Wixon@ssa.gov

Dedication: To Dorothy S. Projector, who brought the use of eligibility microsimulation models to the Social Security Administration and who taught me why we could not simulate social security and disability using surveys alone.

December 11, 1996

Division of Economic Research
Office of Research, Evaluation, and Statistics
Social Security Administration

1. Introduction¹

Much of the literature on data matching appeals to intended uses which seem narrow in scope. For example, matching enables analysts to refine income estimates by comparing estimates from two or more sources. Such an example suggests, however, that what is at stake is a seemingly marginal improvement in existing estimates. On the other hand, for analysts concerned with social security issues, discussions of matched administrative and survey data take on a tone of urgency. Without such data, analysts are largely unable to simulate effects of major structural changes in social security. Nor can they describe those who benefit most from the progressive structure of worker benefits, in terms of such basic characteristics as family income. Nor can those found disabled under SSA's medical-vocational criteria be studied in the light of recent household events which may have encouraged them to apply for benefits. Such analytical tasks represent basic public program analysis--estimating how benefits are targeted and why people apply. Analysts studying most public programs deem such tasks routine and address them, in large part, by using household survey data. Yet due to the data requirements of contributory programs and disability programs, such analyses cannot be undertaken for social security programs without matched administrative and survey data.

I have had first hand experience with these data gaps. Twenty years ago, I worked with a small staff developing a microsimulation model to study the distributional effects of a broad range of programs. These included means-tested programs (for example, Supplemental Security Income or SSI and Aid to Families with Dependent Children) and tax programs (specifically, the social security payroll tax, and the individual income tax). Notably absent from this list, however, were the country's largest cash benefit programs: the social security retirement program (in this case, Old Age and Survivors Insurance or OASI) and the two

¹ I want to thank several people for their comments, including Michael Leonesio and Fritz Scheuren, who served as discussants, as well as Ben Bridges, Barry Bye, Deb Dwyer, Susan Grad, Tom Jabine, David Pattison, and Denny Vaughan.

disability programs (Disability Insurance or DI and the disability portion of SSI).

In this paper, I consider the long standing data gaps underlying these analytical limitations, as well as the potential contributions of exact matching of administrative and survey data. These gaps affect the most basic program analyses--such as evaluating program targeting and application incentives. Neither program data nor household survey data, taken alone, support a reasonable course of distributional and behavioral research. Program data lack a full set of attributes on well being, demographic traits, household composition, and recent household events. Household survey data include such attributes, but have only limited program information. Eligibility simulation models can be used to add programmatic information to survey data; however, this requires that the survey contain program-specific elements of eligibility and benefit determination. In comparing means-tested and contributory programs, I show that household surveys, without administrative data matching, can be used to simulate benefits for means-tested programs, but not for contributory programs. A separate problem unique to disability programs is also discussed: How can survey responses be used to estimate who would be found disabled under SSA criteria? One approach is outlined, involving matching of information on SSA determinations to household survey data. Taken together, the need to simulate social security benefits and to estimate the disabled serves as a rationale for a continued program of administrative and survey data matching.

Section two will discuss some of the basic ideas underlying the use of household data in program analysis, including brief sketches of distributional analysis, behavioral analysis, and eligibility modeling; readers with background in these areas may skip this section. In section three, I consider how household survey data, in combination with eligibility simulation modeling, support basic program analysis for means-tested programs. In section 4, I turn to the social security retirement (OASI) program to explain why--in the case of a contributory program--administrative data on earnings must be matched to survey data in order to undertake distributional or retirement modeling. Section 5 discusses a problem unique to the disability programs and outlines an approach taken by social security analysts to address it. Some concluding thoughts are presented in section 6.

2. Some Basics of Program Analysis Using Household Data

Why Data on Individuals?

Analysis using microanalytic data--data on individuals and households-- prevents the loss of information frequently associated with efforts to model program eligibility or the behavior of individuals with aggregated data. The relationships of interest stem from program provisions (e.g., eligibility for benefits) and behavioral responses (e.g., decisions to apply for benefits). Such relationships depend on *combinations* of traits of particular persons or families.

Nonetheless, because disaggregated data are sometimes not available or because of the tight deadlines associated with policy discussions, estimates based on aggregated or group data are sometimes used. But when only highly aggregated data are available, information on combinations of traits or joint distributions is often lost and assumptions must play a greater role. The use of microanalytic data is intended to avoid this loss of information.

Furthermore, analytic approaches that preserve the heterogeneity in the underlying data also serve the increasing interests of policy makers in understanding program effects on population subgroups. For this reason, when modeling decisions by individuals or program eligibility determinations, analysts often choose individual data--typically from household surveys--if such data are available.

Data Collection Mandates

Those who collect information in the course of program administration or through household surveys have different objectives and constraints. Typically, program administrators collect only the information needed to administer the program, that is, to determine eligibility and benefits.² Program data are limited both with respect to the universe and the attributes collected. On the one hand, information is collected only for those who apply for benefits (and, in the case of a contributory program, for taxpayers). On the other hand, the

² There are times when program administrators collect information "for statistical purposes." However, under resource pressure, collection of some statistical information suffers, since it has no direct bearing on the determination of eligibility and benefits. In many cases, collection of such data degrades over time and the data become unusable.

information collected is mainly limited to what is needed to determine eligibility for benefits and the size of the benefit.

Household surveys, in contrast, collect information on a wider range of attributes. Information on such attributes is needed to investigate (1) distributional effects of programs on subgroups of interest (for example, race, income classes, education, marital status) and (2) incentives to apply for program benefits as an alternative to work and to consider other behavioral effects. And, although surveys are limited by the use of samples, the samples often represent the general population. However, surveys lack detailed program information, such as whether a respondent is eligible or how the respondent's benefit would be affected by a specific program provision. Indeed, while some such program information is uncollectable using standard household surveys, it is nonetheless central to program analysis.

There are other important differences as well. Because of recall constraints, surveys generally collect data relating to the time of collection. Survey cost constraints preclude data collection efforts involving both a large sample and repeated surveys over many years. This is possible in the course of public program administration, however, if such data collection is integral to program administration.

The Microanalytic Toolkit

Distributional analysis and behavioral analysis represent some of the most basic approaches to evaluating the effects of public programs or of proposals to change such programs. As they are now practiced, however, both approaches are often linked to eligibility modeling. Below is a brief sketch of each. It is not my purpose to provide a detailed description, since this paper considers the data requirements of these program analytic tools, rather than the tools themselves. However, I do mean to suggest that distributional analysis and behavioral analysis represent the most basic means by which program analysts estimate whether a public program is targeted to those intended, as well as the behavioral incentives created by the program.

(1) *Distributional analysis*.--Distributional analysis involves demonstrating how eligibility or benefits from a program or a specific program provision affect population subgroups of interest. In one sense, it is not obvious why analysts would examine such effects, since eligibility and benefits are carefully determined based on specific traits, such as advanced age and low income. That is, if benefits are carefully targeted to the low income aged, for example, why should analysts or policy makers be concerned about effects on other subgroups?

Broadly stated, distributional analysis addresses the issue of target efficiency of a public program or of specific program provisions. Target efficiency involves two elements: Ensuring that all intended recipients receive benefits and that *only* intended recipients receive benefits. But implicitly, the "intended" population is defined multidimensionally--not just in terms of explicit eligibility criteria, but in terms of the broad interests of policy makers.³ For that reason, program data have limited value for distributional analysis. The point of distributional analysis, then, is to evaluate *government decisions* with respect to eligibility and benefits from the standpoint of broader criteria than those explicitly used in determining eligibility and benefits.⁴ To support such analysis, the data source must include a comprehensive set of variables on well being and demographic traits and the universe should include nonparticipants to serve as a comparison group. Eligibility simulation models are

³ For example, how the taking of early, reduced retirement benefits might affect poverty rates might be of considerable interest. Yet, because program data do not include measures of income from all sources, it is difficult to investigate such issues using program data alone.

⁴ Because distributional analysis often consists of tabulated characteristics of persons falling into programmatic or socioeconomic categories, it is sometimes called "descriptive" analysis. This characterization is misleading in the sense that, as policy analysts know well, normative conclusions are often drawn from such information. Logically, this is possible because policy makers or analysts bring to the table well established views of the purposes of the programs. A logician would call such views latent normative premises. Let us consider an example.

- policy maker's latent normative premise: a program for the low income aged should treat singles and couples equivalently.
- estimates show: a much higher percentage of single recipients are in poverty
- policy maker's normative conclusion: there is reason to question the design of the program with respect to family composition.

often used to define those affected by a specific program provision.

(2) *Behavioral Analysis*.--Behavioral analysis considers decisions of individuals, such as whether to apply for benefits and to quit or reduce work.⁵ Such decisions are unlike government decisions, for which the determining factors are fixed largely by statute and regulation. The decisions of potential applicants to apply for program benefits are private, psychic events.⁶ For that reason, the analyst uses statistical techniques to infer both the major factors involved in the decision and the roles played by such factors.

Optimally, the data should include information on the major factors hypothesized to determine application decisions and the data should distinguish applicants from nonapplicants. Such factors typically include measures of economic resources, demographic characteristics, recent household events, and the size of the potential program benefit. In the context of applications analysis, the size of the potential benefit has consistently been shown to be a major determinant of the decision. For that reason, eligibility models are frequently used in the analysis of decisions to apply for benefits. One implication of this is that behavioral analysis will be limited, unless the data set includes the information needed to estimate eligibility and benefits.

Once a model has been estimated, the decision factors, their relative importance, and the direction of their effects are known. The model can then be used as a policy simulation tool. In the course of simulating the short term response to a policy change, for example, the model might be used to estimate how many additional persons would apply if benefits were increased. Long term applications, in contrast, involve projecting the changes in applications that are implied by expected changes in the decision factors.

⁵ While behavioral analysis can focus on savings behavior, family formation, and other decisions, I will mainly discuss the decision to apply for benefits.

⁶ Many now believe that some state welfare organizations, hospitals, and firms mandate or encourage application for federal benefits. To the extent this is true, the application decision has become institutionalized, to some degree.

(3) *Eligibility Models.*--In most programs, eligibility determination and the computation of benefits is a lengthy and seemingly arcane process. However, by focussing on program detail in the foreground, it is too easy to lose perspective: program eligibility criteria reflect society's normative decision making--they determine who receives benefits from public programs. Hence, eligibility provisions represent choice variables through which policy makers can refine distributional effects or cut costs, for example. Eligibility determination has two parts, categorical and financial.

Categorical (or nonfinancial) criteria embody an underlying premise--that the programs are not intended for all, but for those not able or not expected to work.⁷ Hence the categorical criteria for public programs correspond to population subgroups broadly considered to be dependent on support from other members of society or entitled to social insurance benefits. The criteria for some major programs are old age, disability, and single parent status.

Financial eligibility and benefits can be estimated using a simulation model if the data set has the needed elements. Financial eligibility and benefits for means-tested programs, for example, are determined by such factors as low income, low assets, family size, and living arrangements. In government decision making of this type, the decision process is detailed in legislation and regulations and should be applied uniformly for all applicants. Assuming the data requirements are met, the decisions are largely replicable using an accounting framework that mimics the eligibility and benefit algorithms.

These analytical tools represent one type of microsimulation model, since they simulate program eligibility using information on persons or families. From a historical perspective, such microsimulation models have provided a framework for incorporating data on individual persons and households into the distributional, behavioral, and cost analyses of benefits and taxes. These models use information on individuals and families on a case-by-case basis, taking each case through an eligibility and benefit determination procedure which mimics the

⁷ Some programs, such as food stamps and the individual income tax, do not have such categorical criteria. However, most benefit programs, including those discussed here, are categorical.

claims process itself.

Survey-based microsimulation models give policy makers an effective means of considering hypotheticals, such as:

- *the program eligibility of individuals who have not applied for benefits.*--Eligibility models are used in conjunction with surveys representing the general population, allowing analysts to estimate the number of eligibles and their hypothetical benefits. Such estimates are useful for evaluating potential program growth. Also, eligibility estimates allow construction of participation rates, allowing analysts to evaluate a means-tested program in terms of how well it reaches its targeted population.
- *the eligibility, benefits, behavioral responses, and costs associated with benefit structures which have been proposed, but not enacted.*--An eligibility model's step by step representation of the eligibility and benefit determination procedures allows analysts to consider "what if" alternatives to the current structure. In this way analysts can either consider broad structural alternatives or incremental changes to specific provisions of the benefit structure.

3. Use of Household Surveys to Analyze Means-Tested Programs

In this section, I will discuss the data requirements for analysis of a public program, focussing on attributes of the sort collected in household surveys, as well as methods for introducing detailed information on eligibility and benefits. I will consider the SSI program (aged portion) which serves as an example of means-tested programs in general. This discussion will show why household data, without administrative data matching, can support an extensive course of research for some public programs. Using the SSI/aged case as a baseline, I will then consider (in the next section) programs with more vexing data requirements.

Attributes

First, it is helpful to distinguish attributes collected from respondents to household surveys

and those collected from program applicants in the course of program administration. But rather than classify a large number of variables, I would like to discuss broad groups of variables, mainly in terms of their roles in the analysis of public programs. Let us consider the two tiers of variables displayed in Table 1. The top tier includes examples--though not an exhaustive listing--of variables which can be used to determine program eligibility and benefits for SSI (aged portion). Categorical eligibility requires information on age, while financial eligibility is determined on the basis of income, assets, family composition, home ownership, and other factors. Such information is, of course, collected from applicants by program administrators. However, many household surveys collect comparable information.⁸ This has a marked impact on the scope of program analysis which can be pursued for means-tested programs.

The second tier of variables, labeled "evaluative," are used differently. From the standpoint of their source, they may be thought of as variables collected by a household interview survey designed to support program analysis. From the standpoint of their use, they are likely to be used by analysts in two ways: (1) to define subgroups used in distributional analysis (for example, age, education, marital status, income class) and (2) as explanatory variables in behavioral analysis (for example, income, job loss, loss of health insurance). That is, these are the variables needed to support distributional or behavioral analysis for any public program. However, since most of these variables play no role in determining eligibility or benefits, they will not be available in program data.

Populations

Public program analysis also entails defining relevant subgroups. For example, program data include information on participants and, in some cases, limited data on denied applicants. General population surveys, taken alone, provide only a distinction between program

⁸ The information collected in surveys is typically not identical to that collected for program administration and surveys often do not collect items with a minor role in determining eligibility and benefits. The Survey of Income and Program Participation (SIPP) was explicitly designed to collect the elements of eligibility and benefit computation for a number of programs.

Table 1
Variables Classified in Terms
of their Analytic Use

Attributes

Program Eligibility Variables

Categorical Eligibility

(age)

Financial Eligibility

(income, assets, family size
home ownership . . .)

Evaluative Variables

Well Being

(income, assets . . .)

Demographic Traits

(age, race, ethnicity,
education, health . . .)

Household Composition, Events

(marital status, loss of spouse,
job loss, loss of health ins. . . .)

beneficiaries and nonbeneficiaries and a measure of the size of the benefit. But when surveys include variables needed to support determination of eligibility and benefits through the use of eligibility models, those eligible--irrespective of whether they receive benefits--can also be estimated. Table 2 suggests the key populations.

Table 2 illustrates how questions in a household survey, combined with an eligibility model, can be used to define these populations. Whether respondents participate or not is based on the response to the survey question on benefit receipt. On the other hand, respondents not receiving benefits may not know whether they are eligible for a program. Program eligibility, then, is an attribute that often cannot be collected in surveys; it must be modeled. As Table 2 illustrates, the survey question on benefits, in conjunction with eligibility modeling, allows analysts to define four populations:

- Eligible participants.--This group includes the great majority of beneficiaries.
- Eligible nonparticipants.--The size and traits of this group are of interest, first, because they suggest the potential for program growth. Second, they represent a key control group vis a vis participants; that is, differences between eligible nonparticipants and participants are important to understand incentives created by the program.
- Ineligible nonparticipants.--This group, which includes most members of the general population, is useful mainly as a control group. The boundary between ineligible nonparticipants and eligible nonparticipants varies depending on the eligibility criteria simulated. Hence, analysis of alternative definitions of eligibility can only be undertaken if nonparticipants are included in the sample.
- Ineligible participants.--Participants can be found ineligible for several reasons including reporting errors, survey/program inconsistencies (if the survey reference period differs from the program accounting period, for example), errors in the modeling of eligibility, or fraud. This is typically the smallest of the four groups and our data often do not allow valid inferences about this group; for that reason, they will be excluded from further consideration here.

Those affected by specific program provisions.--Increasingly, program analysts study not just those receiving benefits from a given program, but those affected by specific provisions of the

Table 2
Populations Relevant for Program Analysis

| | | | |
|----------------------------|------------|---------------------------------|--------------------------------|
| Is Person Eligible? | yes | Is Person Participating? | |
| | no | yes | no |
| | yes | Eligible Participants | Elig. Nonparticipants |
| | no | Inelig. Participants | Inelig. Nonparticipants |

program. That is, we not only look at subgroups defined in terms of, for example, demographic traits or family income--we also look at subgroups affected by provisions such as an asset test or marginal tax rate, to consider distributional and incentive effects of such provisions. This approach is "analytical" in the formal sense of that word, breaking a whole--in this case, the program's benefit structure--into parts.

SSI (Aged)

While the principal focus of this paper is to explain why matched survey data are critical for analysis of social security programs, for purposes of comparison I will discuss how, in the case of means-tested programs, household survey data, unmatched, support a range of program analyses. I will use Table 3, which combines the populations and attributes discussed above, to consider how well program data and household survey data, each used independently, satisfy the data requirements for distributional analysis, behavioral analysis, and eligibility modeling. The fundamental requirement involves having a comprehensive set of evaluative attributes for population subgroups defined in terms of fairly detailed programmatic criteria. That is, information on benefit receipt is not sufficient for purposes of program analysis. One approach is to link evaluative information available in surveys to detailed programmatic information for the same individuals.

On the face of it, data linkage might seem an obvious solution. However, data linkage (in this case, exact matching) is problematic for two reasons. First, it is useful only for those who have applied for benefits; it would tell us nothing about the program eligibility of those who have not applied. Second, matching is useful for some analyses of the existing program, but would not allow analysis of many proposed alternatives. Eligibility simulation offers a more flexible approach. Because it involves a representation of the eligibility process itself, eligibility models can be used to determine eligibility or incentive effects for those who have not applied or the effects of benefit structures which have not been enacted. For such analyses, data--whether matched or unmatched--must be supplemented with models that estimate eligibility and benefits.

Table 3 allows us to reflect on the data requirements for analysis of SSI (aged) by considering program data and household survey data in turn. Each can be considered in terms of attributes and populations. Program data, represented by the area outlined with the broken line, include only participants and only the attributes needed to determine categorical and financial eligibility.⁹ However, both distributional analysis and behavioral analysis require data that: (1) juxtapose program information with evaluative variables for the same individuals and (2) include nonparticipants (as a control group for behavioral studies or for simulating eligibility alternatives). These factors limit the analytic potential of program data when used alone, although such data remain useful for describing program trends and predicting costs.

On the other hand, household survey data, unmatched, have considerable potential for analysis of means-tested programs. By virtue of the mandate under which much survey data are collected, they include evaluative variables basic to both distributional and behavioral analysis and they collect data on nonparticipants as well as participants. Taken alone, however, the program information included is too limited; it does not indicate, for example, how participants' benefits were affected by specific program provisions. Fortunately, in the case of means-tested programs, such program details can be simulated using eligibility models. The large shaded area in table 3 illustrates, then, that eligibility models can be used to define key programmatic populations and that the survey offers comprehensive evaluative variables. As a result, an unmatched household survey, such as the SIPP, can support a broad range of program analyses--including distributional analysis, behavioral analysis, and eligibility modeling--for most means-tested programs.

It is important to understand why means-tested programs can be simulated using household surveys. Both household surveys and means-tested programs are instruments of public policy

⁹ In table 3, as well as in other tables used below, I discuss the analytical potential of data sets, used in conjunction with models, by illustrating to what extent the data set "fills" the attribute/population space shown on the table. In these tables, I show that the most useful data sets (1) allow estimates for all program-related populations shown and, (2) for members of each such population, include not only detailed program eligibility variables, but also a comprehensive set of evaluative variables.

Table 3
Supplemental Security Income (Aged):
Household Survey Data with Benefit Simulation

| <u>Attributes</u> | <u>Populations</u> | |
|--|--|-------------------------|
| | Program Participants | Nonparticipants |
| | Eligible | Ineligible |
| Program Eligibility Variables | | |
| <i>Categorical Eligibility</i> (age) | (Eligibility Simulated) | (Eligibility Simulated) |
| <i>Financial Eligibility</i> (income, assets, family size home ownership . . .) | | |
| Evaluative Variables | Household Survey Data (with Benefit Simulation) | |
| <i>Well Being</i> (income, assets . . .) | | |
| <i>Demographic Traits</i> (age, race, ethnicity, education, health . . .) | | |
| <i>Household Composition, Events</i> (marital status, loss of spouse, job loss, loss of health ins. . . .) | | |

and, in some respects, they are focussed similarly. Household surveys collect information on the financial well-being of demographic groups, among other things. Means-tested programs are typically targeted to specific demographic groups with low income and low assets. And, both the survey and the program relate to the *current period* (typically, in a well designed survey, the survey reference period subsumes the program accounting period). These coinciding features allow the program simulation.

4. Special Case One: Social Security and Matched Survey Data

In section 3, I suggested that the data requirements for basic microdata analysis of means-tested programs are substantially met by household survey data if those data are used in conjunction with an eligibility simulation model. That discussion also serves as a benchmark for considering why household survey data, taken alone, do not support basic program analysis for the nation's largest cash benefit program, social security.¹⁰ But before considering this in more detail, let me discuss the general criteria underlying the social security benefit structure, that is, how the program is targeted.

Dual Targeting of Social Security Benefits

The report of the 1979 Advisory Council on Social Security included a statement of the principles underlying the OASDI benefit structure:

From its beginning those responsible for the design of social security have sought to assure, on the one hand, a reasonable relationship between the social security taxes paid by individuals and the benefits they receive and, on the other hand, at least a minimally adequate income for long-term low-wage workers. Maintaining a reasonable relationship between taxes and benefits has been described as the goal of individual equity. Assuring a basic level of

¹⁰ The discussion in this section relates to the social security retirement (OASI) program and to financial eligibility for the disability insurance program. A separate problem unique to the major disability programs, DI and SSI, is discussed in the next section.

income has been called the goal of adequacy.¹¹

While covered earnings (and resulting contributions) represent the major determinant of benefit amount, a number of provisions are motivated by the adequacy goal. Major adequacy provisions include the "progressive" structure of worker benefits and dependent benefits. Other smaller-scale provisions include the windfall benefit provision limiting benefits for some federal workers, the minimum benefit (now repealed), and a special benefit for those over the age of 85 (occasionally proposed). Because of the scale of the OASDI program, some adequacy-related provisions have major effects. For example, over twelve million persons receive dependent benefits as spouses, children, or widows.¹²

Adequacy provisions under the structure of social security benefits are not explicitly targeted so as to take into account family income, as are provisions of a means-tested program. Some provisions are designed to provide additional benefits to those with low levels of covered wages or low benefits. Examples include progressive worker benefits, the repealed minimum benefit, and the proposed benefit for those over age 85. Analogously, family composition provides the targeting mechanism for dependent benefits. For adequacy provisions, then, low wages, low benefits, and family composition can be considered "proxies" for low levels of financial resources. And if they serve as proxies, then such provisions are target efficient to the extent that a broad, stable relationship exists between, for example, low covered wages and low levels of all financial resources. However, the denial of windfall benefits to insured federal workers serves as a reminder that targeting mechanisms such as low wages should be subjected to scrutiny. Matching administrative and survey data allows such scrutiny.

Let us consider Table 4. In terms of the variables needed, financial eligibility and benefits are determined differently for a contributory social insurance program than for the means tested program considered in section 3. Financial eligibility and benefits are determined based on such factors as the period working in a job covered under social security, earnings

¹¹ See page 55, Advisory Council on Social Security (1979).

¹² See Table 5.A1, Social Security Administration (1995).

Table 4
Social Security Benefits:
Program Data versus Household Survey Data

| <u>Attributes</u> | <u>Populations</u> | |
|---|------------------------------|-----------------|
| | Program Participants | Nonparticipants |
| <p><i>Program Eligibility Variables</i> <i>Categorical Eligibility</i> (age) <i>Financial Eligibility</i> (earnings history, quarters covered, family comp., marital history . . .)</p> | Program Data | |
| <p><i>Evaluative Variables</i> <i>Well Being</i> (income, assets . . .) <i>Demographic Traits</i> (age, race, ethnicity, education, health . . .) <i>Household Events, Composition</i> (marital status, loss of spouse, job loss, loss of health ins. . . .)</p> | Household Survey Data | |

histories, marital history, age at retirement, and current family composition. And, in terms of evaluative variables, in order to deal with dual targeting, analysts require information on *both*: (1) family income and other financial resources as well as (2) past earnings and contributions.¹³ In addition to an estimate of potential benefits, behavioral analysis--for example, retirement modeling--requires information on variables such as household composition and recent household events.

Table 4 illustrates why, in the case of the social security program, *neither* program data nor household surveys, taken alone, support some of the most basic distributional and behavioral analyses. As suggested by the smaller block, program data include detail on benefits for those receiving benefits; however, they lack the evaluative variables needed to undertake distributional or behavioral analysis. Social security program data do not include information on total family income, assets, or recent household events, for example.

Household surveys, however, cannot support social security analysis as they can for means-tested programs. The impediment is that household surveys do not collect the information needed to simulate financial eligibility and benefits for social security, namely, earnings histories and the period working in a job covered by social security.¹⁴ This limitation affects social security analysis in fundamental ways. Retirement modeling requiring an estimate of the retirement benefit for all eligibles cannot be undertaken. In terms of distributional analysis, characteristics can be tabulated for participants and nonparticipants, but not for more detailed beneficiary groups, such as workers with low lifetime wages or those receiving only

¹³ Earnings, then have more than one role in evaluating a contributory program. On the one hand, earnings histories allow simulation of a social security benefit. On the other hand, analysts and policy makers will want to consider how specific provisions or proposals affect long term, low wage workers, as opposed to high wage workers.

¹⁴ While some of these items could be collected in household surveys, it seems problematic to collect information such as the period working under social security or earnings histories, due to recall constraints. This also seems wasteful, in view of the fact that such information has already been collected administratively.

spouse benefits. Nor can analysts simulate alternative benefit structures.¹⁵

Table 5 illustrates what many analysts of social security take to be the solution--matching program data on earnings to household survey data. Assuming the survey has information on family composition and marital history, benefits can then be simulated for survey sample members. As the large shaded area illustrates, matching earnings histories and simulating benefits fulfill the critical data requirement: On the one hand we can simulate eligibility and benefits for sample members and, on the other hand, the survey provides evaluative attributes.

5. Special Case Two: SSA's Disability Programs

The Problem of Categorical Eligibility

Eligibility for major public programs involves two types of criteria, *categorical* and *financial*. As discussed above, for the social security retirement (OASI) program the obstacle to distributional and behavioral analysis involves the inability to simulate *financial* eligibility and benefits without using matched data. The problem relating to financial eligibility also exists for one of the disability programs--Disability Insurance--because the program is financed through contributions from past earnings. However, in this section I consider *categorical* eligibility, which poses a problem unique to the disability programs, DI and SSI (disabled). This problem, as addressed in recent SSA research, has involved the use of matched data other than information on earnings histories.

¹⁵ A small number of program provisions can be simulated using survey information alone, such as how changes in cost of living adjustments might affect poverty. The key is that in a few such cases, simulated benefits can be estimated without information on lifetime earnings, by using benefit information reported in the surveys. For other issues requiring information on lifetime earnings, such information can be added through statistical matching or by imputing earnings streams. For most purposes, however, exact matching of the observed earnings is clearly preferred, especially when distributional analysis is needed.

Table 5

**Social Security Benefits:
Matched Survey Data with Benefit Simulation**

| <u>Attributes</u> | <u>Populations</u> | | |
|---|--|-------------------------|-------------------------|
| | Program Participants | Nonparticipants | |
| | | Eligible | Ineligible |
| <i>Program Eligibility Variables</i> | Program Data | (Eligibility Simulated) | (Eligibility Simulated) |
| <i>Categorical Eligibility</i> (age) | | | |
| <i>Financial Eligibility</i> (earnings history, quarters covered, family comp., marital history . . .) | | | |
| <i>Evaluative Variables</i> | Matched Survey Data (with Benefit Simulation) | | |
| <i>Well Being</i> (income, assets . . .) | | | |
| <i>Contributions</i> (earnings history, taxes) | | | |
| <i>Demographic Traits</i> (age, race, ethnicity, education, health . . .) | | | |
| <i>Household Events, Composition</i> (marital status, loss of spouse, job loss, loss of health ins. . . .) | | | |

Disability determination, which determines categorical eligibility, serves a gatekeeping function for both the DI and SSI programs, distinguishing allowances from denials. With the total number of applicants for the two programs now having reached 2.5 million per year, its budgetary and income distributional effects are undeniable. However, using survey information to estimate those categorically eligible--those the Social Security Administration would consider disabled--has proven especially problematic with respect to disability programs. Other programs--those for the aged or for single parents, for example--use criteria relating to easily observed traits and, as a consequence, survey responses on such traits permit reasonable estimates of those categorically eligible.

But, how to use survey responses on health and activity limitations to represent program disability criteria is by no means self evident. Judgmental factors are involved in two respects. First, the critical survey responses are judgmental. When surveys ask respondents about their health and the extent to which impairments limit activities, the responses are self-evaluative or self-rated. Under one hypothesis, for example, of those with a given impairment, persons with a low tolerance for pain or a weak attachment to the labor force report more severe limitations. Second, the criteria used by SSA to determine disability status are complex and also involve judgmental elements. For example, early in the complex determination process, applicants are denied if their impairments are considered nonsevere. Also, at a later step, there is an evaluation to determine whether the applicant has the residual capacity to perform substantial work, after taking into account the applicant's impairment, age, education, and past work. Hence, both the program criteria and the responses to health questions in household surveys involve judgmental components.

It is not surprising, then, that survey estimates of the size of the disabled population cover a wide range. Table 6, which summarizes estimates from a recent Bureau of the Census report, illustrates this point.¹⁶ The five estimates, each using conventional definitions, cover a disturbingly large range—from 3.4 million to 29.5 million working age adults. By comparison, the number of persons receiving disability benefits under DI and SSI during the

¹⁶ See McNeil, 1993.

Table 6.—Selected estimates of the number and percentage of the working age population¹ with disabilities² and the number of adults receiving Social Security or Supplemental Security Income benefits based on their own disability, late 1991

| Disability Definitions | Persons (in thousands) | Percent of total |
|--|---------------------------|---------------------|
| Total persons..... | 165,040 | 100.0 |
| With a disability ³ | 29,482 | 17.9 |
| With a severe disability ⁴ | 13,171 | 8.0 |
| With a work disability and prevented from working ⁵ | 7,588 | 4.6 |
| With one or more severe functional limitations..... | 6,596 | 4.0 |
| Has difficulty performing one or more activities of daily living..... | 3,442 | 2.1 |
| Total receiving DI or SSI disability benefits..... | 5,702 | 3.5 |
| DI beneficiaries ⁶ | 3,878 | 2.3 |
| SSI recipients ⁷ | 1,824 | 1.1 |

¹Aged 15-64 except as noted.

²As estimated from the 1990 and 1991 panels of the Survey of Income and Program Participation. Interviews took place from October 1991 through January 1992.

³Persons identified as: (1) having a work disability or housework disability; (2) having difficulty in performing one or more of six functional activities, six activities of daily living (ADLs), five instrumental activities of daily living (IADLs); (3) using a wheelchair, crutches, a cane, or walker; (4) having one of five classes of impairments (a learning disability such as dyslexia, mental retardation, other developmental disability such as autism or cerebral palsy, Alzheimer's disease/senility/dementia, or any other mental or emotional condition); or (5) a nonaged beneficiary of the Medicare or SSI program.

⁴A person identified as: (1) prevented from working at a job or business or from doing work around the house; (2) unable to perform 1 of 17 functional, ADL, or IADL activities; (3) having one of three classes of impairments (mental retardation, a developmental disability, or Alzheimer's disease/senility/dementia); or (4) using a wheelchair or a long-term user of crutches, a cane, or walker.

⁵Aged 16-64.

⁶Number of Social Security beneficiaries, aged 18-64, receiving benefits as of December 1991 based on their own disabilities, for example, disabled workers, disabled widows and widowers, and disabled adult children.

⁷Number of adults aged 18-64 receiving a federally administered payment but not also receiving DI benefits as of December, 1991.

Source: John McNeil, "Americans with Disabilities: 1991-92," U. S. Bureau of the Census, *Current Population Reports, Series P-70-33*, U.S. Government Printing Office, Washington, DC, 1993. Beneficiary and recipient estimates based on SSA administrative data.

period was about 5.7 million. Hence, depending on the definition of disability selected, the number of working age disabled can be up to five times the number of beneficiaries.

One Approach: Estimating a Model of Disability Determination with Matched Data

Although how survey responses relate to disability determinations is not self evident, it is a relationship that can be modeled statistically provided the necessary data are available. In the end, efforts to use household survey data to study disability programs depend on a central methodological question: How, if at all, can survey responses relating to health, demographic factors, activity limitations, and work be used to identify those who would be considered disabled under SSA criteria?

Recent research at the Social Security Administration has addressed this issue using SIPP data exact-matched to SSA records on disability determinations.¹⁷ This study models the outcomes of SSA disability determinations for adult applicants, using as explanatory variables survey responses on health, functional limitations, demographic traits, and work experience. These variables are shown to be systematically related to the determinations of SSA adjudicators and the relationships estimated are plausible to those familiar with the disability determination process. This effort will support estimation of the number of persons in the general population eligible for the disability programs. Beyond that, it will allow development of a comprehensive model involving both government decisions (that is, eligibility determination) and individual behavior (that is, applications decisions), to explain program growth. This approach will allow us to distinguish the effects of categorical criteria, financial criteria, and the application incentives faced by households.

Table 7 deals with the data and modeling requirements for program analysis of the DI program. DI is doubly challenging in that it poses obstacles in terms of both the categorical

¹⁷ See Lahiri, Vaughan, and Wixon (1995). All matching activities for this study were carried out as part of a joint SSA-Bureau of the Census statistical project under the aegis of the agencies' Memorandums on the Exchange of Statistical Information and Service. All work involving the development and analysis of the matched data set at SSA was carried out by SSA employees (or on-site contractors) acting as special sworn agents of the Bureau of the Census.

Table 7

**Disability Insurance Benefits: Matched Survey Data,
Determination Model, and Benefit Simulation**

| <u>Attributes</u> | <u>Populations</u> | |
|---|--|--------------------------|
| | Program Participants | Nonparticipants |
| | | Eligible Ineligible |
| <i>Program Eligibility Variables</i> | Program Data | (Eligibility Simulated) |
| <i>Categorical Eligibility</i> (health, activity limitations, . . .) | | (Eligibility Simulated) |
| <i>Financial Eligibility</i> (earnings history, quarters covered, family comp., marital history . . .) | | |
| <i>Evaluative Variables</i> | Matched Survey Data (with Disability Determination Model, Benefit Simulation) | |
| <i>Well Being</i> (income, assets . . .) | | |
| <i>Contributions</i> (earnings history, taxes) | | |
| <i>Demographic Traits</i> (age, race, ethnicity, education, health . . .) | | |
| <i>Household Events, Composition</i> (marital status, loss of spouse, job loss, loss of health ins. . . .) | | |

and financial criteria for eligibility, necessitating matching of both disability determination data and earnings history data. The problem of categorical eligibility, discussed in this section, also applies to the disability portion of the SSI program.

Neither program data nor household surveys--when used in stand alone mode--support a basic course of program analysis for the disability programs. Program data provide a wealth of detail on the impairment, SSA's determination and the criteria for it, earnings, the period working in a job covered by social security and so on. But, once again, such data do not allow analysis of the program's distributional effects in terms of comprehensive measures of well-being or in terms of a broad set of demographic traits. Nor do program data include information on, for example, recent household events, information needed for behavioral analysis. And, while household surveys have a full array of attributes, they lack necessary program detail. Without data matching and eligibility modeling, household surveys allow only analysis exploiting the most basic program category--whether the survey respondent receives benefits or not.

The administrative data on disability determinations serve two functions in this modeling effort. First, they permit identification of survey sample members who applied for benefits during the period proximate to the survey, including those denied as well as those allowed. Second, they enable separate analyses of applicants allowed or denied under different decision criteria. For example, some applicants are allowed because their impairments met SSA's medical criteria (called the medical listings), while for others the decision takes into account not only medical criteria, but age, education, work experience, and an assessment of the applicant's capacity for work. Combined, these two alternative criteria provide the basis for the allowance decisions of the eight million persons currently receiving benefits. The administrative data permit distributional and behavioral analyses of each group separately, whereas survey information would permit only analyses of "composite" effects of the two distinct criteria (that is, analysis of all disabled beneficiaries together). This illustrates how matching allows analysts to focus on specific normative criteria underlying eligibility and benefit decisions. Moreover, the same analytical opportunities would not be available by designing better surveys, since survey respondents would not know under which of the two

criteria they receive benefits.

6. Conclusions

There is a data requirement central to the analysis of all public programs considered here: What is needed to support microanalytic study is a data set which combines evaluative variables (such as measures of well being, demographic characteristics, household composition, and household events) with detailed programmatic information (such as program eligibility, benefit size, and how an individual's eligibility or benefit is affected by a specific program provision). Juxtaposing such attributes for individuals represented in the data set is basic to analysis of how the program is targeted and the incentives it poses.

Juxtaposing these two types of attributes, however, is problematic. Neither survey data nor program data, taken alone, offer both. Moreover, these limitations are not accidental--they reflect basic features of administrative data collection and survey data collection. While what I have called evaluative variables are frequently collected in surveys, few are collected in the course of program administration, since, by statute and regulation, most such variables have no role in determining eligibility or benefits. In addition, administrative data have detailed program information, but only for participants. Conversely, while surveys collect a more comprehensive set of evaluative variables, in terms of program information, they often ask about benefit receipt and size only. More detailed program information (such as the eligibility of nonapplicants or whether affected by specific benefit criteria) are likely uncollectable using standard household surveys.

Fortunately, a reliable means of replicating government decisions on eligibility and benefits has been in use for over two decades. Eligibility models, sometimes referred to as microsimulation models, can be used to simulate eligibility and benefit decisions, often using surveys which have collected the necessary elements. Such models estimate both categorical eligibility (such as advanced age or single parent status) and financial eligibility. Because such models can be used to add program information to the attributes collected in surveys,

they represent a linchpin for program analysts. That is, such surveys can only be fully exploited for distributional and behavioral analysis if program eligibility can be simulated. This implies, in turn, that basic program analysis using household surveys depends on whether the survey, matched or unmatched, meets the data requirements for eligibility modeling.

I have illustrated that for the SSI (aged) program--considered as an example of means-tested programs generally--eligibility and benefits can be simulated using information collected in household surveys having comprehensive measures of financial resources. And, as a consequence, household surveys, without administrative data matching, can support modeling of program participation decisions and simulation of many alternative benefit structures, for example. This success story follows from a coincidence between principal features of means-tested programs and household surveys. Means-tested programs use family financial status in the current period as a basic criterion, while for cross sectional surveys, such as the SIPP, the financial well being of the family just prior to the survey represents a primary focus. As a result, such surveys include the elements of eligibility for most means-tested programs.

But household surveys do not confer equivalent analytical opportunities for all public programs. Contributory programs have particularly vexing data requirements. In the case of social security, the obstacle is that surveys do not include lifetime earnings, which are needed for benefit simulation. This precludes much retirement modeling. It also prevents distributional analysis, such as evaluating program provisions having adequacy objectives in the light of the family income of those affected. Matching of earnings data from administrative sources allows simulation of the social security benefit, enabling much basic analysis of program targeting and applications behavior. By comparison, analysts of means-tested programs can undertake an analogous course of study using unmatched household surveys.

Finally, there is a separate problem unique to SSA's disability programs. On the basis of information collected in surveys, how can analysts estimate who would be found disabled under the criteria used by SSA? The inability to estimate those eligible in terms of categorical (nonfinancial) criteria has limited the use of household data in analyzing program

growth. One solution is presented here. SSA analysts have matched administrative information on disability determinations to household survey data; as a result, the outcome of SSA determinations can be modeled using survey responses on health, activity limitations, demographic traits, and work behavior. This approach will allow analysis of both government decisions (that is, eligibility determination) and individual behavior (that is, applications decisions) in a framework that distinguishes the effects of categorical criteria, financial criteria, and the incentives faced by households.

References

- Advisory Council on Social Security (1979).
Social Security Financing and Benefits: Report of the 1979 Advisory Council on Social Security.
- Kilss, Beth and Frederick J. Scheuren (1978). *The 1973 CPS-IRS-SSA Exact Match Study*, **Social Security Bulletin**, 41:10:14-22.
- Lahiri, Kajal, Denton R. Vaughan, and Bernard Wixon (1995). *Modeling SSA's Sequential Disability Determination Process Using Matched SIPP Data*, **Social Security Bulletin**, 58:4:3-42.
- McClung, Nelson (1986).
The Art of Transfer Policy Analysis, in G. H. Orcutt, J. Merz, and H. Quinke, editors, **Microanalytic Simulation Models to Support Social and Financial Policy**. North-Holland.
- McNeil, John M. 1993. *Americans with Disabilities: 1991-92, Data from the Survey of Income and Program Participation*. U.S. Bureau of the Census, Current Population Reports, Series P-70, No. 33, Washington, DC: U.S. Government Printing Office, December.
- Projector, Dorothy S. and Ellen G. Murray (1978). *Eligibility for Welfare and Participation Rates, 1970*. **Studies in Income Distribution**, number 7.
- Social Security Administration. 1995. *Annual Statistical Supplement to the Social Security Bulletin*. U.S. Government Printing Office, Washington, DC.
- Vaughan, Denton R. and Bernard Wixon. 1991. "Two Papers on a New SIPP-Based Microsimulation Model of SSI and OASDI." ORS Working Paper Series, Number 54, Division of Economic Research, Office of Research and Statistics, Social Security Administration, Washington, DC.
- Wixon, Bernard, Benjamin Bridges, Jr. and David Pattison (1987).
Policy Analysis Through Microsimulation: The STATS Model, **Social Security Bulletin**, 50:12:4-12.

December 11, 1996

wixon-c:\usrfiles\wp\model\match5.ppr

November 12, 1996
M.V. Leonesio

DISCUSSANT'S REMARKS
Using Administrative Records for Statistical Purposes

Before getting into the substance of the two papers, let me say a little about my own work, which will indicate how sympathetic I am to the positions of the two authors. I am an economist at the Social Security Administration whose research concerns the interactions between Social Security's programs and the labor market. That is, how do American work and retirement patterns affect Social Security revenues and expenditures and, in turn, how do our programs influence work patterns and trends. Some of my research uses the matched SIPP files to which Bernie Wixon has referred. In addition, for several years I worked on a project with a handful of people at the Census Bureau and SSA to match our administrative data to a data set we haven't mentioned here: the National Longitudinal Survey of Women. Just this year we have finally created this matched file, which is the primary data resource being used in research we are conducting on the Social Security program's treatment of women.

Let me say a little more about this project to supplement the discussion in the two papers we've heard. In recent years a number of policy questions have arisen regarding women and the Social Security program. Just a few of these:

- How does Social Security treat women who make different choices regarding marriage and work? For example, a woman who works her entire life and pays substantial Social Security taxes could end up receiving smaller Social Security benefits than a married, full-time homemaker who pays little or no Social Security taxes and receives benefits on the basis of the earnings record of a high-earning husband. Over time have spousal and survivor benefits paid to women been increasingly paid to higher income families where wives can afford not to work outside the home?
- Although the poverty rate for persons aged 65 or older has fallen substantially in the past 3 decades, and is now lower than the rate for the population aged 18-64, the rate for older women -- particularly those who live alone -- remains high. Are there appropriate Social Security policy changes that might remedy

this?

- Do labor market trends in hours and earnings for women suggest that their lifetime earnings patterns will eventually closely resemble those of men? How will the type and size of women's benefits differ in the future? And, what are the consequences for the long term finances of the Social Security program?

When we began to think about a research program to address these types of questions, an attractive data source seemed to be the National Longitudinal Survey of Women. Back in 1967 the Bureau of Labor Statistics began interviewing two samples of women: the NLS Mature Women who were then aged 30-44, and the NLS Young Women who were then aged 13-21. These samples have been regularly interviewed ever since about their work experiences, education and training, family backgrounds, child rearing activities, health status, incomes, living expenses, saving and investment, retirement planning, and a host of other information relevant for explaining their eventual economic status in old age. Of course, we would like to see precisely how these women do and will interact with the Social Security program. The Mature Women are now in their 60s and 70s and for the most part have reached their retirement years. The Young Women are in their 40s and we can begin to see how their generation will bring quite different earnings and marital histories into their retirement years.

Just this year we have linked the Social Security Administration's record data on earnings and benefits to the NLS survey data for these two groups of women (and their husbands) which enables us to explore a wide range of policy-relevant research questions. For example, what factors explain the incidence of poverty for older women? Are younger women preparing for retirement any better than the previous generation? How are the type and amount of Social Security benefits likely to change as successive cohorts of women retire? Are the determinants of the timing of retirement for women the same as for men, who have been much more extensively studied? Most of these types of questions cannot be pursued

without survey data linked with the Agency's administrative data. On their own, surveys -- even those as comprehensive as the NLS Women -- do not have the accurate Social Security covered earnings information needed to compute the basis and amount of a woman's retirement benefit. And, the Agency's administrative data lack nearly all the information that social scientists require to explain behavior. That is, what explains a woman's lifetime earnings pattern, why did she apply for benefits when she did -- or when is she likely to -- and how important is her Social Security benefit to her economic wellbeing? A matched survey file of this type offers enormous potential for policy research.

Now, on to the two papers.

Specific remarks on Ron Prevost's paper:

Ron Prevost has laid out many of the important questions that will have to be addressed if a variety of administrative record files are to be used as key inputs into the decennial census planning and information gathering. Our experience at SSA with trying to use our own administrative files suggests that Ron's research agenda will take a small army to implement. Administrative data are often ambiguous, inaccurate, awkwardly managed, and poorly documented -- if documented at all. Keep in mind that these shortcomings do not necessarily impede an agency from carrying out its basic administrative functions effectively. Rather, they simply demonstrate that recordkeeping procedures that might work for program administrators can impose thorny problems for the researcher wanting to use these data. It will no doubt be difficult to access, interpret, and merge information from many different administrative sources.

These concerns based on the experience of my research colleagues at Social Security lead me to pose this question. Keep in mind that I'm an economist, not a survey statistician. My understanding is that survey methodologies are now so sophisticated that sampling

techniques allow one to get a more accurate estimate of population characteristics than efforts to count the population in an exhaustive manner, as in the decennial census. If the decennial census has problems due to escalating costs and an uncooperative public, is greater reliance on administrative data from many sources the most cost-effective solution? Why this option as opposed to a greater reliance on sample surveys? I would be curious to hear the answer.

Specific remarks on Bernie Wixon's paper:

I am well acquainted with the disability research project that serves as a basis for many of the ideas in my colleague Bernie Wixon's paper. This is an excellent example of how administrative data linked to survey information can open up a set of questions for investigation that were otherwise impossible to address.

In many instances an Agency's administrative data will enable a researcher to create a statistical description of specific elements of an Agency's operations that cannot be accomplished via any alternative means. But, matched data can also be used to improve the scope and quality of behavioral research, as Bernie's paper has emphasized. Think of what a researcher can learn from an Agency's administrative files. Often we can recover key data elements from administrative forms that were processed. We can review applications, administrative decisions, outcomes, payments, and the like. This is all valuable stuff and can be illuminating in its own right. But, if we are interested in questions such as: Why did this person submit an application, and why then? Why did other individuals with ostensibly similar circumstances not apply? Will particular changes in our program rules redistribute benefits to financially needy families? Would that new policy cause people to behave differently? Administrative files rarely contain the necessary information. In contrast, what can a researcher learn from household surveys such as the CPS, SIPP, and PSID files, but without access to an agency's administrative files? Often there's plenty of background information on

the demographic and socioeconomic characteristics of individuals and their families. But, from the perspective of policy research and analysis, there is usually insufficient detail about how the respondents interacted with a specific government program. When survey and administrative data are matched, the result is an information base that can support considerably improved research projects that assess the effectiveness of current and proposed government programs.

Although there are clearly grounds for optimism on this score, even with the creation of high quality matched survey data sets, the research problems are formidable. The centerpiece of behavioral social science research is usually the statistical estimation of some model intended to represent and explain some behavioral phenomenon. So, one might hypothesize a simple model with general functional form:

$$y = f(X, \beta, \epsilon)$$

where y is the behavior to be explained or predicted, X is a set of explanatory factors, β is a vector of coefficients or parameters to be estimated, and ϵ is an error term that specifies the relationship's stochastic properties. The objective is to estimate the true values of β by applying appropriate statistical methods to our data.

The list of the main things that can go wrong in this process is collectively known as *specification error*. Specification error can be of four basic types:

1. Incorrect functional form. The researcher usually has to make decisions and choices about $f(\cdot)$ and it is very easy to get things wrong. Social science theories -- even economic ones -- have little specific to offer on this count. In the large number of cases where we really don't have much of a theory at all, it is very easy to wander far astray at this point.

2. Incorrect set of explanatory variables. These are of two types of mistakes: including variables that don't belong and omitting ones that do. It is easy to include

factors that seem plausible but simply don't belong, thereby imposing inefficiency in the estimation. Even worse is the case of leaving out explanatory variables that belong in the model, thereby suffering from inefficient estimation as well as admitting bias into the estimation of the model's parameters. The main reasons for omitting variables that belong in our model are flawed theories and data sets that simply have not collected the relevant information -- in some cases because the phenomena are unobserved or unmeasurable (*e.g.*, motivation, tolerance for pain).

3. Errors in variables. The explanatory variables are measured with error. For example, we might think that income is involved in our story, but the survey has measured income inaccurately.

4. Incorrect specification of the model's error term. Empirical researchers can be remarkable cavalier about this aspect of model development, leaving the stochastic specification as a maintained hypothesis. The ease of using today's canned statistical packages makes this error extremely easy to commit.

All of these forms of specification error potentially invalidate the model results and their interpretation. So, what contribution do matched surveys make to this process? They help potentially with the second and third types of problems. Sometimes the addition of administrative data allows a researcher to include information in the model that is available from no other source. An example might be the absence of a worker's earnings history in a survey such as SIPP, that can be obtained from SSA's administrative records. Or, sometimes we have imprecise measurement of certain explanatory factors that can be improved by consulting an Agency's records. For example, during what time interval in the past did an individual receive a particular type of government transfer and in what amount? This is the type of information that is frequently inaccurately reported in surveys.

In any event, the availability of matched data offers the promise of improving the quality of policy research on behavioral issues. In light of the daunting problems that credible research of this type must surmount, it falls considerably short of a panacea.

Session 12

VALIDATION OF COGNITIVE QUESTIONNAIRE PRETESTING METHODS

Evaluating the Generalizability of Cognitive Interview Findings

Paul Beatty, Gordon Willis, and Susan Schechter
National Center for Health Statistics

I. Introduction

One purpose of cognitive interviewing is to confirm that terms are understood consistently; therefore, if we were to conduct a cognitive interview about cognitive interviewing, we might well ask "what does the term *cognitive interview* mean to you?" Since the term may take on different meanings to different people in different organizations, it is probably sensible to begin by deciding upon a working definition of the term. For our purposes, cognitive interviewing entails asking a participant to either "think out loud" while answering survey questions, or respond to probes about question interpretation and thought processes, or both. Survey researchers have generally accepted this technique as a legitimate, valid, effective means to quickly identify and correct questionnaire problems (Jobe and Mingay, 1991; Willis, Royston and Bercini, 1991; DeMaio and Rothgeb, 1996).

Is all the faith we have placed in the method justified? Based on anecdotal evidence, it seems to be. There is no shortage of examples of "bad" survey questions that were identified and improved based on cognitive interview findings. However, there have not been many systematic attempts to demonstrate that cognitive interviewing finds legitimate survey problems. It seems reasonable to ask: does cognitive interviewing find valid results that generalize to field settings?

That question needs to be answered in several steps. First, it is important to ask more basic questions about the goals of cognitive interviews-- what do we expect them to accomplish? After answering that, we can evaluate how well cognitive interviewing meets those goals. We will do that through reviewing some of our recent studies along those lines; in addition, we will outline our plans for continued research.

II. Addressing the problem of small and non-representative samples

One of the most common criticisms regarding cognitive interviewing concerns sample size and representativeness: how can one infer anything from a dozen interviews from a convenience sample? This criticism is closely related to our question about the goals of cognitive interviewing.

First, we should note that this lack of representativeness is very much by design. The idea is to select participants from particular age groups, those with certain health conditions, or whatever characteristics interest us the most. If, for example, we were interested in people over 65 with asthma, we could recruit

them specifically. A random sample would be an inefficient approach to finding them-- in fact, several hundred respondents from the general population might tell us less than a dozen selected specifically from a target group of interest. Furthermore, questions nestled into complex skip patterns might rarely or never be administered during the test.

This efficiency is certainly important-- but how can we find a representative sample of cognitive problems that respondents will have when answering these questions, without a representative sample of the people who will be answering them? It is important to note that we are not claiming that we do this. Rather, we are proposing that cognitive interviewing does something much more modest-- provide clues regarding potential sources of survey error. In other words, rather than claiming that we found something that "will be a problem" when a survey is fielded, we would claim that it *might* be a problem, based on interactions with relevant survey participants.

There is also the issue of recognizing the difference between a legitimate questionnaire problem and an "odd case"-- which could be particularly difficult if we only interview a small sample of people. But actually, logic can usually distinguish the odd cases from likely problems. Cognitive interviews usually suggest not only what the problem is, but what aspect of the question creates the problem.

As an example, consider this question that was recently tested in our cognitive laboratory:

During the past year, on average, on how many days did you drink alcoholic beverages, that is beer, wine, or liquor?

- _____ days
- a. per week
 - b. per month
 - c. per year

One laboratory subject expressed confusion. Probing during the cognitive interview revealed the source of this confusion: the question asks for a "number of days in the past year" and also an "average." It would make sense to ask the average number of days in a typical year-- or, it would make sense to ask about the number of days in the last year, dropping "average"-- but as it is, the question asks for both a "one year total" estimate, and an average over an unspecified time period. Apparently for this reason, our subject asked "do you want days last year, or what?"

This example has three vital characteristics: (1) the clue of a potential problem, (2) a reasonable explanation for the source of the problem, and (3) possible solutions. The identification of this potential problem is valuable because it is logical that respondents could stumble on this problem, and it can be avoided. Whether we discover this with one subject or fifty, the merit of the insight is really determined through a logical judgment. Thus,

the interview was not a mechanism for proof, but rather an idea generator about potential problems.

III. Evaluating the value of cognitive interviewing "clues"

Traditionally, the debate about cognitive interview validity has focused on whether or not they uncover the "true" cognitive processes of respondents (Nisbett and Wilson, 1977). That is, are respondents capable of telling us how they figure out their answers? When they think out loud or respond to probes, are they telling us what is really happening in their minds, or is it actually a re-creation of their thought processes, which is therefore less valid?

It may not be critical to answer this question at this point. Participants provide us with clues that seem to have great value for discovering the sources of survey problems. The challenge for cognitive researchers is to demonstrate that these clues are actually useful to survey research, whether or not they reflect "true" cognitive processes. A more pressing concern is: what if the clues are wrong, or misleading, or otherwise steer us astray?

Until now, we have simply assumed that this is not the case. Researchers who conduct cognitive interviews have made several implicit assumptions about the value of cognitive interviewing clues. At face value, these assumptions seem reasonable, but they have generally not been challenged in a serious manner. The four major assumptions are as follows:

Assumption #1: Cognitive interviewing finds problems that will carry over to actual surveys.

In other words, the findings of cognitive interviews are not "artifacts" deriving from the method. These interviews, we assume, tell us something that has practical utility.

Assumption #2: The response process when answering questions in a cognitive laboratory is more or less the same as in a survey interview.

For example, question comprehension processes should be similar enough in a laboratory to a survey setting to be applicable. In other words, using laboratory findings is not comparing cognitive apples and oranges.

Assumption #3: Cognitive interviewer behavior does not have an undue effect on the content of the interview.

Some interviewer variation is inevitable, of course. We are simply assuming that cognitive interviewer behavior does not radically alter the way subjects answer survey questions, or affect the basic value of our findings.

Assumption #4: The cognitive interviewing process is basically reliable-- if repeated, it would yield similar results.

That is, if one group of cognitive interviewers identified problems with a particular questionnaire, a different group of interviewers should find compatible (though probably not identical) results.

Some may feel that these assumptions have been made too lightly. Our own research has attempted to explore their veracity, focusing in particular on the first and second assumptions. Two studies described below investigate the assumptions through distinctive approaches.

IV. Two studies on the generalizability of cognitive interview findings

STUDY ONE (Willis and Schechter, 1996)

Anecdotally, if one compares survey questions before and after a round of cognitive interviews, it often seems obvious that the new question is "better" than the previous version. But what about actual survey data? Can we show that changes from cognitive interviews have positively impacted actual survey data?

Consider the following survey question, designed to measure time spent performing strenuous physical activity:

On a typical day, how much time do you spend doing strenuous physical activities such as lifting, pushing, or pulling?
(hand card)

- a. None
- b. Less than 1 hour
- c. 1-4 hours
- d. 5 or more hours

When tested in a cognitive laboratory, many subjects selected the "1-4 hours" response. When they were probed, however, they often admitted that they worked in offices and performed typical office tasks-- not what we would define as strenuous.

The question seemed to produce a bias-- reporting "none" clearly makes one appear sedentary. Given the available response options, it was much more desirable to report some level of activity than absolutely none.

Our clue of a potential problem was the preponderance of "1-4 hours" responses, which disagreed with probe responses. Our explanation of this discrepancy is the undesirability of appearing to be completely sedentary. A possible solution, then, would be to provide respondents with a more socially desirable "out."

A first step was to draft a question that eliminated this problem. An alternative version was written with this additional screener question:

On a typical day do you spend any time doing strenuous activities such as lifting, pushing, or pulling? (Yes/No)

A "no" response counted as zero; only subjects who answered "yes" received the original frequency question. When we tested this new version, many subjects were perfectly willing to respond "no," sometimes adding comments such as "I work at a computer all day." The screener question may be an improvement because it presents a balanced choice of equally legitimate responses: some people do strenuous activities and others do not. The previous question implied a continuum ranging from sedentary to vigorously active. Respondents' desire not to appear at the low end of this continuum might have influenced their responses.

The next logical question is: does this new version actually make a difference in the field, improving the accuracy of statistics? To test that, both versions were administered in a split ballot-- one with the screener and one without. The following results were observed in a relatively small field pretest, and repeated in a study on the health of women of child-bearing age:

Table 1: Field Pretest Results: Versions Before and After Cognitive Interview Modifications

| Hours | <u>Test 1: NHIS Field Pretest</u> | | <u>Test 2: Women's Health Study</u> | |
|-------|-----------------------------------|--------------|-------------------------------------|--------------|
| | <u>Ver 1</u> | <u>Ver 2</u> | <u>Ver 1</u> | <u>Ver 2</u> |
| 0 | 32% | 72% | 4% | 49% |
| <1 | 32% | 18% | 42% | 16% |
| 1-4 | 35% | 10% | 50% | 27% |
| 5+ | 0% | 0% | 4% | 8% |
| | ---- | ---- | ---- | ---- |
| | n=37 | n=39 | n=93 | n=94 |

As predicted, the distributions of answers are quite different, with many more respondents falling into the "zero" category when a yes/no screener is used (Version 2, in both tests). We presume that the Version 2 responses are more accurate. We do not know that for certain, but given the apparent tendency to overestimate time spent performing strenuous activity, a good case can be made for this conclusion.

This process was repeated using other survey questions over several different split ballot experiments. The results generally matched these findings: hypotheses from cognitive interviews were borne out by field data. This suggests that cognitive interview findings were relevant and applicable to a field setting.

STUDY TWO (Beatty, Schechter, and Whitaker, 1996)

This study was a follow up to cognitive interviews about subjective health assessments. Questions were based on feelings in the last 30 days-- for example, "During the past 30 days, how many days has your physical health been not good?" The questions called for numeric responses between 0 and 30 days, but many subjects had difficulty providing them. Some provided general answers, such as "I feel that way a lot"; others objected to the premise of the question, arguing that "I can't put it in days."

It seemed clear that the questions had problems, since a large proportion of responses were not given in the expected format. However, the survey sponsors had administered these questions in the field with no reports of trouble from interviewers, and very low item nonresponse. Their alternative theory was that the conversational tone and frequent probing in cognitive interviews actually created the appearance of problems.

The purpose of our study was to examine the relationship between probing style and subjects' answers. Using transcripts of cognitive interviews, we first coded each subject's response to each survey question, or the statement that most clearly resembled a legitimate response.

Second, we developed a code for how closely this response conformed to the expected response format-- that is, a number between 0 and 30. We labeled this "precision," recording it on a scale from 0 to 3 as follows:

Code 0: The response was clear, requiring virtually no rounding, judgment, or interpretation from a coder. Example: "Four days."

Code 1: The response required minimal interpretation from a coder, such as a moderately qualified answer, or answers given in a narrow range. Examples: "Probably every day," "Six or eight days."

Code 2: The response required considerable interpretation from a coder, such as broad ranges. Examples: "Six to ten days," "More than 15 days."

Code 3: The response could not be coded in the expected format. Examples: "I can't put it in days," "For a while I was in horrible pain," etc.

Third, we coded the type of probes that preceded each response. We distinguished between "re-orienting" and "elaborating" probes. *Reorienting probes* encourage subjects to re-focus on answering the survey question, such as "So how many days out of 30 is that?" *Elaborating probes* are more typical of cognitive interviews, designed to get information beyond the answer to the survey question-- for example, "Tell me what you were thinking

about while answering" (which encourages the subject to discuss the answer).

Our analytic goal was to determine how probing style was related to response precision. We found that probing style had considerable influence. When re-orienting probes preceded responses, 24% of responses were "precise"; when elaborating probes preceded responses, only 5% of responses were precise. Similarly, the percentage of "uncodeable" responses changed considerably depending on probing style: 60% were uncodeable following elaborating probes, whereas only 27% were uncodeable following reorienting probes. These results appear in Table 2, below:

Table 2: Response precision, by types of probes preceding response

| Precision | Elaborating probes before response | Re-orienting probes before response |
|----------------|---------------------------------------|--|
| 0 (Precise) | 4.8% | 24.4% |
| 1 | 21.4% | 34.1% |
| 2 | 14.3% | 14.6% |
| 3 (Uncodeable) | 59.5% | 26.8% |
| | ----- | ----- |
| | n=42 | n=41 |

(Table excludes cases in which no probing preceded response. Because re-orienting probes and elaborating probes were used in 23 cases, columns are not mutually exclusive.)

Next, we conducted additional interviews, this time training interviewers to use only re-orienting probes. This was done to evaluate whether response imprecision could be reduced by curtailing interviewer behavior that led to increased discussion. Interviewers discussed the meaning of subjects' answers only at the end of the interview session, during a debriefing. A comparison of results from the first and second round of interviews appears below:

Table 3: Precision of responses, compared across interview rounds

| Precision | Round 1 | Round 2 |
|----------------|---------|---------|
| 0 (Precise) | 36.3% | 82.3% |
| 1 | 32.6% | 14.6% |
| 2 | 8.1% | 0.0% |
| 3 (Uncodeable) | 23.0% | 3.2% |
| | ----- | ----- |
| | n=135 | n=158 |

(Table includes all responses, whether preceded by probes or not.)

In the second round of interviews, 82% of responses were precise, and only 3% were uncodeable. At first, it might seem that the charge against cognitive interview findings was correct-- if one removes conversational probes, subjects' responses are much more straightforward. However, post-interview debriefings revealed that subjects still had many of the same misgivings about answering the questions that they had in the earlier cognitive interviews-- they were reluctant to answer in terms of days, or felt their answers were inaccurate. In the later round, however, interviewers denied subjects the opportunity to *express* uncertainty about their answers. If subjects tried to explain or qualify their responses, the interviewer asked them to respond numerically. Thus, we suggest that cognitive interviewing does not create the appearance of problems, but rather that conventional interviewing *suppresses* the expression of response difficulties.

The fact that some subjects deviate from question format in cognitive interviews, in and of itself, is not particularly illuminating-- interviewers ask them to do this. However, the amount of deviation from format, which varies across questions, may provide a useful measure of relative difficulty answering the questions. A greater desire to discuss the nuances of answers is probably informative. Nevertheless, analyses needs to be performed with sensitivity to the fact that a cognitive interview is quite different from a survey interview.

V. Future directions for empirical work

Several of the assumptions mentioned earlier-- regarding cognitive interviewer effects, and reliability of conclusions-- have not yet been addressed. We have initiated several studies that explore those assumptions, however, and expect to present data in the near future.

Staff at NCHS recently constructed a "methodological questionnaire" to serve as the basis for additional research. The questionnaire was constructed from drafts of questions from various surveys, but the methodological questionnaire will not actually be fielded. It will therefore be possible to explore hypotheses by maintaining complete control of questionnaire content, question wordings, and so on. ("Real survey" pressures often make it difficult to implement this type of methodological work).

NCHS staff conducted 40 cognitive interviews using this questionnaire, which will serve several purposes. First, cognitive interviewer behavior will be coded: we will explore how much interviewer behavior varies, and in what manner. As of this writing, it is too early to tell exactly how much individual interviewers' styles differ, but it is clear that there is a wide variety of activity during cognitive interviews. A preliminary taxonomy of cognitive interviewer behavior distinguishes between numerous types of probes (probes about thought-processes, question interpretation, question difficulty, and probes for information

beyond the scope of the survey question); types of feedback (feedback on subject performance, and feedback on content of responses); and other remarks (transitional statements, confirmation of subject responses, and so on).

In addition to coding what interviewers do, we will investigate what interviewers *conclude* about the nature and extent of questionnaire problems. An important component of reliability assessment is determining whether interviewers reach the same conclusions about problems in a particular questionnaire. Also, a contractor will conduct 60 cognitive interviews using the same questionnaire. That will enable comparison of how two independent groups go about evaluating a questionnaire, and comparison of the conclusions they reach.

Finally, the analysis will extend to other pretesting methods. Twenty questionnaire designers have provided "expert reviews," of the methodological questionnaire; also, field pretest interviews were behavior-coded (see Fowler and Cannell, 1996). Comparing the results of these appraisals should provide a sense for how the methods complement each other, rather than demonstrating which techniques are "best."

In summary, much work remains in investigating the generalizability of cognitive interview findings. However, we also have good preliminary indications that cognitive interviews are effective clue-finders that greatly help questionnaire designers perform their jobs. We look forward to sharing more results of our evaluations in the future.

REFERENCES

Beatty, P., Schechter, S., and Whitaker, K. (1996). "Evaluating Subjective Health Questions: Cognitive and Methodological Investigations." Proceedings of the Section on Survey Research Methods, American Statistical Association, in press.

DeMaio, T.J. and Rothgeb, J.M. (1996). "Cognitive Interviewing Techniques: In the Lab and in the Field." In Schwarz, N., and Sudman, S., eds., Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research. New York: Jossey-Bass.

Fowler, F.J., and Cannell, C.F. (1996). "Using Behavior Coding to Identify Cognitive Problems with Survey Questions." In Schwarz, N., and Sudman, S., eds., Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research. New York: Jossey-Bass.

Jobe, J.B., and Mingay, D.J. (1991). "Cognition and Survey Measurement: History and Overview." Applied Cognitive Psychology, 5, 175-192.

Nisbett, R.E., and Wilson, T.E. (1977). "Telling More Than We Can Know: Verbal Reports on Mental Processes." Psychological Review, 84, 231-259.

Willis, G.B., Royston, P.N., and Bercini, D. (1991). "The Use of Verbal Report Methods in the Applied Cognitive Laboratory." Applied Cognitive Psychology, 5, 251-267.

Willis, G., and Schechter, S. (1996, July). "Evaluation of Cognitive Interviewing Techniques: Do the Results Generalise to the Field? Paper presented at the Fourth International Sociological Association Conference on Social Science Methodology, Essex, England.

PREDICTING TEST-RETEST RELIABILITY FROM BEHAVIOR CODING¹

Jennifer C. Hess, Center for Survey Methods Research, Bureau of the Census,
Washington, DC 20233

Eleanor Singer, Survey Research Center, University of Michigan

¹ The authors wish to acknowledge the contribution of John Bushery, Richard McGuinness and their colleagues in the Demographic Statistical Methods Division of the Census Bureau who calculated the reliability data used in this paper.

Introduction

In attempting to move questionnaire design from art to science, researchers use different evaluation techniques to help determine how well questions are working. Techniques such as behavior coding, respondent debriefing, interviewer debriefing, cognitive interviewing, and nonresponse analysis all provide information to help the questionnaire designer assess whether respondents understand questions as intended and whether they are able to provide adequate answers to them. In 1994, Presser and Blair evaluated some of these methods, concluding that behavior coding provided more reliable diagnoses of question difficulties than conventional pretests involving a small number of interviewers followed by an interviewer debriefing.

However, with the possible exception of some types of respondent debriefing questions, these techniques do not actually measure question reliability. Reliability data, such as those that could be obtained in a test-retest experiment (reinterview), are rarely collected as part of pretest activities because they are time-consuming, labor intensive and very costly to collect. Of course, the goal of good questionnaire design is to produce reliable and valid information, not simply questions that are easy for respondents to answer. But it is assumed that questions that pass the screen of the questionnaire evaluation techniques described above are also more likely to produce data that are reliable and valid.

How well do question evaluation techniques in fact predict reliability and validity? Data reported by Belli and Lepkowski (1995) suggest that interviewer behaviors have little predictive value for response accuracy, though respondent behaviors are somewhat more predictive of response accuracy. Recently, the U.S. Department of Agriculture's Food and Consumer Service fielded a new survey, designed to measure the subjective experience of hunger in the United States. This survey provided an opportunity to examine how well some traditional question evaluation techniques predict test-retest reliability. The Census Bureau was asked to help develop the questionnaire, using some of the evaluation methods listed above. In addition, a reinterview was conducted with a sample of households following the survey. In this paper, we use behavior coding data to predict how reliably questions are answered, as measured by an index of inconsistency developed by the Census Bureau.

Methods

Sample

The Food Security Supplement to the Current Population Survey (CPS) was conducted from April 16-25, 1995 on a nationally representative sample of approximately 54,000 interviewed households. Respondents were asked both the CPS labor force questions and the Food Security Supplement questions. The response rate for the CPS was 92.9 percent and for the supplement was 85.4 percent. Approximately 90 percent of the cases were conducted in the field using computer assisted personal interviewing (includes both personal visit interviews and telephone interviews from field representatives' homes) and 10 percent were conducted at the Census Bureau's centralized telephone facilities using computer assisted telephone interviewing.

Approximately 34 percent of the households in the sample were "low income," which, for the purposes of this study, is defined as at or below 185 percent of the poverty level.² Three-quarters of the sample households were urban and one-quarter rural. Approximately 85 percent of the households were White, 10 percent were Black, and 6 percent were Hispanic (could be of any race).³

The questionnaire included five different sections: food expenditures, program participation, food sufficiency, coping mechanisms and food scarcity, and concern about food sufficiency.⁴ Food expenditures were asked of all households. These questions collect information on the actual amount the household spent for food last week and the usual amount the household spends on food per week. The program participation section asks about food stamp reciprocity and participation in other government and private programs that provide food, such as the school lunch program and WIC. The food sufficiency section contains questions used to assess whether respondents clearly have enough to eat or whether there are times when their resources are strained and they have difficulty providing themselves or their families with a nutritionally adequate diet. These questions are used to screen respondents either into or out of the remainder of the questionnaire. The coping mechanism and food scarcity section measures the extent of food insecurity in the household as do the questions in the section on concern about food sufficiency.

Behavior Coding

Behavior coding is the systematic coding of the interactions between an interviewer and a respondent (Cannell, Lawson, and Hausser, 1975; Cannell et al., 1989). Interviewers at the Census Bureau's Hagerstown and Tucson Telephone Centers tape recorded a total of 147 cases of which 136 were subsequently behavior coded. (Eleven cases were not used because permission to record the interview was not on the tape.) We used a quota sample for behavior coding, not a random sample. The telephone centers were instructed to tape record interviews with the first 75 low income households.

We coded the first exchange between the interviewer and the respondent for each question. Coders assigned one interviewer code and up to two respondent codes per question. (Two respondent codes were most often assigned when the respondent interrupts the question reading to provide an answer. Thus, one of the codes is a "break-in" and the other may be any of

²Our measure of "185 percent of poverty" in this survey is based on family size and family income. The measure, however, is rather imprecise, because the only measure of family income in the CPS is based on a single question about family income in the previous calendar year and is a categorical variable composed of income ranges.

³Race of the household is measured by the unweighted race of the reference person. The reference person is the first person listed on the household roster and is the name of the person or one of the persons who owns or rents the house/apartment.

⁴Contact the authors for a copy of the questionnaire.

the remaining respondent codes.) Four experienced coders from the Hagerstown Telephone Center behavior coded the tapes. (See Appendix A for a description of interviewer and respondent behavior codes.)

To assess coder reliability, each coder was asked to complete the same five cases (in addition to the regular workload). The coders averaged 87 percent agreement on interviewer codes, 92 percent agreement on at least one of the two respondent codes, and 83 percent agreement on both respondent codes. The kappa statistics, which take into account the probability that two coders will agree on a code by chance, ranged from .68 to .80 for between coder agreement on interviewer codes, .74 to .93 on at least one of the two respondent codes, and .55 to .84 on both respondent codes. Kappa values above .75 represent excellent agreement and values from .40 to .75 represent fair to good agreement beyond chance (Fleiss, 1981). Thus, our statistics indicate fair to excellent agreement between coders.

An evaluation of the supplement questionnaire based on behavior coding data indicated that the food expenditures section caused the most problems of any section (see Table 1). Eighty-three percent (N=18 questions) of the questions in this section were flagged as problematic by behavior coding. Approximately 60 percent of the questions in the food sufficiency section (N=10 questions) and the concern about food sufficiency section (N=6 questions) were problematic. The remaining two sections, the program participation section and coping mechanisms and food scarcity section, caused fewer problems. Twenty percent of the questions in the program participation section (N=10 questions) and 28 percent of the questions in the coping mechanisms and food scarcity section (N=36 questions) were problematic. However, 15 of the 36 questions in the latter had less than 7 responses. When these cases are excluded, the percentage of problematic cases in this section drops to 10 percent. (Results are for both categorical and continuous variables.)

Table 1. Percentage of Problematic Supplement Questions By Section

| Section | Question numbers | Total number of questions in section | Percent problematic questions |
|-------------------------------------|------------------|--------------------------------------|---|
| Food expenditures | 1-8 | 18 | 83 percent |
| Program participation | 9-9G | 10 | 20 percent |
| Food sufficiency | 11A-16 | 10 | 60 percent |
| Coping mechanisms and food scarcity | 17-52 | 36 21 | 28 percent 10 percent (excluding questions with less than 7 cases) |
| Concern about food sufficiency | 53-58 | 6 | 67 percent |

Reinterview

The Food Security Supplement reinterview was conducted from April 17-29, 1995 by CPS supervisors, senior field representatives, and interviewers. Approximately 90 percent of the reinterviews were conducted within 7 days of the original interview, but in some cases, there was up to a 10 day lag.⁵ The reinterview was conducted on a nationally representative sample of 1,827 with a response rate of 63.6 percent (1,162 completed interviews). The reinterview was conducted with the same respondent who had answered the original survey. The sample was split between households with family incomes at or below 185 percent of the poverty level and those with family incomes above 185 percent of the poverty level; 929 reinterviews were conducted with the former group and 233 with the latter. This sample was drawn in order to test two important features of the questionnaire: 1) the reliability of the screening questions that determined whether a respondent was asked the remaining questions that measure degree of food

⁵The number of days between the original interview and the reinterview may account for some of the unreliability measured in the index of inconsistency.

insecurity, and 2) the reliability of the questions on food insecurity. Because of cost constraints, most reinterviews were conducted by telephone.⁶

The major objective of the reinterview was to measure response variance, that is, to determine the degree of inconsistency between the original survey answer and the reinterview answer. The reinterview data contain several measures of response variance. We will use the index of inconsistency in this paper. This is a relative measure of response variance that estimates the ratio of response variance to total variance for each question. In general, an index of less than 20 indicates that response variance is low; an index between 20 and 50 indicates that response variance is moderate; and one over 50 indicates that response variance is high (McGuinness, forthcoming).⁷

Table 2 shows the mean and median index of inconsistency by section of the questionnaire for categorical variables.

⁶ Approximately 35 percent of the cases in the original interview were conducted by personal visit and 65 percent were conducted by telephone either from the field representatives' homes or from a centralized telephone facility. Personal visit interviews are primarily month-in-sample one and five cases, that is, those cases that are in sample for the first time or those cases that are returning to the sample after a four-month hiatus. Thus, as much as 35 percent of the sample may be subject to a mode effect and some of the variation in the index may be due to a mode effect. Based on differences in survey data resulting from personal visit vs. telephone mode effects, the consensus at the Census Bureau is that these differences are quite small and would contribute little to the variation in the index.

⁷ The index of inconsistency is the simple response variance divided by the total variance. Computationally it is the proportion who change answers between the original interview and the reinterview divided by $(P1*Q2) + (P2*Q1)$
where P1= the proportion in category from the original interview
where Q1= the proportion not in category from the original interview
where P2= the proportion in category from the reinterview
where Q2= the proportion not in category from the reinterview

Table 2. Mean and Median Index of Inconsistency for Each Section of the Questionnaire

| <u>Section</u> | <u>Mean</u> | <u>Median</u> |
|-------------------------------------|-------------|---------------|
| Food expenditures | 52 | 52 |
| Program participation | 25 | 19 |
| Food sufficiency | 46 | 47 |
| Coping mechanisms and food scarcity | 44 | 44 |
| Concern about food sufficiency | 53 | 52 |

In general, these data indicate that four of the five sections of the questionnaire are producing moderately to highly unreliable data, with the notable exception of the program participation section.

Results

Behavior coding guidelines generally state that a question is considered problematic if less than 85 percent of the time interviewers read questions exactly as written or with only slight changes that do not affect question meaning, or if less than 85 percent of respondents give adequate or qualified answers to the question (Oksenberg, et al., 1991). Our analysis is limited to questions with a minimum of 7 cases in the behavior coding data.

We compare the results of behavior coding to those of the reinterview data at the question level. That is, we compare the diagnostic utility of behavior coding in predicting which questions will yield reliable data on reinterview. We do not have matching datasets at the level of the individual respondent, since the samples for behavior coding and for reinterview were drawn independently.

The questionnaire contained 75 questions, plus one split ballot item. There were 55 categorical questions of the "mark one answer" type, 20 continuous questions, and one question that was a "mark all that apply" type. This question had 5 possible responses and is treated as five separate questions in this analysis.

We were unable to use all questions in our analysis for two reasons. First, 3 questions were excluded because they had less than seven cases in the behavior coding data, 16 were excluded because of an unreliable index of inconsistency, and 15 were excluded because of both reasons. In most cases, the index was unreliable because the characteristic of interest is rare in the population and too few respondents were reinterviewed to provide reliable estimates. Thus,

46 questions were available for analysis. Second, because the index of inconsistency is calculated differently for categorical and continuous variables and the small number (N=9) of continuous variables made it impossible to carry out separate analyses for them, we decided to restrict the analysis to categorical variables.⁸ The analysis in this paper is, therefore, restricted to the 37 categorical variables for which we have reliable behavior coding and reinterview data.

Table 3 shows the three models we used to test the predictive utility of the behavior coding data. The dependent variable is the index of inconsistency, a continuous variable that, in theory, ranges from 0 to 100.⁹ All three models include the two independent variables for the behavior coding data. These variables are percentages ranging from 0 to 100. The respondent behavior code is the percentage of times respondents provided an adequate or qualified answer to the question. The interviewer behavior code is the percentage of times interviewers read the question exactly as worded or with only slight changes that didn't affect question meaning. In addition to the two behavior coding variables, Model 2 includes three dummy variables representing the sections of the questionnaire. Although the questionnaire contains five sections, two of them--food sufficiency and coping mechanisms and food scarcity--are similar in content and are differentiated in the questionnaire only because the former is used to screen respondents either into or out of the remainder of the questions. Accordingly, these two sections were collapsed for the present analysis. The omitted category is the concern about food sufficiency section. The sections of the questionnaire were included in the model since we knew from both the behavior coding data and the reinterview data that not all of the sections performed equally well. Model 3 includes interactions between the respondent behavior code and the sections of the questionnaire.

⁸We did, in fact, run a general linear model separately for the numeric data. Because of sample size only the behavior coding variables could be used to predict the index of inconsistency. Neither the respondent nor the interviewer behavior coding variable was significant.

⁹It is possible for the index of inconsistency to be greater than 100 if the number of observed agreements is less than chance. See Perkins, 1971 for details.

Table 3. General Linear Models for Predicting the Index of Inconsistency (Standard errors in parentheses)

| | <u>Model 1</u> | <u>Model 2</u> | <u>Model 3</u> |
|--|---------------------------|---------------------------|---------------------------|
| <u>Variable</u> | <u>Parameter Estimate</u> | <u>Parameter Estimate</u> | <u>Parameter Estimate</u> |
| Intercept | 155.7 (57.1) | 76.7 (48.0) | -4.9 (69.0) |
| Respondent behavior code (RBC) | -0.6* (0.2) | -0.5* (0.2) | 0.3 (0.8) |
| Interviewer behavior code | -0.6 (0.6) | 0.2 (0.5) | 0.4 (0.4) |
| Food expenditure (Food) | | 15.3* (6.8) | 268.7** (75.5) |
| Program participation (Program) | | -26.5** (7.7) | 201.1* (91.0) |
| Food sufficiency, coping mechanisms and food scarcity (Coping) | | -7.5 (6.5) | 34.5 (67.4) |
| RBC*Food | | | -3.1** (0.9) |
| RBC*Program | | | -2.7* (1.1) |
| RBC*Coping | | | -0.5 (0.8) |
| Model r-square | 0.20* | 0.61** | 0.83** |
| Degrees of freedom | 2 | 5 | 8 |
| N | 37 | 37 | 37 |

** : p<.01 * : p<.05

Model 1 indicates that the respondent behavior code significantly predicts the index of inconsistency. The sign of the parameter estimate is in the expected direction; that is, as the percentage of respondents who provide adequate or qualified answers increases, the index of inconsistency decreases, indicating lower response variance (higher reliability). Interviewer behavior, however, is not significantly related to the index of inconsistency. These results are similar to those found by Belli and Lepkowski (1995).

The lack of association between interviewer behaviors and question reliability is not surprising. Very few questions were identified as problematic based on interviewer reading errors. Interviewer and respondent behavior coding data for the 37 questions of interest is included in Appendix B. Using the 85 percent threshold for determining whether a question was problematic indicates that only 2 of the 37 questions would be considered problematic based on interviewer reading errors. These same two questions plus an additional 12 were determined to be problematic based on respondent codes.

Model 2 includes the dummy variables for the sections of the questionnaire. (The omitted category is the concern about food sufficiency section.) The two behavior coding variables perform similarly in Model 2 as in Model 1. The parameter estimate for the respondent behavior code remains significant and inversely correlated with the dependent variable, and the interviewer behavior codes are not significant. Addition of the three dummy variables contributed significantly to the model R^2 . The results indicate that questions in the food expenditures section were associated with higher levels of response variance (more unreliable) and questions in the program participation section were associated with lower levels of response variance (more reliable) than questions in the omitted section. These findings are consistent with the behavior coding data. Using the 85 percent threshold, five of the seven questions from the food expenditures section of the questionnaire that are included in this analysis were identified as problematic based on respondent codes, whereas only one of the five questions in the program participation section of the questionnaire was identified as problematic based on respondent behavior codes.

Model 3 includes interaction terms between the respondent behavior coding data and the section of the questionnaire. The increase in the R^2 value between Model 2 and Model 3 is significant, indicating that the interaction terms contribute significantly to the amount of variation explained in the dependent variable. The interaction terms indicate that the ability of the respondent code to predict the dependent variable is contingent on the section of the questionnaire. The respondent code is significantly associated with the index of inconsistency only in the food expenditures and program participation sections. The respondent code was not significantly associated with the index in the combined food sufficiency/coping mechanisms sections. Appendix B shows that questions in this section performed well according to respondent behavior coding data, but produced relatively unreliable data according to the index. And respondent behavior coding data for the concern about food sufficiency section were mixed, whereas the index indicated the questions were uniformly unreliable.

Discussion

Why does behavior coding predict reliability of response in some sections of the questionnaire but not in others? On a purely statistical level, the lack of variation in the independent variable (respondent behavior code) in the combined food sufficiency/coping mechanisms and food scarcity section or the dependent variable in the concern about food sufficiency section is probably sufficient to preclude a significant effect of the behavior coding variable in those sections. The more interesting question, however, has to do with how these sections of the questionnaire differ from the others either in terms of the content of the questions, or in terms of their structure.

One way in which these sections differ from the others is that questions in the food expenditures and program participation sections are of a more clearly factual nature than those in other sections. The food expenditure section includes questions on whether the respondent shopped at various locations (supermarkets and grocery stores, other stores, and restaurants), whether they included all purchases regardless of how they paid for them, how often they shop at supermarkets and grocery stores, and whether the amount they spent last week is the usual amount they spend per week. The program participation questions ask about food stamp reciprocity, and participation in other food-related programs such as the school lunch and breakfast program and WIC. The remainder of the questionnaire measures the extent of food insecurity in the household. Questions in the concern about food sufficiency section are intended to measure a more subjective dimension of food insecurity than questions in the food sufficiency/coping mechanisms section. However, one could argue that several of the questions in the latter section are subjective as well (see particularly questions 32, 33, 35, 38 in the questionnaire).

A second difference is the reference period used in the questions. The food expenditure questions ask about shopping "last week," and the program participation questions ask about the "last 30 days." Questions in the other sections of the questionnaire have either long or nonexistent reference periods. Out of 25 questions, 19 ask about the "past 12 months," 3 ask about the "past 30 days," and 3 mention no reference period. Perhaps the long reference period results in respondents using recall strategies that produce unreliable data. Unfortunately, the data collected in this study do not allow us to investigate these hypotheses further.

Conclusions

For a long time, researchers have used behavior coding as a guide in questionnaire development, on the assumption that when respondents and interviewers are able to ask and answer questions without difficulty, the quality of the information obtained will be better. This assumption has been based largely on faith rather than empirical evidence. The findings in the present paper provide empirical support for the assumption, but they also appear to qualify it in some important respects. First, interviewer behavior coding has no predictive value for reliability, at least in a study such as this one, where interviewers perform at a uniformly high level. These findings might well differ in studies with greater variability among interviewers. Second, respondent behavior coding data do not appear to predict all types of reliability equally well. Prediction appears to be better for factual questions, and/or for questions with a relatively

short recall period. When these conditions are not met, people may be able to answer the questions--and, therefore, behavior coding data may give no indication of difficulty--but the reliability of answers (and, hence, their validity) may nevertheless be low. Clearly, more research is needed into the characteristics of questions for which behavior coding is a valid predictor of test-retest reliability.

In concluding, we would also like to draw attention to some limitations of our data that make us offer these conclusions with a great deal of caution. First, our results are not generalizable. The behavior coding data were not drawn from a random sample of households. They are primarily low income households from the first 75 low income cases interviewed at two of the Census Bureau's centralized telephone facilities. Moreover, the samples for behavior coding and reinterview are different. The reinterview sample is nationally representative, but was oversampled for low income households and suffers from a low response rate (64 percent). Second, because of differences in sample design and sample size, our analysis is at the question level, not the individual level. This analysis would be more precise if we had matched individual level data. Third, the number and type of questions contained in this analysis are very small and the questions are not constructed to deliberately vary either content or structure. Although there were 80 questions in the original survey, we were only able to include 37 questions in our model. Questions were excluded primarily because the characteristic of interest is so rare in the population that the reinterview sample was too small to produce a reliable index of inconsistency. Moreover, we had to exclude continuous variables from the model because the index is calculated differently for categorical and continuous variables and there were too few continuous variables to produce a separate model. Fourth, although approximately 90 percent of the reinterviews were done within seven days of the original interview, the elapsed time between the original interview and the reinterview may account for some of the unreliability measured in the index of inconsistency, and the impact of the elapsed time may not affect all questions equally. It is possible that questions with shorter reference periods, such as those asking about behaviors occurring "last week" in the food expenditures section, were more adversely affected by the elapsed time between interviews than questions with longer reference periods. Respondents may be answering the food expenditure questions about a different week during the reinterview than in the original interview.¹⁰ Thus, the index may not be speaking to reliability in the food expenditure questions and may be correlating with the behavior coding data for the wrong reason. Given these caveats, our results suggest that respondent behavior coding is associated with one measure of reliability; however, its ability to predict reliability in our study was not uniform throughout the questionnaire. Additional research is needed to understand the characteristics of questions for which behavior coding is a valid indicator of reliability and those for which it is not.

¹⁰The questionnaire was modified during the reinterview to prompt respondents to report for the week before the original interview.

REFERENCES

- Cannell, C., Lawson, S. and Hausser, D. (1975). *A Technique for Evaluating Interviewer Performance*. Ann Arbor, University of Michigan.
- Cannell, C., Oksenberg, L., Kalton, G., Bischooping, K., and Fowler, F. (1989). "New Techniques for Pretesting Survey Questions." Report submitted to the National Center for Health Statistics. Ann Arbor, Survey Research Center, University of Michigan.
- Hess, J., Singer, E., Ciochetto, S., "Evaluation of the April 1995 Food Security Supplement to the Current Population Survey." Report prepared by the U.S. Bureau of the Census, Center for Survey Methods Research for the U.S. Department of Agriculture Food and Consumer Service, Alexandria, VA, January 26, 1996.
- McGuinness, R., "Reinterview Report: Response Variance in the 1995 Food Security Supplement." Report prepared by the U.S. Bureau of the Census, Demographic Statistical Methods Division/QAEB for the U.S. Department of Agriculture Food and Consumer Service, Alexandria, VA, forthcoming.
- Perkins, Walter M., "On the Index of Inconsistency." Memo for the Center for Research and Measurement Methods, U.S. Bureau of the Census, 1971.
- Presser, S., and Blair, J. "Survey Pretesting: Do Different Methods Produce Different Results?" *Sociological Methodology*, Vol. 2, No. 12, pp. 73-104.

APPENDIX A

Interviewer and Respondent Behavior Codes¹

Interviewer Codes

- E: Exact question reading
- S: Slight change in question reading
- M: Major change in question reading
- V: Verify
- O: Other

Respondent Codes

- A: Adequate answer
- Q: Qualified answer
- I: Inadequate answer
- C: Requests clarification
- B: Break-in or interruption of question reading
- D: Don't know answer to question
- R: Refuses to answer question
- O: Other

¹ For a fuller description of the behavior codes, see Hess, Singer, and Ciochetto (1996), Attachment E.

APPENDIX B

Interviewer and Respondent Behavior Coding Data
and the Index of Inconsistency
for Questions Included in the Regression Models

| <u>Question Number</u> | <u>Interviewer Behavior²</u> | <u>Respondent Behavior³</u> | <u>Index of Inconsistency</u> |
|------------------------|---|--|-------------------------------|
| Food Expenditures | | | |
| 1 | 97.1 | 86.0 | 68.7 |
| 1A1 | 93.2 | 67.0 | 97.5 |
| 1C | 77.9 | 83.1 | 47.6 |
| 2 | 97.8 | 82.4 | 55.1 |
| 3 | 99.3 | 90.4 | 33.9 |
| 4 | 91.4 | 76.4 | 100.0 |
| 6 | 95.2 | 81.8 | 79.8 |
| Program Participation | | | |
| 9 | 96.6 | 92.1 | 9.6 |
| 9C | 100.0 | 92.9 | 19.4 |
| 9D | 93.3 | 86.7 | 32.0 |
| 9E | 96.9 | 78.1 | 47.1 |
| 9F | 95.4 | 88.4 | 15.1 |
| Food Sufficiency | | | |
| 11A | 100.0 | 52.0 | 46.8 |
| 11 | 98.2 | 83.3 | 47.1 |
| 12 | 99.0 | 61.2 | 52.3 |
| 15 | 97.0 | 85.2 | 42.1 |
| 16 | 97.0 | 94.0 | 41.3 |

²Percent exact or slight readings.

³Percent adequate or qualified answers.

| <u>Question Number</u> | <u>Interviewer Behavior</u> | <u>Respondent Behavior</u> | <u>Index of Inconsistency</u> |
|-------------------------------------|-----------------------------|----------------------------|-------------------------------|
| Coping Mechanisms and Food Scarcity | | | |
| 17 | 95.8 | 95.8 | 43.4 |
| 18 | 96.8 | 93.6 | 35.0 |
| 19 | 97.5 | 100.0 | 35.9 |
| 20 | 98.9 | 87.1 | 43.5 |
| 21 | 100.0 | 93.6 | 35.6 |
| 22 | 98.9 | 92.5 | 39.5 |
| 24 | 96.8 | 96.8 | 41.0 |
| 25 | 92.9 | 50.0 | 56.1 |
| 26 | 100.0 | 92.9 | 46.2 |
| 28 | 97.9 | 94.6 | 54.2 |
| 32 | 100.0 | 94.6 | 36.0 |
| 33 | 100.0 | 93.9 | 49.2 |
| 35 | 98.9 | 91.4 | 47.4 |
| 38 | 98.9 | 98.9 | 48.2 |

Concern About Food Sufficiency

| | | | |
|----|------|------|------|
| 53 | 81.7 | 77.4 | 54.1 |
| 54 | 90.3 | 82.6 | 48.7 |
| 55 | 92.5 | 79.6 | 54.2 |
| 56 | 95.0 | 87.5 | 50.1 |
| 57 | 97.5 | 85.0 | 65.9 |
| 58 | 97.5 | 75.0 | 48.0 |

A Discussion of Cutting Edge Research in Cognitive Interviewing and Behavior Coding
Robert F. Belli
University of Michigan

I would like to thank the speakers for presenting insightful papers that illustrate cutting edge research in cognitive interviewing and behavior coding. Before offering comments on this research, I would first like to review the purposes and problems of cognitive interviewing and behavior coding in order to frame my comments.

Cognitive Interviewing

The purpose of cognitive interviewing is to precisely assess the cognitive processes that affect the quality of survey reports. By engaging participants to explore in-depth their cognitive processes while answering questions, cognitive interviewing in principle is designed to determine whether survey questions pose problems in comprehension, retrieval, judgment, or answer formatting, and to specify the exact nature of these problems.

However, the findings of cognitive interviews may not generalize to actual field surveys. One issue is that cognitive interviews are largely based on convenience samples and small samples which do not mirror field surveys that typically involve large probability samples. A second issue is that the cognitive processes encouraged in cognitive interviews may not mirror those encountered in field surveys. For one thing, cognitive interviews are conducted in different settings and contexts than field surveys--they usually involve bringing participants into a controlled laboratory setting, whereas field surveys are administered to respondents who are in their own homes (if a household survey). In addition, the techniques used in cognitive interviews are largely unstandardized to allow the freedom to explore various cognitive processes, whereas the techniques of field surveys are typically standardized regarding the rules of interviewing. Finally, a third issue is that the interpretation of cognitive interviews are more subjective and based on the insights of the researcher rather than following publicly verifiable principles of scientific objectivity. Thus, conclusions drawn from various researchers and laboratories may not be reliable, a prerequisite for any ability to generalize outside the realm of the cognitive interviewing process.

Behavior Coding

The purpose of behavior coding is to identify those survey questions that pose the most threat to the ideals of standardized interviewing, both with respect to interviewer and respondent verbal behaviors. Interviewers are expected to read questions exactly as worded and to adequately and nondirectively address respondent misunderstandings with question content. Respondents, for their part, are expected in ideal conditions to be motivated to answer survey questions to the best of their ability and to express areas of misunderstanding if they occur.

Yet, with respect to data quality, it's not clear that the questions behavior coding prioritizes as problematic, that is, those questions that are most illustrative of being in variance with the ideals of standardized interviewing, are those that actually threaten the quality of survey report. Additionally, the problem codes, in and of themselves, do not precisely identify the

kinds of cognitive and interviewing processes that are posing problems.

Beatty Paper on Cognitive Interviewing

Beatty conducts sensible and clever research in seeking to show that problems revealed in cognitive interviews are generalizable to field surveys. For the most part, Beatty seeks to discover whether the same problems revealed in cognitive interviewing also demonstrate themselves in survey situations. Importantly, it is demonstrated that both the problems and solutions revealed in cognitive interviews are (at times) mirrored in field surveys, and that the cognitive problems revealed in survey interviews do affect the quality of survey report even if the style of interviewing in field settings tends to mask these problems. Additionally, Beatty is working toward establishing that different interviewing techniques in cognitive interviewing does not affect the conclusions drawn, that researcher interpretations are not merely subjective, and that cognitive interviewing results are reliable across laboratories. Finding evidence in support of these hypotheses will go a long way toward demonstrating the utility of cognitive interviewing in improving the quality of survey report.

I have a couple of comments regarding this work that I believe characterizes its potential. As shown by Beatty in his presentation, there will be a need for objective coding measures in this work to assess issues of reliability and validity. Blixt, Dykema, and Lepkowski, in their presentation, have illustrated the benefits of using objective coding schemes in assessing which questions, and what aspects of questions, pose the greatest threat to data quality. Beatty also provides an illustration of the benefits of such coding with the analyses of the question dealing with respondents' assessment of how many days during the past month their health was not good. Beatty provides codes both for independent and dependent variables by coding whether the interviewer engaged in an elaborating probing style as is typical for cognitive interviews, or by engaging in a re-orienting probing style as is more typical for field survey interviews, and by coding the precision of responses. Beatty finds that the elaborating probes revealed problems in cognitive processes that were masked by the re-orienting probes. No doubt that in extending this work the coding of cognitive interviews will be needed to assess whether different interviewing styles and different cognitive interviewing staff identify the same questions as problematic, and for the same reasons.

Related to this need for objective coding measures, the determination of whether laboratories that conduct cognitive interviews on small sample sizes will provide results that are generalizable to field surveys appears to remain as an intractable problem. Survey practitioners are not interested in any problem that may uniquely appear, after all, every survey question is likely to pose problems to some of the respondents some of the time. Rather, interest centers on those questions that pose the greatest threat, those that consistently reveal cognitive problems. As discussed by Blixt et al, the benefit of coding schemes is that they offer such an ability to identify the most problematic questions, but at the cost of requiring fairly large sample sizes (certainly beyond the tendency in cognitive interviewing to use sample sizes of 5-10 participants). Beatty in this research agenda will also require fairly large sample sizes to gain an understanding of the extent to which different interviewing techniques and different interviewers or laboratories are consistently finding the same problems.

Papers by Hess & Singer; Blixt, Dykema, & Lepkowski on Behavior Coding

Both of these papers indicate that variance from the ideals of standardized interviewing as revealed by behavior coding do affect the quality of survey reports. Importantly, quality of survey reports is measured in two different ways, by reliability of survey answers across the same survey questions administered on two occasions (Hess & Singer), and by the agreement of survey answers with external records (Blixt et al). Such a consistency of findings across different types of measures of data quality is reassurance regarding the authenticity of the results.

In comparing these different measures of data quality, both studies are able to ascertain the quality of factual data, but only the reliability measure (Hess & Singer) is able to determine whether there exist associations between behavior codes and the quality of answers to subjective questions. Interestingly, whereas both types of measures show that behavior coding is associated with survey quality with factual questions, Hess and Singer did not find reliable associations with subjective questions. Perhaps the fluid nature of subjective questions in the face of many competing contextual factors is responsible for the lack of findings.

Surprisingly, neither study found that interviewer question reading changes were associated with poorer data quality, in fact, Blixt et al. have counter intuitively found that interviewer variance from reading questions as written is associated with improvements in the exact matches between survey reports and medical records. In related work based on the same data, Belli and Lepkowski (1996) had not found any improvement in data quality associated with question wording changes. The difference between Blixt et al. and Belli and Lepkowski involves the manner in which comparisons of survey responses and external records were measured. Blixt et al. used a dichotomous dependent measure that distinguished between exact agreements and any disagreement, Belli and Lepkowski used a continuous measure based on the absolute value of the difference between reports and records. A possible explanation for the inconsistency of findings is that question-reading changes may be potent in affecting survey reports in opposite directions, on occasion being effective in leading to improved remembering, but at other times being counterproductive by encouraging poorer quality retrospective reports.

With regard to respondent behavior, both Hess and Singer and Blixt et al are consistent in showing that problem behaviors are associated with poorer quality reports. However, there are inconsistencies in that there are no reliable indications regarding the circumstances in which significant associations between the occurrence of problem behavior codes and data quality measures will appear. As one example, Blixt et al found that qualified answers are significantly associated with the occurrence of fewer exact matches in the reports and records for hospital stays over a 12 month reference period and office visits over a 4 week reference period, but qualified answers do not reveal significant associations in the quality of reports for office visits with either 6 month or 2 week reference periods. As another example, Blixt et al found that any respondent code problems were associated with fewer exact matches between reports and records for office visits that involved 12 month and 4 week reference periods, but not for 12-month hospital stays, 6-month office visits, or 2-week office visits. In short, there is no consistency in the appearance of significant findings based upon type of report or length

of reference period, and just as Alice in Wonderland's Cheshire cat, the effects show themselves at unpredictable times. Another noteworthy inconsistency is that Blixt et al particularly found poorer data quality associated with qualified responses whereas Hess and Singer noted significant data quality decrements associated with problem codes other than qualified answers, since they treated qualified responses in the same way as adequate answers, as an indication of nonproblematic respondent behavior. Overall, these inconsistencies highlight the continuing problem as to how to interpret the precise relationships between behavior codes, interviewing dynamics, and the quality of survey response. Our present level of understanding only permits very tentative explanations for the associations that do appear.

One message that is particularly clear about this research is that respondent behaviors are more indicative of compromises to data quality than anything that the interviewer has direct control. Whereas respondent problem behaviors are associated with poorer data quality, interviewer problem behaviors are not. Belli and Lepkowski (1996) found additional evidence that interviewer behavior does not affect the quality of report by finding that with reports on 12-month hospital stays, regardless of whether an interviewer probed adequately or inadequately, the occurrence of probing was associated with greater discrepancies between reports and records. Apparently, it was the need to probe following respondent behavior that is driving this effect, since the manner in which interviewers probed did not matter.

Results suggest that improvements in data quality are less likely to be promoted by concentrating on interviewer adherence to standardized procedures, and more likely to be promoted by devoting attention to respondent needs that will facilitate the effective answering of survey questions. Such advances will depend on improvements in questionnaire design principles that maximize the ability of respondents to answer accurately and consistently.

Reference:

- Belli, R. F., & Lepkowski, J. M. (1996). Behavior of survey actors and the accuracy of response. Health Survey Research Methods: Conference Proceedings (pp. 69-74). DHHS Publication No. (PHS) 96-1013.

Discussion: Validation of Cognitive Questionnaire Pretesting Methods

Theresa J. DeMaio
U.S. Bureau of the Census

I'd like to thank the authors for three very good papers. I enjoyed reading them all. And I think that the general topic of validating cognitive questionnaire pretesting methods is a very important one that deserves more attention than it receives. I'm going to focus my remarks today on the Beatty, Willis, and Schechter paper, since my research experience focuses more heavily on cognitive interviewing than on behavior coding.

I'd like to organize my comments around the four assumptions about the value of the methodology that were included in the paper. Beatty et al presented evidence from their work about these assumptions. I'm going to discuss the assumptions and present evidence relevant to them from my work at the Census Bureau. As with Beatty's examples, they are not derived from controlled experimental comparisons, but they are illustrative nonetheless.

Assumption #1. The cognitive interviewing method finds problems that will carry over to surveys. This is an important assumption, and one for which we have quite a bit of anecdotal evidence, I think. Beatty et al presented some evidence in their paper, and my work also substantiates this assumption. At the Census Bureau, we've done some testing of forms being developed for the 2000 census. The testing focused on the design aspects of the forms, rather than their content. The forms are self-administered, and that provides a bit of a twist to the average cognitive interview in ways that I will get back to later.

A well-planned and well-executed research program would incorporate preliminary stages of testing such as cognitive interviews prior to field testing. However, we all know how the constraints of operational schedules wind up squeezing the testing. In this case, I think there were definite advantages to the fact that the cognitive testing of three proposed census short forms took place simultaneously with a nationally representative field test that included these forms along with others.

Cognitive interviews showed that respondents thought two of the mailing envelopes were too flashy and didn't look official enough. The message that the census is mandatory, which was included on all the envelopes, was not imparted to respondents in some cases. There were differences in design aspects of the questionnaires, too, that were noted differentially as problematic by respondents, who completed all three forms. There was no roster on any of the forms and the item that requested the number of people living in the household had different, and in some cases unacceptably high, rates of item nonresponse. And the concept of the census including everyone in the household was not adequately conveyed on any of the forms.

When the nationally representative field test results came in (in the 1996 National Content Survey), the mail return rates for the envelopes that were viewed as flashy and unofficial suffered in comparison to the official envelope. Item nonresponse rates for the item requesting the number of household members were high. And many forms were received at the processing office that contained a single household member's name and information repeated in the answer spaces for up to five persons. In short, what we found in the laboratory was also experienced in the field.

Assumption #2. The response process when answering questions in a cognitive laboratory is more or less the same as in a survey interview. The very wording of this assumption assumes that cognitive interviews are conducted with interviewer-administered interviews, and all of the research reported here today deals with that type of interview. I agree that this assumption is inherent in the cognitive interview method, and the Beatty et al paper provides one clever attempt to provide evidence about this assumption. However, I think the assumption needs to be reworded to encompass self-administered interviews as well. In a self-administered interview, there is even more reason to question whether this assumption is a reasonable one. The respondent sits across the table from the cognitive interviewer, and while the respondent is completing the questionnaire, the interviewer in a concurrent interview is frequently asking probing questions.

One of the consistent findings we have noted in our research on self-administered questionnaires is that respondents invariably have problems with skip instructions. We have done interviews with different types of respondents, different questionnaire content, different formats for skip instructions and problems with skip instructions seem to be a constant. One possibility, of course, is that the cognitive interview situation, in requiring the respondents to focus both on the interviewer and the questionnaire, affects the respondent's ability to concentrate on the printed document and thus introduces skip pattern errors that would not occur otherwise. Although I can't say one way or the other whether this hypothesis is correct, I think it is an important research issue.

I and my colleague Cleo Jenkins have been considering this issue. While we haven't had an opportunity to collect information about skip instructions, we have collected data that may shed light on other aspects of the self-administered completion process. In the census form research that I mentioned previously, we built in a controlled experiment in which a random half of the interviews were conducted using concurrent think aloud methods and the other half were conducted using retrospective think aloud methods. (In retrospective interviews, the probing is conducted after the form is completed, while the probing in a concurrent interview takes place while the form is being completed.) We haven't had a chance to analyze these data yet, but I think it will provide a good opportunity to learn about the kinds of errors that respondents make in a concurrent vs. a retrospective interview, as well as the kinds of information that can be obtained through each. Unfortunately for us, but perhaps fortunately for the general public, the census short form does not contain skip instructions! A retrospective interview is not the same as completing the form at home, but at least we're taking incremental steps in the right direction.

Assumption #3. Cognitive interviewer behavior does not have an undue effect on the content of the interview. The Beatty et al paper notes that this assumption refers to two different things: first, that the interviewer's behavior affects the respondents' answers to the survey questions themselves, and second, that the interviewer's behavior affects the number of problems, types of problems, etc., that he/she encounters with the questions. I'm not sure that this assumption is really needed. I think the first aspect of the statement seems to overlap with assumption #1: that is, if cognitive interviewer behavior has an undue effect on the survey responses, it seems to me this would mean that cognitive interviewing results would not carry over to the actual survey. On the other hand, the second aspect refers to nonsystematic cognitive interviewer behaviors that could affect the research results they receive. And this seems to overlap with assumption #4, which I'll talk about next.

Assumption #4. The cognitive interviewing process is basically reliable--if repeated, it would yield similar results. This very important assumption is largely untested. Presser and Blair (1994) compared results across various pretesting methods, including cognitive interviewing, and within multiple trials of each one using a questionnaire that was a composite of various National Health Interview Survey supplement questionnaires in the early stages of development. They found that the results across three trials of cognitive interviewing were not totally consistent. They correlated the overlap between the questionnaire problems that were identified during the three independent sets of cognitive interviewing, and found that the correlations ranged from .4 to .6. This is the only systematic attempt I know of to compare the results of cognitive interviewing across interviewers or interviewing organizations. I know there have been other instances where, for example, the Census Bureau and the National Center for Health Statistics have conducted interviews on the same projects, but there has been no attempt to conduct comprehensive systematic analysis to compare the results. I think this is an area that is in need of future research, and I'm glad to see that Paul and his colleagues have some plans in that area.

Those are the assumptions that are presented in the Beatty et al paper. I also think there is another basic assumption that underlies the cognitive research we do that is mentioned but not given much prominence in the Beatty et al paper. I think it is important that we try to confront this issue, so I'd like to add another assumption to the list.

Assumption #5. Respondents have sufficient access to their thought processes that they can verbalize how they go about answering survey questions. This is in some ways related to assumption #1, but I think it goes deeper than that. We take what our respondents tell us as accurate renditions of their thought processes. Yet those of us who have conducted cognitive interviews know that there are distinct differences among respondents in their ability to verbalize what they are thinking about. Failure to verbalize a problem is not necessarily an indicator that no problem exists. Eleanor Gerber and Tracy Wellens (1996) have suggested that respondents may not be aware of cultural factors that come into play during the response process. And respondents may not appreciate the influence of the visual aspects of self-administered forms when they are completing them.

In some of our recent interviews on the census form, one of the objectives was to evaluate how respondents reacted to icons, or pictures with benefits messages, that were included on the form to provide information about why census questions are asked. Two kinds of things happened. In a few cases, respondents actually read some of the icons while they completed the form, but when asked whether they had noticed them, they said no. One respondent offered as an explanation, "Well, it might have gotten into my conscious but it never got into my subconscious." However, the more frequent occurrence was that respondents didn't appear to notice the icons at all, but when they were asked about them later, it was clear they had processed them, even though they never mentioned them during the think aloud. My point in bringing this up is to note that I think we need to investigate this assumption, like the others that are included in the Beatty et al paper.

In conclusion, I want to thank Beatty and his colleagues for their attempt to specify the assumptions that underlie the cognitive interview research that we do, and for giving me the opportunity to think about them.

REFERENCES

Gerber, E. and T. Wellens. 1996. "Perspectives on Pretesting: 'Cognition' in the Cognitive Interview." Paper prepared for presentation at the Essex '96 Fourth International Conference on Social Science Methodology, July 1996.

Presser, S., and Blair, J. 1994. "Survey Pretesting: Do Different Methods Produce Different Results?" Sociological Methodology, Vol. 2, No. 12, pp. 73-104.

TOWARD AN AGENDA FOR THE FUTURE

TOWARD A FUTURE AGENDA FOR THE FEDERAL COMMITTEE ON STATISTICAL METHODOLOGY

Graham Kalton

Westat, Inc. and Joint Program in Survey Methodology, University of Maryland

The main focus of my remarks is on a future agenda for the Federal Committee on Statistical Methodology (FCSM), a committee headed up so effectively for many years by Maria Gonzalez, to whom this conference is dedicated. Since the activities of the FCSM need to be responsive to the needs of the federal statistical system, I shall also comment on some likely future needs for that system, based in part on the presentations made in earlier sessions of the conference.

The current climate for federal statistics is one of increased demands and decreased resources. I believe that this climate calls for increased collaboration between the federal statistical agencies, and between these agencies and other bodies involved in statistical data collection and dissemination. This collaboration, which will be a recurring theme in my remarks, relates both to the production of statistics and to methodological research. In particular, in the current context, the climate enhances the need for contributions from the FCSM. In times of scarce resources, methodological research is often (mistakenly) a major victim. The FCSM can play the vital roles of sharing methodological advances across the whole of the federal statistical system and of promoting cross-agency contacts to facilitate the pooling of resources for methodological research.

The Personal Responsibility and Work Opportunity Act of 1996 will have a major impact on federal statistics. Welfare reform, with the devolution of responsibility for many welfare programs to states, calls for significant changes in many of the large federal surveys (e.g., the Current Population Survey, the Survey of Income and Program Participation, the

Medical Expenditure Panel Survey, and the National Health Interview Survey) and will affect the availability of welfare program data from administrative data systems. Information needs to be compiled on the programs offered by the individual states, and sometimes substate units, and survey questionnaires will need to be changed to reflect the new programs and their eligibility criteria. There will be greater need for state level estimates to monitor the effects of the different programs.

Welfare reform presents many methodological challenges for federal statistical programs. These include, for example, the difficulties of questionnaire design when respondents may not be familiar with the names of welfare programs, the need for small area data, and the development of linkages between state administrative program data and survey respondents. These challenges again call for collaboration between federal statistical agencies to pool their resources and develop unified methods of handling the problems that will be encountered. They also call for collaboration between federal and state statistical agencies so that data can be generated to satisfy both national and state needs. Such collaboration may, for example, establish compatible administrative data systems across the states. It may also involve partnerships between a federal statistical agency conducting a major national survey and some states to supplement the survey's sample in those states in order to produce state estimates of adequate precision.

Several of the sessions at this conference have indicated the benefits of close collaboration between the federal statistical agencies. An obvious example is the session on sharing data for statistical purposes. Another is the session on survey integration for health data, in which several health surveys are linked to the National Health Interview Survey. Other linkages may also be contemplated. In particular, linking large-scale surveys using inexpensive data collection methods - like the proposed American Community Survey to be conducted largely by mail and the National Immunization Survey conducted by telephone - to smaller scale

surveys collecting more extensive data in some form of two-phase sample design holds promise for the production of small area estimates.

Another session that points to the benefits of collaboration was the one on electronic dissemination of federal statistics. This field is experiencing rapid advances in technology, and there are clear benefits from collaboration between agencies. The benefits relate to keeping abreast of the technology, to the adoption of a common standard to aid users, and to the placement of data from several agencies in a single location, providing the user with one-stop shopping.

Panel surveys received little mention at this conference, but I believe they warrant further attention from federal statistical methodologists. There are nowadays many panel surveys and great strides have been made in addressing the additional complexities they present. Nevertheless, I think there remains much to be done, especially in the area of the longitudinal analysis of panel survey data. I see the methodology of panel surveys as a fruitful area for an FCSM activity, extending the work of the earlier subcommittee on this topic.

The issues of quality improvement, quality measurement, and quality reporting need to be kept on the agenda for the federal statistical agencies despite the tight financial situation. I interpret quality in broad terms, to include accuracy, relevance, timeliness, accessibility, and cost-efficiency. I am pleased to note the production of several quality profiles recently, and I hope this activity will continue, including both quality profiles for new surveys and updates of existing quality profiles. The FCSM Subcommittee on Quality Reporting might usefully develop guidelines for these quality profiles.

My last general suggestion for an area of attention for FCSM was stimulated by the comments of Victor Cohn in his keynote speech at this conference. The area is that of

communication skills, both oral and written. It is critically important that federal statisticians be able to communicate effectively, yet training of statisticians in communication is often limited. There is no quick remedy for this situation and it is not a problem that is methodological in a narrow sense. Nevertheless, it is a problem that FCSM might consider addressing.

I would now like to turn to the mission of the FCSM and its methods of operation. Margaret Martin (1991) provides a good review of the FCSM's mission, which includes exchanging information across agencies, promoting the advancement of best current methods, and obtaining consensus on definitions. As I argued at the previous seminar (Kalton, 1995), I think that special attention should be given to disseminating information to the smaller statistical agencies (whose needs are perhaps greatest), and to taking account of methodological advances developed in other countries and in organizations outside government.

The traditional way for the FCSM to carry out its work is by means of subcommittees appointed to study and produce working papers on specific topics. Recent working papers have, for instance, covered electronic dissemination, statistical disclosure limitation methodology, indirect estimates for small areas, and computer assisted survey information collection. Such working papers serve a very useful function. To attain full benefit from the work put into their production, they need to be distributed widely within the federal statistical service. They could also usefully be sent to government statistical agencies in other countries and to private survey research organizations.

While the working paper approach is effective in meeting some of the FCSM goals, it has its limitations. A working paper usually takes two or more years to complete, and once it is published the subcommittee disbands. Since working papers usually deal with topics of current research interest, they are likely to become rapidly outdated. A mechanism is needed to provide new information in a timely manner. One possibility is that FCSM could set up hot links

between the working papers (which are available on the internet) and other on-line information that provides news of recent developments. Another is that FCSM could establish list serves for continuing on-line discussion groups on working paper, and other, topics. Government statistical agencies in some countries produce journals or newsletters on statistical methodology. Running a full-fledged journal is a major undertaking, but newsletters, perhaps also on the internet, may be much more manageable and provide an effective way to communicate material rapidly among federal statisticians.

The FCSM might also convene workshops of federal statisticians and others to discuss topics of current concern, like the workshop held a couple of years ago on the use of incentives in surveys. Such workshops may be small-scale, informal, and convened at relatively short notice.

Finally, I believe that FCSM should continue to support these biennial symposia, well organized by Ed Spar and COPAFS. There are many conferences these days, and probably too many. However, this one, which brings together federal statisticians from many agencies, and which serves as a forum for the presentation of FCSM activities, is unique. I believe it serves a valuable function and I look forward to the next one.

REFERENCES

- Kalton, G. (1995). Improving the quality of federal statistics. In U.S. Office of Management and Budget *Seminar on New Directions in Statistical Methodology*, pp. 1-15. Statistical Policy Working Paper 23. Washington D.C. Statistical Policy Office.
- Martin, M. E. (1991). Towards an Agenda for the Future. In U.S. Office of Management and Budget *Seminar on Quality Federal Data*, pp. 462-464. Statistical Policy Working Paper 20. Washington, D.C. Statistical Policy Office.

Toward an Agenda for the Future
Nancy J. Kirkendall
Office of Management and Budget

Over the past two days we have heard many innovative ideas and suggestions for the Federal statistical system -- many of which resulted from the efforts of the Federal Committee on Statistical Methodology (FCSM), and its able leader Maria Gonzalez. Maria initiated the FCSM 20 years ago and was instrumental in seeing that it was a success. This conference is a special tribute to Maria; we miss her.

I am honored to have been selected to try to fill Maria's shoes, and to continue the important efforts of the FCSM. The FCSM is a very important collaborative effort of the Federal statistical agencies. In light of budget constraints and public distrust of government, collaboration among the Federal statistical agencies is more important than it ever has been.

I am interested in continuing and building on the good work the FCSM has done in the past. But I am also interested in seeing how we can make the FCSM even more valuable in the future. Please send me your ideas.¹

With that introduction, let me point out some of the themes for the future I heard during the past two days. John Rolph and Graham Kalton also identified major themes, and my themes are on their lists as well: first, communication, and second, observations about the Federal statistical system.

Communication

The issue of communication came up in many of the talks in this conference: examples include Vic Cohn's keynote speech, "What We the Public Needs to Know;" the session "Training Federal Statisticians;" and the session "What the Public Needs to Know About Federal Statistics." In the latter session, the comments by Terri Ann Lowenthal were particularly relevant. Terri Ann suggested that Federal statisticians need to establish useful working relationships with legislators and staff, the media, and other professional societies (users of our data). She suggested that we need to get involved in and understand policy debates, the data needed, and how the data will be used. In short, we need to broaden our focus beyond statistics, so that we can better communicate with our customers.

Communication is the key to trust, as well as to understanding. As Vic Cohn so aptly offered "If we want to gain public confidence, we must confide in the public." We have fallen short in communication, and as a result we have lost trust. This is a problem that stretches beyond the Federal statistical community. It is also a problem in the academic community. We do not teach elementary statistics courses so that they are interesting. David Grier, a professor at George

¹ Nancy Kirkendall: phone (202) 395-7313, fax (202) 395-7245, email kirkendall_n@a1.eop.gov.

Washington University, used to talk about meeting people at Washington cocktail parties. When he told them he was a statistician, they typically put their hands in their pockets, looked down, shuffled their feet and said something like "You're in statistics? I took a statistics course once. It was awful." If these basic courses were more interesting, people might remember more about them than that they were very difficult and incredibly boring. Ideally these courses should instill an appreciation for and understanding of statistics and statisticians.

There has been some relatively recent work within the American Statistical Association on making elementary statistics courses more interesting and informative. Novel ideas and hands-on experiments enhance communication. The Federal statistical system needs to identify and use convincing new ways to demonstrate the value of statistical methods and concepts.

How do we go about improving communication? I am not sure. But I believe that it is critically important for our profession as well as for the Federal statistical system.

Federal Statistical System

Connie Citro had many important recommendations for the Federal statistical system. Her suggestions embody one of my earlier points: interagency coordination will be absolutely necessary in the future. Connie's suggestions, augmented with additional insights from other speakers follow. OMB's Statistical Policy Office working in concert with the statistical agencies should:

1. Commission an interagency group to address emerging policy issues and data needs, and to decide how best to satisfy them.
2. Charge interagency forums and working groups with specific charters --identifying problems to be solved, and working together to solve them.
3. Designate a lead agency for immigration (also recommended by Jeff Passel). One of the problems with the data on immigration is that no agency feels responsible for this area.
4. Establish a cross-cutting advisory group (in addition to the ongoing advisory groups for specific agencies.)
5. Consider implications of the Government Performance and Results Act, not only for our own agencies, but for the system as a whole. This supports Joe Wholey's comment in the session on "Performance Measurement in Statistical Agencies" that in 15 months, the government will be required to have a performance plan. We should strive to see that the statistical agencies are represented as a unit in that plan.
6. Use the INTERNET to advantage. There is a social statistics briefing room on the White House home page. It could be used to discuss cross-cutting issues such as kids, immigration, etc. We should consider ways to use the INTERNET to facilitate dialogue on cross-cutting issues.

7. Develop standards of good documentation for INTERNET products. For example, we need ways to describe data limitations in a form or format that people would be willing to read. In the session "Quality in Survey Data" Renee Miller talked about coming up with short, interesting descriptions of survey data and their limitations. I believe she said that the Energy Information Administration's descriptions were modeled after those prepared by the National Center for Education Statistics.

8. Be aware of the sociology of the Federal statistical system. As noted by Dan Melnick in the session "What the Public Needs to Know About Federal Statistics," each agency's approaches to problems -- its strengths and weaknesses -- result from its own history, staff, etc. A cross cutting approach can generate synergy, taking advantage of our diversity.

All of the sessions were interesting and valuable, and most pointed to a continuing need for interagency cooperation. For example, in the session "Sharing Data for Statistical Purposes," Tom Mesenbourg's paper made an implicit assumption that Congress will pass the data sharing legislation. We certainly hope that will be the case! As Tom observed, if the Statistical Confidentiality Act that would implement data sharing is passed, the statistical community will need a plan for implementing it in an orderly way. We will need to consider how to phase in this new approach and how to establish priorities. Kathy Wallman observed that we will need to work together to develop standards and guidelines for implementing the act.

An Update on the FCSM

About a year ago two new subcommittees of the FCSM were established to pursue training federal statisticians and to review and report on quality in survey data. Two of the sessions at this conference were organized by the new subcommittees.

The session on Training Federal Statisticians presented the efforts of a team lead by Cynthia Clark. They have done a lot of work, shared many ideas and made great progress toward their goal. In their session, the panel discussion that followed the formal subcommittee presentations focused on the need for training in new fields: large data bases, handling messy data, INTERNET, performance measures (timeliness, accuracy, relevance of products), political skills, ethical issues, and communication.

The second new FCSM subcommittee is not as far along, and needs to be revitalized. Nevertheless, there were two very interesting papers in the session on Reviewing and Reporting Quality in Survey Data. The papers by Claes Andersson from Statistics Sweden and by Renee Miller from the Energy Information Administration described approaches to measuring data quality and identified issues. The discussants pointed out that measures of total quality are most likely too expensive in times of shrinking budgets, and that it is not clear what users need or want in terms of the quality of documentation. Hopefully, the FCSM subcommittee will take on some of the challenges identified during the session. I would personally like to see this subcommittee identify common approaches for the Federal statistical system to use in measuring quality and timeliness. These are two fundamental attributes of information, and a more integrated statistical

system require will require common measures to support evaluation and comparison.

I also want to take special note of a particular recent success. In April, 1996, the FCSM along with many of the statistical agencies cosponsored a seminar on data editing, which was attended by about 500 people. The proceedings volume from that seminar appeared as Statistical Policy Working Paper 25 in December 1996.

Conclusions

This has been a very valuable conference with many excellent talks and discussions. Perhaps conferences like this are one of the best outcomes of interagency cooperation. I would like to thank Ed Spar, Helen Peck, and Susan Cohen from COPAFS for their part in organizing this conference. I did not realize until today that Helen came back from retirement to help out.

Thanks to all participants and attendees. What would a conference be without you? Please remember to send me your ideas for the future of the Federal statistical system in general and for the FCSM in particular.

**Reports Available in the
Statistical Policy Working Paper Series**

1. *Report on Statistics for Allocation of Funds*, 1978 (NTIS PB86-211521/AS)
2. *Report on Statistical Disclosure and Disclosure-Avoidance Techniques*, 1978 (NTIS PB86-2115/AS)
3. *An Error Profile: Employment as Measured by the Current Population Survey*, 1978 (NTIS PB86-214269/AS)
4. *Glossary of Nonsampling Error Terms: An Example of a Semantic Problem in Statistics*, 1978 (NTIS PB86-211547/AS)
5. *Report on Exact and Statistical matching Techniques*, 1980 (NTIS PB86-215829/AS)
6. *Report on Statistical Uses of Administrative Records*, 1980 (NTIS PB86-214285/AS)
7. *An Interagency Review of Time-Series Revision Policies*, 1982 (NTIS PB86-232451/AS)
8. *Statistical Interagency Agreements*, 1982 (NTIS PB86-230570/AS)
9. *Contracting for Surveys*, 1983. (NTIS PB83-233148)
10. *Approaches to Developing Questionnaires*, 1983 (NTIS PB84-105055/AS)
11. *A Review of Industry Coding Systems*, 1984 (NTIS PB84-135276)
12. *The Role of Telephone Data Collection in Federal Statistics*, 1984 (NTIS PB85-105971)
13. *Federal Longitudinal Surveys*, 1986 (NTIS PB86-139730)
14. *Workshop on Statistical Uses of Microcomputers in Federal Agencies*, 1987 (NTIS PB87-166393)
15. *Quality in Establishment Surveys*, 1988 (NTIS PB88-232921)
16. *A Comparative Study of Reporting Units in Selected Employer Data Systems*, 1990 (NTIS PB90-205238)
17. *Survey Coverage*, 1990 (NTIS PB90-205246)
18. *Data Editing in Federal Statistical Agencies*, 1990 (NTIS PB90-205253)
19. *Computer Assisted Survey Information Collection*, 1990 (NTIS PB90-205261)
20. *Seminar on Quality of Federal Data*, 1991 (NTIS PB91-142414)
21. *Indirect Estimators in Federal Programs*, 1993 (NTIS PB93-209294)
22. *Report on Statistical Disclosure Limitation Methodology*, 1994 (NTIS PB94-165305)
23. *Seminar on New Directions in Statistical Methodology*, 1995 (NTIS PB95-182978)
24. *Report on Electronic Dissemination of Statistical Data*, 1995 (NTIS PB96-121629)
25. *Data Editing Workshop and Exposition*, 1996 (NTIS PB97-104624)
26. *Seminar on Statistical Methodology in the Public Service*, 1997 (NTIS PB97-162580)

Copies of these working papers may be ordered from NTIS Document Sales, 5285 Port Royal Road, Springfield, VA 22161; telephone: (703) 487-4650. The Statistical Policy Working Paper series is also available electronically through the Bureau of Transportation Statistics World Wide Web home page (<http://www.bts.gov>).