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:

- The full series of interviews should be completed in approximately 14,600 households.
- 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.
- The sample should produce approximately unbiased national estimates and estimates at the census region level.
- 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 of 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	
1996 MEPS Panel		**Round 1A**	**Round 2A**	**Round 3A**	**Round 4A**	**Round 5A**	**Round 6A**
Responding Households (by Round)	10,800	9,500 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	
1997 MEPS Panel				**Round 1B**	**Round 2B**	**Round 3B**	**Round 4B**
Responding Households (by Round)	5,600			5,057 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 heath 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 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 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:

Yij = 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 Yij be estimated as

a model based function of Xij, or

Yij = f(Xij) where

Xij = 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, Xij, will be recalibrated to a predicted provider reported response

Yij = f(Xij).

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

Yij = Xij.

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 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 ares 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 careful 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 theses 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.