Statistical Policy
Working Paper 23

Seminar on New Directions in Statistical Methodology

Part 2 of 3

Federal Committee on Statistical Methodology

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

June 1995

MEMBERS OF THE FEDERAL COMMITTEE ON
STATISTICAL METHODOLOGY

(June 1995)

Maria E. Gonzalez, Chair
Office of Management and Budget

M. Denice McCormick Myers, Secretary
National Agricultural Statistics Service

Susan W. Ahmed
National Center for
Education Statistics

Yvonne M. Bishop
Energy Information
Administration

Cynthia Z.F. Clark
National Agricultural
Statistics Service

Steven Cohen
Administration for Health
Policy Research

Lawrence H. Cox
Environmental Protection
Agency

Zahava D. Doering
Smithsonian Institution

Daniel Kasprzyk
National Center for
Education Statistics

Nancy Kirkendall
Energy Information
Administration

Daniel Melnick
Substance Abuse and Mental
Health Services Administration

Robert P. Parker
Bureau of Economic Analysis

Charles P. Pautler, Jr.
Bureau of the Census

David A. Pierce
Federal Reseerve Board

Thomas J. Plewes
Bureau of Labor Statistics

Wesley L. Schaible
Bureau of Labor Statistics

Rolf R. Schmitt
Bureau of Transportation
Statistics

Monroe G. Sirken
National Center for
Health Statistics

Robert D. Tortora
Bureau of the Census

Alan R. Tupek
National Science Foundation

Denton R. Vaughan
Social Security
Administration

Robert Warren
Immigration and Naturalization
Service

G. David Williamson
Centers for Disease Control
and Prevention

PREFACE

The Federal Committee on Statistical Methodology was organized by the Office of Management and Budget (OMB) in 1975 to investigate issues of data quality affecting Federal statistics. Members of the committee, selected by OMB on the basis of their individual expertise and interest in statistical methods, serve in a personal capacity rather than as agency representatives. The committee conducts its work through subcommittees that are organized to study particular issues and prepare working papers presenting their findings. The subcommittees are open by invitation to Federal employees who wish to participate. This is the 23rd Statistical Policy Working Paper published under the auspices of the committee since its founding.

On May 25-26, 1994, the Council of Professional Associations on Federal Statistics (COPAFS) hosted a "Seminar on New Directions in Statistical Methodology." Developed to capitalize on work undertaken during the past fifteen years by the Federal Committee on Statistical Methodology and its subcommittees, the seminar focused on a variety of topics that have been explored thus far in the Statistical Policy Working Paper series and on work on statistical standards undertaken by the Statistical Policy Office at OMB. The subjects covered at the seminar included:

Economic Classification Revisions
Disclosure Limitation Methodology
Customer Surveys
Advances in Data Editing
Time Series Revision Policies
Incentives in Surveys
Computer Assisted Survey Information Collection
Longitudinal Surveys
Cognitive Testing and Self-Administered Questionnaires
Statistical Uses of Administrative Records
Small Area Estimation
Nonresponse in Surveys

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

Statistical Policy Working Paper 23, published in three parts, presents the proceedings of the "Seminar on New Directions in Statistical Methodology." In addition to providing the papers and formal discussions from each of the twelve sessions, the working paper includes Graham Kalton's keynote address, "Improving the Quality of Federal Statistics," and comments by Norman M. Bradburn, Robert M. Groves, and Katherine K. Wallman at the closing session, "Toward an Agenda for the Future."

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

Table of Contents

Wednesday, May 25, 1994

Part 1

KEYNOTE ADDRESS

IMPROVING THE QUALITY OF FEDERAL STATISTICS 1
Graham Kalton, Westat, Inc.

Session 1 - ECONOMIC CLASSIFICATION REVISIONS

ECONOMIC CLASSIFICATIONS IN THE NEW NORTH AMERICAN INDUSTRY CLASSIFICATION SYSTEM (NAICS). 17
Jack E. Triplett, Bureau of Economic Analysis

REVISING THE UNITED STATES STANDARD OCCUPATIONAL CLASSIFICATION (SOC) SYSTEM. 29
Thomas J. Plewes, U.S. Bureau of Labor Statistics

COMMENTS ON THE REVISIONS OF THE STANDARD INDUSTRIAL AND OCCUPATIONAL CLASSIFICATIONS. 34
Joel Popkin, Joel Popkin & Company

COMMENTS ON ECONOMIC CLASSIFICATION REVISIONS. 37
Joe Mattey, Federal Reserve Board

Session 2 - DISCLOSURE LIMITATION METHODOLOGY

RESTRICTED DATA VERSUS RESTRICTED ACCESS: A PERSPECTIVE FROM "PRIVATE LIVES AND PUBLIC POLICIES". 43
George T. Duncan, Carnegie Mellon University

STATISTICAL DISCLOSURE LIMITATION METHODOLOGY. 57
Nancy J. Kirkendall, Energy Information Administration

DISCUSSION OF PRESENTATIONS ON STATISTICAL DISCLOSURE METHODOLOGY. 68
Stephen E. Fienberg, Carnegie Mellon University

SELECTED ASPECTS OF RESTRICTED DATA. 81
Tore Dalenius, Brown University

Session 3 - CUSTOMER SURVEYS

QUALITY MANAGEMENT FOR CUSTOMER SATISFACTION SURVEYS. 87
Richard M. Devens, Jr., U.S. Bureau of Labor Statistics

COMPARABILITY IN CUSTOMER SATISFACTION SURVEYS: PRODUCTS, SERVICES AND GOVERNMENT AGENCIES. 99
Michael D. Johnson, University of Michigan

CUSTOMER SURVEYS: DISCUSSION. 121
Robert M. Groves, University of Michigan

DISCUSSION. 126
Elizabeth Martin, U.S. Bureau of the Census

Session 4 - ADVANCES IN DATA EDITING

IMPROVING OUTLIER DETECTION IN TWO ESTABLISHMENT SURVEYS. 137
Julia L. Bienias, David M. Lassman, Scott A. Scheleur, and Howard Hogan, U.S. Bureau of the Census

TIME SERIES AND CROSS SECTION EDITS WITH APPLICATIONS TO FEDERAL RESERVE DEPOSIT REPORTS 152
David A. Pierce and Laura Bauer Gillis, Federal Reserve Board

DISCUSSION. 172
Sandra A. West, U.S. Bureau of Labor Statistics

DISCUSSION. 177
Brian V. Greenberg, U.S. Bureau of the Census

Session 5 - TIME SERIES REVISION POLICIES

TIME SERIES REVISIONS: THE EFFECTS ON GROSS DOMESTIC PRODUCT. 183
Robert P. Parker and Teresa L. Weadock, U.S. Bureau of Economic Analysis

RAISING THE NATION'S UNEMPLOYMENT RATE. 209
John E. Bregger, U.S. Bureau of Labor Statistics

COMMENTS ON PARKER AND WEADOCK, TIME SERIES REVISIONS: THE EFFECTS ON GROSS DOMESTIC PRODUCT. 220
Murray F. Foss, American Enterprise Institute

Part 2

Session 6 - INCENTIVES IN SURVEYS

TIME, DOLLARS, AND DATA: SUCCEEDING WITH REMUNERATION IN HEALTH SURVEYS. 225
Trena M. Ezzati-Rice, Andrew A. White, William R. Mosher, National Center for Health Statistics, Maria Elena Sanchez, Agency for Health Care Policy and Research

THE USE OF INCENTIVES TO SURVEY "HARD-TO-REACH" RESPONDENTS: A BRIEF REVIEW OF EMPIRICAL RESEARCH AND CURRENT RESEARCH PRACTICE. 256
Richard A. Kulka, National Opinion Research Center

DISCUSSION. 290
Diane K. Willimack, National Agricultural Statistics Service

DISCUSSION. 295
W. Sherman Edwards, Westat, Inc.

Thursday, May 26

Session 7 - COMPUTER ASSISTED SURVEY INFORMATION COLLECTION

REDESIGNING A QUESTIONNAIRE FOR COMPUTER-ASSISTED DATA COLLECTION: THE CURRENT POPULATION SURVEY EXPERIENCE. . . 301
Cathryn Dippo and Anne Polivka, U.S. Bureau of Labor Statistics, Kathleen Creighton, Donna Kostanich, and Jennifer Rothgeb, U.S. Bureau of the Census

AN OVERVIEW OF THE 1992 CENSUS OF AGRICULTURE CATI SYSTEM. 311
Jeanette K. Mon, U.S. Bureau of the Census

DISCUSSION: WHAT CAN CAI LEARN FROM HCI?. 363
Mick. P. Couper, Joint Program in Survey Methodology

DISCUSSION OF TWO PAPERS ABOUT CASIC. 378
Sandra Sperry, Westat, Inc.

Session 8 - LONGITUDINAL SURVEYS

PANEL DESIGN AND ESTIMATION STRATEGIES IN THE NATIONAL MEDICAL EXPENDITURE SURVEY. 381
Steven B. Cohen, Agency for Health Care Policy and Research

METHODOLOGICAL ISSUES ENCOUNTERED IN FOLLOWING A COHORT OF EIGHTH GRADERS. 428
Steven J. Ingels, National Opinion Research Center, Jeffrey A. Owings, National Center for Education Statistics

DISCUSSION. 446
Gary M. Shapiro, Abt Associates

DISCUSSION. 449
James M. Lepkowski, University of Michigan

Part 3

Session 9 - COGNITIVE TESTING AND SELF-ADMINISTERED QUESTIONNAIRES

LITERACY LIMITATIONS AND SOLUTIONS FOR SELF-ADMINISTERED QUESTIONNAIRES. 453
Judith T. Lessler and James O'Reilly, Battelle Memorial Institute

THE LANGUAGE OF SELF-ADMINISTERED QUESTIONNAIRES AS SEEN THROUGH THE EYES OF RESPONDENTS. 470
Cleo R. Jenkins, U.S. Bureau of the Census and Don Dillman, U.S. Bureau of the Census and Washington State University

DISCUSSION. 517
Jared B. Jobe, National Center for Health Statistics

DISCUSSION. 524
Roger Tourangeau, National Opinion Research Center

Session 10 - STATISTICAL USES OF ADMINISTRATIVE RECORDS

IMPROVING DATA QUALITY THROUGH INCREASED DATA SHARING. . .531
Edward A. Trott, Bureau of Economic Analysis

HEALTH REFORM INFORMATION SYSTEMS: GREAT EXPECTATIONS, UNCERTAIN PROSPECTS. 540
Edward L. Hunter, National Center for Health Statistics

DISCUSSION. 549
Miron L. Straf, Committee on National Statistics, National Academy of Sciences

Session 11 - SMALL AREA ESTIMATION

SMALL AREA ESTIMATION FOR THE NATIONAL HEALTH INTERVIEW SURVEY USING HIERARCHICAL MODELS. 555
Donald Malec, National Center for Health Statistics, J. Sedransk, State University of New York, Albany

THE ROLE OF DESIGN BASED VARIANCES AND COVARIANCES IN SMALL DOMAIN ESTIMATION. 569
Robert E. Fay, U.S. Bureau of the Census

DISCUSSION. 595
Phillip S. Kott, National Agricultural Statistics Service

DISCUSSION. 599
David A. Marker, Westat, Inc.

Session 12 - NONRESPONSE IN SURVEYS

EXPLORING NONRESPONSE IN U.S. FEDERAL SURVEYS. 603
Maria Gonzalez, Office of Management and Budget, Dan Kasprzyk, National Center for Education Statistics, Fritz Scheuren, Internal Revenue Service

MODEL-BASED REWEIGHTING FOR NONRESPONSE ADJUSTMENT. . . . 625
David A. Binder, Sylvie Michaud and Claude Poirier, Statistics Canada

DISCUSSION. 646
J. Michael Brick, Westat, Inc.

DISCUSSION. 649
Joseph L. Schafer, Pennsylvania State University

TOWARD AN AGENDA FOR THE FUTURE

Norman M. Bradburn. 657
National Opinion Research Center

Robert M. Groves. 660
The Joint Program in Survey Methodology

Katherine Wallman. 665
Office of Management and Budget

Part 2

Session 6

INCENTIVES IN SURVEYS

TIME, DOLLARS, AND DATA: SUCCEEDING WITH REMUNERATION IN HEALTH SURVEYS

Trena M. Ezzati-Rice, Andrew A. White, William D. Mosher
National Center for Health Statistics, CDC

Maria Elena Sanchez
Agency for Health Care Policy and Research

Introduction

All surveys experience nonresponse in spite of the fact that the goal for all sample surveys is to maximize survey response, and thus minimize nonresponse bias in the survey estimates. Over the years, survey incentives (both monetary and non-monetary) have been used in conjunction with other survey methodologies to obtain complete and accurate information for the largest number of sample units.

Incentive use in surveys has spanned a wide variety of survey types, sponsors, respondents, and survey topics. As the title suggests, we present and discuss the use of remuneration (monetary incentives) in selected health surveys conducted by the National Center for Health Statistics, CDC (NCHS), the Agency for Health Care Policy and Research(AHCPR), and Project HOPE.

Experimental and field results have demonstrated that incentives can have a positive impact on survey response rates, and we will concentrate on describing selected incentive applications and experiments. Included here are new results from a field trial of remuneration in NCHS's National Survey of Family Growth, Cycle 5 Pretest.

Additional information in the area of remuneration in health surveys not reviewed in this paper can be found in the literature[1, 2, 3, 4]. In particular, the paper by Kulka

---

[1] Kulka, R.A. (1992, October). A Brief Review of the Use of Monetary Incentives in Federal Statistical Surveys, presented at the COPAFS Symposium on Incentives in Surveys, Boston, MA.

[2] Willimack, D.K., Petrella, M., Beebe, T., and Welk, M. (1992, August). The Use of Incentives in Surveys: Annotated Bibliography, Survey Research Center, Institute for Social Research, University of Michigan.

included in this volume gives a brief background on the use of monetary incentives and references numerous articles this topic across various types of surveys. Kulka also addresses sociological models proposed to describe the operation of incentives. The paper by Willimack et al. is an unpublished review of published literature however they do make the observation : "The bulk of the published literature regards the effects of incentives in mail surveys. No doubt incentives have been and/or are being used in both telephone and face-to-face surveys, but there is a lack of documentation of tests in the published literature. Perhaps incentives have been implemented in telephone and face-to-face surveys based on the mail survey results and on 'common sense,' without specific testing within mode. Based on a brief look at conference abstracts, it appears that documentation of incentive testing and/or use in survey modes other than mail may be found in non-published literature, such as professional association conference presentations."

Overview of Health Survey Issues and Remuneration

The use of monetary incentives is not new in health surveys and over the years those of us working in the field have learned a great deal about their use. Much of the material in this section was presented at an October 1992 COPAFS/OMB Symposium on Providing Incentives to Survey Respondents held at the John F. Kennedy School of Government, Harvard University, Cambridge, MA.

Although most surveys conducted by the Federal Government are based on unpaid, voluntary participation, monetary or gift incentives for participation may be justified for certain types of household and establishment based surveys in order to increase participation rates, encourage accurate record keeping, and/or keep expenses down. In addition for health surveys, remuneration may be justified for surveys which involve a physical examination and the drawing of a blood sample in order to maximize response rates.

High response rates bring the benefits of increased validity through increased precision and reduced potential for bias in survey estimates Incentives and

---

[3] Kulka, R. A. (1994, May). The use of Incentives to Survey "Hard-to-Reach" Respondents: A Brief Review of Empirical Research and Current Research Practice, presented at the COPAFS Seminar on New Directions in Statistical Methodology, Bethesda, MD.

[4] Dillman, D. A. (1991). The design and administration of mail surveys. Annual Review of Sociology, 17, p. 225-249.

remuneration can be considered appropriate whenever respondents are asked to devote time and effort (sometimes a considerable amount) to assisting the government in obtaining high quality data for research and policy related issues which will impact the entire nation.

For population based surveys, incentives can result in a higher motivation to participate, increased effort to give accurate (honest) responses, greater acceptance of government surveys, increased response rates, lower item nonresponse rates, and cost savings through better data quality and fewer call backs and/or canceled appointments. Further, cash incentives enhance the importance of the survey to respondents and provide tangible evidence of the value of their input. Incentives may stimulate otherwise reluctant respondents to participate and to more readily make themselves available to participate. Gift incentives for children could reinforce the value and importance of their participation.

The largest potential for net benefits from remuneration or incentives will be found: among surveys that experience higher refusal or item non-response rates; among surveys where persons at higher risk of being targeted respondents are less likely to participate (thus biasing the results); in situations where respondents are easy to locate, but initial or continuing cooperation is hard to gain; among those surveys requiring repeated contact in a short time period (like a month) or with other significant burden involved (like a physical examination or drawing of blood); or among those that require respondents to do something on their own, like complete a mail questionnaire or keep diaries, or participate in a followup survey of initial nonrespondents.

In particular incentives are most likely to have an effect in surveys that: require the respondent to travel; are lengthy or have a longitudinal component; are focussed at hard to reach populations (like adolescents or young black males), or that ask questions about sensitive topics (like income, drug use, risk behaviors related to HIV/AIDS). Remuneration may also gain survey participation when sample persons do not perceive an immediate benefit to themselves and/or society by participating.

Preparation and pilot testing of questionnaires (especially laboratory based testing) often requires respondent travel and takes large blocks of respondent time. The potential benefit to the survey is so large from this type of testing that remuneration is well worth it.

For institutional surveys, incentives can result in money and time savings (e.g., hospital versus government staff abstracting hospital records), removal of a barrier to participation (financial loss), greater acceptance of government surveys, cost savings

through fewer call backs and/or canceled appointments, increased response rates, and for some businesses, lower item nonresponse rates, and better data quality.

Remuneration in institutional surveys is often seen as compensation rather than as an incentive to participate in voluntary surveys. Institutions think of time as money and may consider compensation a requirement to engage in substantial continuing activities. Some may set that standard for any voluntary survey. In this sense, remuneration may make a survey possible. In addition to participation, remuneration to the institution can be cost beneficial to the government if summary data must be compiled from business records for survey purposes. The institution can often do it for less than having government field representatives comb the records. In businesses where staff energies must be diverted to complete a survey, the thoroughness and accuracy of the response may be improved if the business is being remunerated.

In institutional surveys, the continuing nature of data collection from particular respondents is a critical factor. There may be no other way to obtain the data (e.g. the is a charge for access to hospital records and state vital records already exist, why should the government be excepted?). Remuneration may actually be less expensive than providing the person hours needed to compile the data from institutional records even if the institution is willing to grant access.

A review of these issues and the use of incentive methods took place at the 1992 COPAFS/OMB Symposium[5]. This led to a recommendation, "that OMB seriously consider an agency's request to use incentives in a limited number of specific situations in which a survey violates the norm of what is considered the standard survey." The participants defined a standard survey as: a cross-sectional survey of the household population done in about an hour in a single session at the respondent's convenience; and done in the respondent's home with non-intrusive, nonsensitive questions. Symposium participants suggested that incentives be considered in the following special situations:

Surveys of hard-to-reach or special population subgroups;

Surveys involving unusual demands or respondent intrusions such as -

---

[5] Council of Professional Associations on Federal Statistics (COPAFS). (1993), September). Providing incentives to survey respondents: Final report (Contract No. GS0092AEM0914). Washington, DC: Regulatory Information Service Center, General Services Administration.

Lengthy interviews,
Keeping a diary,
Taking physical or physiological tests, or
Going somewhere special to participate;

Surveys involving sensitive questions and/or topics;

Surveys involving a commitment to participate over time such as for a panel survey; and

Surveys where respondents are not a household respondent such as a physician, hospital, or nursing home.

Specific Experiences with Remuneration in Health Surveys

This section highlights some major institutional and population based surveys which successfully use or have used remuneration to increase survey response and/or data quality. For some of the summaries we have paraphrased or reproduced language in the original references for the sake of accuracy.

NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY

The National Health and Nutrition Examination Survey (NHANES) is a periodic survey designed to assess the health and nutritional status of the noninstitutional population of the United States. The NHANES consists of a household interview with adult, youth, and family medical history questionnaires, followed by a 3-4 hour standardized physical examination in specially equipped mobile examination centers (MEC's). The NHANES is based on a stratified multistage cluster probability sample design[6]. The on-going Third NHANES or NHANES III is the seventh in a series of surveys using health examination procedures that have been conducted since 1960 by NCHS.

As for most large-scale Federal surveys, the success of the NHANES surveys depends upon voluntary participation of individuals selected in the sample.

---

[6] Ezzati, T., Massey, J., Waksberg, J., Chu, A., and Maurer, K. (1992). Sample Design: Third National Health and Nutrition Examination Survey. National Center for Health Statistics. Vital Health Statistics. Series 2, No. 113.

Remuneration has been used in all NHANES surveys and has been shown to be necessary for attaining adequate response rates. For the three early Health Examination Surveys (HES) conducted in the 1960's of adults, children, and youths, respectively, the examination response rates were excellent ranging from 87 percent to 96 percent[7]. However, with the beginning of NHANES I(1971-74), the examination response rate (64 percent) was much lower than those in the earlier HES surveys. After extensive efforts to improve the miserable response rate through interviewer re-training, increased publicity, and community outreach, the response rates remained at an unsatisfactory level. Therefore, it was proposed that a monetary incentive be considered to reduce examination nonresponse[8]. Since there was little information available from previous studies to show the effect of paying respondents to participate in health surveys, a field experiment was undertaken. First, of course, justification had to be submitted to OMB and approval obtained. It was reasoned that remuneration for NHANES was justified since participation in the survey required several hours of the respondent's time (thus lost time from work) and paid assistance for child care might also be required. It was hypothesized that the cost of the remuneration would be offset by a reduction in the number of contacts to a household to obtain respondent participation. Further, if the response rates increased significantly, the overall validity of the survey results would outweigh the remuneration costs.

The experimental design for the study was superimposed upon the within primary sampling unit (PSU) design for NHANES I[9]. The study was undertaken during 1972 in the San Antonio, Texas, PSU. The segments within the PSU were randomly paired by segment size and median family income. All of the sample persons in one segment of each pair were told about the $10 remuneration. The sample persons in the other segment of the pair were not told of the $10 remuneration. It should be noted however, that all persons who were examined received $10. The difference was that persons in the "not told" segments did not know about the remuneration until after they had been examined, while those in the "told" segments knew of the $10 remuneration before being examined.

Telling sample persons that they would be given $10 after completing the examination phase of the survey had a positive effect on the response rate in San

---

[7] Bryant, E., Kovar, MG., and Miller, H. (1975). A Study of the Effect of Remuneration Upon Response in the Health and Nutrition Examination Survey. National Center for Health Statistics. Vital Health Statistics. Series 2, No. 67.

[8] Ibid.

[9] Ibid.

Antonio. Among the 303 persons in the experimental segments (told) who were contacted, 82 percent were examined (see Table 1). On the other hand, among the 292 persons in the control segment (not told), only 70 percent were examined. Thus, the NHANES I experiment showed that the offer of $10 to sample person increased the response rate by 12 percentage points. Also, there was some evidence that sample persons were more cooperative and that less effort was required to obtain response when remuneration was offered as evidenced by the number of persons making an examination appointment at the first interview contact. Also, a larger proportion of the "told" group kept their appointments than the "not told" group. Only 2.1 contacts per examined person were required for the "told" group as compared with 2.5 such contacts per examined person for the "not told" group.

Table 1. Examination Response Rates from NHANES I Remuneration Experiment (sample sizes shown in parentheses)

	Not Told of Payment	Told of Payment
Experiment	70%	82%
	(292)	(303)
Actual Survey*	68%	78%
	(7335)	(6035)

*Represents results from the first 35 stands of NHANES I.

The findings of the NHANES I remuneration study were considered conclusive enough to include remuneration in the remainder of NHANES I. The overall response rate at the 45 survey locations where remuneration was offered in NHANES I (including San Antonio) was 78 percent as compared to 68 percent for the 19 survey locations where remuneration was not offered.

The $10 remuneration used in NHANES I was continued in NHANES II. However, about midway through NHANES II, the response rate dropped to about 70 percent. Therefore, another field experiment was used to assess the effect of increased remuneration on survey response. The study for NHANES II (1976-80), looked at the impact of increasing the $10 remuneration used in NHANES I to $20. Three survey locations were selected for the study. Each segment was paired with another segment similar to it with regard to poverty/non-poverty status and distance from the

examination center. Sample persons in one of each pair of segments were "told" that they would receive $20 for participating in the examination, while the sample persons in the other paired segment were told they would receive $10. However, all persons who were examined were given $20 no matter which monetary amount they were originally told..

There was a significant positive effect on response rates with the $20 incentive[10]. Out of 720 persons offered $20 for participation, 79 percent were examined (see Table 2), while among the 716 persons offered $10, only 74 percent were examined. An important finding from NHANES II which did not show up for NHANES I was the increase in the response rate by number of persons in a household for the $20 group versus the $10 group. There was no effect in households with only one sample person. However, in households with two and three or more sample persons, payment of $20 rather than $10 increased response rates by 8 and 16 percentage points, respectively[11].

Table 2. Examination Response Rates from NHANES II Remuneration Experiment (sample sizes shown in parentheses)

	Told of $10 Payment	Told of $20 Payment
Experiment	74%	79%
	(716)	(720)
Actual	72%*	76%**
	(NA)	(NA)

*Represents the final 44 stands of NHANES II.
**Represents the first 16 stands of NHANES II.

---

[10] Findlay, J. and Schaible, W.L. (1980) A Study of the Effect of Increased Remuneration on Response in a Health and Nutrition Examination Survey. Proceedings of the Section on Survey Research Methods of the American Statistical Association, pp. 590-594.

[11] Ibid.

The results indicated that there was a potential cost saving associated with the fact that the higher paid group was generally more cooperative. For example, it was found that a larger proportion of the examined respondents in the higher paid groups went to the examination center as a result of the first contact (68 percent versus 61 percent). Also, the higher paid group required, on average, fewer contacts overall (1.77 vs. 2.09)[12].

Overall, for NHANES II, for the 44 survey locations at which $10 was offered to sample persons, the response rate was 72 percent. But for the 16 stands at which $20 was given, the response rate was 76 percent.

In the Puerto Rican phase of Hispanic HANES (1984), the response rates were unsatisfactory, so an increase in remuneration from $20 to $50 was made. A cross-tabulation of response rates (see Table 3) by method of payment before and after the increased remuneration was instituted indicated a higher response rate in the $50 group (83 percent) than for the $20 group (76 percent)[13]. The results, however, should be viewed with caution since the increased payment was not randomized within the various survey locations.

Table 3. Hispanic HANES Examination Response Rates for New York City Metro Area (Puerto Rican Phase) Stands by Payment Amount*, 1984

Payment Amount	Examination Rate
$20	76% (3101)
$50	83% (576)
DK**	0% (116)

*Unpublished data from J. Findlay, NCHS.
**Cases could not be classified into either payment category due to lack of information.

[12] Ibid.

[13] Unpublished data from Jean Findlay, National Center for Health Statistics.

In the on-going NHANES III, the basic remuneration payment is $30 for all persons who come to the examination. However, for adults 20 years and older there is an additional incentive of $20 if they come for their examination at the "right time". Time of day of the examination and fasting status need to be controlled for the analyses of many of the biochemical tests including the Oral Glucose Tolerance Test (OGTT - a test for diabetes). Thus a random half-sample of adults is designated to have blood drawn in the morning, while the remaining 50-percent is selected for afternoon or evening appointments.

The NHANES III which includes an oversample of both Mexican-Americans and Blacks and includes no upper age limit (in contrast to previous NHANES) is experiencing higher examination response rates than any other previous NHANES survey. There is a six percentage difference in the overall examination response rate between NHANES II and NHANES III-Phase 2 (79% in the on-going NHANES III-Phase II versus 73% in NHANES II - see Table 4). This increase is due not only to the increased remuneration amount but also to the fact that NHANES III samples multiple persons per household and includes on average two persons per household. Clearly, there is a strong monetary incentive for a household as a whole if several members are selected into the sample.

Table 4. Examination response Rates for NHANES II, Hispanic HANES, and NHANES III

	MEC Examined	MEC + Home Examined
NHANES I	74	NA
NHANES II	73	NA
HHANES	73	NA
NHANES III - PHASE 1	77	78
NHANES III - PHASE 2*	79	80

*On-going.

NHANES III examination of nonresponse rates by household size show that the nonresponse rates decrease significantly with increasing household size. Another important finding related to participation in NHANES III is the increase in the

response rate for the OGTT. This test requires the sample person to fast 10-16 hours and requires two blood draws. The OGTT response rate in NHANES II among examined persons was only 66 percent thus requiring the results to be interpreted with extreme caution. However, for NHANES III, the OGTT response rate, again among examined persons, has increased by 11 percentage points (77 percent response rate for phase 1) and by 23 percentage points so far for phase 2 (89 percent response rate)[14]. The increased remuneration in NHANES III is felt to be responsible in large part for this important increase in the OGTT response rate for NHANES III.

NHANES Summary

Previous research associated with NHANES I and NHANES II has involved several field experiments to assess the impact of monetary incentives on achieving an adequate response rate so reliable estimates can be produced and generalized to the total population. The NHANES I remuneration experiment generally showed that paying sample persons $10 rather than nothing at all increased response rates about 12 percent.

The NHANES II study looked at the effect on the response rate of increasing the remuneration from the $10 used in NHANES I to $20. The average increase in the response rate was about 5 percent.

The results from both the NHANES I and NHANES II had two other findings pertinent to remuneration. First, the results indicated that there were potential cost savings associated with the fact that the higher paid groups were generally more cooperative. For example, it was found that a larger proportion of the examined respondents in the higher paid groups went to the MEC as a result of the first contact. They also required, on the average, fewer contacts overall. They also had fewer broken appointments.

Further, both of these experimental studies showed a marked increase in cooperation in households with more than one sample person, and the differences between the remuneration groups went up with household size. A positive relationship between household size and response rate was also observed in the Hispanic HANES. This trend is continuing in NHANES III as well. These results, presumably, are due to the fact that the monetary award increases substantially for the household as a whole when multiple persons are selected per household. Finally, the response rates in

[14] Unpublished data from Meena Khare, National Center for Health Statistics.

NHANES III are higher than those for previous NHANES surveys and the response rates for the two minority subgroups are higher than for whites and all others. In addition, the response rate for the OGTT in NHANES III is higher than for NHANES II.

The NHANES experiences demonstrate the power of remuneration to stimulate respondent participation in the face of inconvenience and financial disincentives; and in addition to motivate respondents to provide highly personal and sensitive information including physical and biological characteristics. The use of remuneration in NHANES III shows how phased payments can improve scheduling logistics as well. The remuneration serves as a motivator, and also to offset financial disincentives, and gives evidence of the importance of participation and survey goals. NHANES data have been significantly improved through incentive use.

NHANES I EPIDEMIOLOGIC FOLLOWUP STUDY

As mentioned in the previous section, examinees were paid $10 to participate in the baseline NHANES I study. They were again paid $10 to participate in the first NHANES Epidemiologic Followup Study (1982-84 NHEFS interview). Although no remuneration was paid in the 1986 followup of the elderly (conducted using computer assisted telephone interviews: CATI), subjects with reported hospital or nursing home admissions in the 1987 Followup were paid $5 as an incentive to sign and return the Medical Authorization Form (MAF). Payment was needed in the 1987 Followup because of the sharp decrease in the rates at which MAF's were returned in the 1986 Followup. Remuneration of the 1987 respondents increased the MAF return rate in the elderly group by 11 percentage points from 75.0% in the 1986 Followup to 85.7% in the 1987 Followup.

For the 1992 wave, a $5 remuneration was made subject to reported hospital or nursing home admissions since the last contact. Respondents are interviewed by telephone. However, the form which authorizes the hospital or nursing home to release patient information is mailed to the respondent and must be signed and returned. The remuneration is paid to the respondent or to the individual who signs the MAF, if this person is different from the respondent, as an incentive to return the signed authorization.

In addition, due to the difficulty of gaining the cooperation of some hospitals and nursing homes, if a nursing home or hospital requests reimbursement for work performed in abstracting or photocopying selected information from the admission and discharge records or abstracts, a small amount of money is provided to defray their

expenses. During 1987, a total cost of $2,170 or .0012 percent of the total contract cost was paid. This remuneration mostly involved nursing homes.

The NHEFS uses remuneration to gain commitment to continuing participation in a longitudinal study.

NATIONAL HEALTH INTERVIEW SURVEY YOUTH RISK BEHAVIOR SUPPLEMENT

In 1991 the Survey Research Center of the University of Michigan and the Bureau of the Census helped the National Center for Health Statistics assess the impact of financial rewards on respondent participation and motivation in a Youth Risk Behavior Supplement (YRBS) to the 1992 National Health Interview Survey (NHIS). This study[15] employed cognitive interview techniques and a traditional split sample field experiment.

The YRBS contained a number of sensitive questions requiring answers that could be perceived as being socially undesirable or threatening to self-image. Such topics included questions on the use of alcohol, drugs, sexual activity, over- or underweight status, and exercise program involvement.

Interviews were conducted both with groups of parents and youth before field trials and a pretest were conducted. The field trials were then held where youth were interviewed in their homes, paid $20, and debriefed after the interview about their views on paying respondents for participating in surveys. The pretest of the full survey procedures included a split sample in which the interviewers mentioned the $20 payment half of the respondents and did not mention the payment to the other half.

In the split sample experiment, complete interviews were obtained in 90% of the households where the $20 was mentioned, but in only 79% of the remaining households. The cognitive interviews conducted with the groups and during the field trial debriefings helped the researchers develop insight concerning reasons for this difference in response and for a potential increase in data quality when payment is

---

[15] Kalton, G., Cannell, C., Camburn, D., Oksenburg, L. and Holland, L. (1991). The Effect of Financial Incentives on Respondent Participation. Final report of the Association of Schools of Public Health Cooperative Agreement: "Applied Research on the Conduct of Adolescent Health Behaviors and Characteristics", University of Michigan, pp. 47-56.

mentioned in advance.

The cognitive interviews indicated that unwillingness to report events or behaviors is only partially caused by concerns over privacy or confidentiality. In fact, youth respondents could see no reason to be diligent in answering survey questions. They voiced skepticism about any benefits that may accrue from survey participation. Therefore techniques typically used with adult respondents, such as appealing to their obligations as citizens or emphasizing the societal benefits may be unproductive. The researchers concluded that a $20 remuneration offer, linked with a signed commitment to participate and give accurate answers, would be an effective way to motivate respondents to participate and report accurately. The debriefing interviews confirmed this conclusion.

A further interesting note from this study is that the success of respondent payments in obtaining YRBS interviews may be partly due to the reaction of interviewers to the payments. During interviewer debriefings, their comments suggested that respondent payments also have a forceful, positive influence on the attitudes and expectations of interviewers. Interviewers with such a positive outlook may feel they are likely to obtain an interview, rather than expecting a refusal, prior to contacting a potential respondent and subconsciously may convey to potential respondents a more positive view of the YRBS study. The researchers postulate that the total impact of respondent payments on participation rates is the sum of the positive direct effect on respondents and the indirect effect that payments have on the attitudes of interviewers.

YRBS incentive use focused on the motivational aspects of incentives for a non-traditional target population in a study that collected sensitive data.

NATIONAL HOUSEHOLD HIV SEROPREVALENCE SURVEY FEASIBILITY STUDY

A feasibility study for a National Household HIV Seroprevalence Survey (NHSS), based on a probability sample of households was conducted in Dallas County, Texas, in the fall of 1989. One of the major concerns of a household survey attempting to estimate the prevalence of HIV infection is that a high proportion of persons who are at the greatest risk of HIV infection may refuse to participate. This possibility of differential rates of response between those at higher risk and lower risk of HIV infection means that estimates derived from the survey have the potential to be biased. Among procedures to maximize the response rate in the NHSS, a $50 incentive payment was provided to all sample persons who provided a blood sample to be tested

for HIV antibodies and completed a self-administered risk behavior questionnaire. The $50 incentive seemed to have a positive impact on survey participation with a higher than expected response rate for a highly sensitive survey involving the collection of HIV risk behavior data and the collection of a blood specimen in the home. The response rate for the combined questionnaire and blood sample was 84 percent, and 90 percent for the questionnaire only[16]. Respondents in the NHSS were asked to check all reasons for their participation in the survey. Among all respondents, 47 percent stated "helping with AIDS research", while another 39 percent stated the "$50 payment" (see Table 5).

Table 5. Reasons for participation in the Dallas County Household HIV Survey

Reasons	Percent
Helping with AIDS research	46.8
U.S. Public Health Service Sponsorship	2.9
$50 payment	38.5
Videotape presentation	1.2
Assurance of privacy	3.3
Other factors, unspecified	7.2

In addition to estimating the prevalence of HIV infection in Dallas County, another major objective of the survey was to evaluate various methods for assessing and reducing nonresponse bias. A standard survey method for assessing bias due to nonresponse is to conduct a followup survey with a sample of initial survey nonrespondents with different incentives for participation. In the NHSS, a special followup study of a sample of nonrespondents was conducted in which half of the sample persons were offered an incentive of $100 to complete the self-administered risk behavior questionnaire only, and the other half was offered $175 to complete both the questionnaire and provide a blood sample. The followup survey of nonrespondents increased the questionnaire only response rate by 10 percentage points (80% vs. 90%)

[16] National Household Seroprevalence Survey Feasibility Study Final Report. Research Triangle Park, NC: Research Triangle Institute, April 30, 1990. Research Triangle Report RTI/4190-01F.

and the blood and questionnaire by four percentage points (80% vs. 84% - see Table 6). Of particular importance was the increased reporting of risk behavior in the followup study. The prevalence rates among male respondents for three major HIV risk behaviors (intravenous drug use, receptive anal intercourse, and multiple sex partners) were 3 to 5 times higher in the followup survey than in the regular survey (see Table 7).

In summary, results from the NHSS feasibility study followup survey indicated that a high proportion of persons who initially refused to participate, when recontacted and offered an increased incentive, completed the risk behavior questionnaire. A lower proportion of persons who initially refused to participate provided both a blood sample and completed the risk questionnaire, when recontacted and offered an increased incentive. Persons at higher risk for HIV infection participated at higher levels in the followup survey than in the regular survey. The followup survey effectively increased the total number of persons who participated in the Dallas HIV survey. The increase in risk reporting among the sample of regular survey nonrespondents that were followed-up allowed for a significant reduction in nonresponse bias in the HIV estimate produced for Dallas County.

Table 6. Sample persons response rates in the Dallas County Household HIV Survey, 1989

Survey component	Regular survey	Regular + followup survey	Overall*
Screening	97	98	98
Blood & Questionnaire	80	84	82
Questionnaire only	80	90	88

*Product of screening and sample person rate.

Remuneration in the NHSS feasibility study demonstrated the power of incentives even in the most sensitive topic and invasive data collection situations, however establishing the exact mechanics of the reasons for success would require further study.

Table 7. Prevalence of selected HIV risk behaviors in the regular and followup survey, Dallas County male population, 18-54 years, 1989

Risk Behavior Since 1978	Regular Survey	Followup Survey	Total Estimate
Intravenous drug use	3	12	4
Receptive anal intercourse	3	11	5
5+ male partners	2	10	3
1+ male partner	5	17	8

THE NATIONAL SURVEY OF FAMILY GROWTH PRETEST

The National Survey of Family Growth (NSFG) is done periodically by NCHS to collect national data on the factors that affect the U.S. birth rate and women's reproductive health--factors that include sexual activity, marriage and divorce, contraception, sterilization, infertility, miscarriage, and abortion. Previous cycles of the NSFG have interviewed about 8,000 women 15-44 years of age in the noninstitutional population of the United States with response rates ranging from 75 percent to 80 percent.

Interviewing for the next NSFG, called Cycle 5, will be conducted in January-July of 1995. Three of the principal challenges for Cycle 5 of the NSFG will be (1) increasing response rates to make it possible to conduct a telephone reinterview in 1997 with as many of the original respondents as possible, (2) improving reporting of HIV-related sexual behavior, and (3) improving the reporting of abortion.

Response rates.--Most recently, in 1988, the NSFG used a list sample of households interviewed in the NHIS. Using a list sample saves nearly a million dollars on sample design and selection costs, but it makes it necessary to find women who move between the NHIS interview and the NSFG interview. Some are never found, so response rates are reduced somewhat. Response rates have been between 75 and 80 percent in recent cycles, despite the intrinsic appeal of the subject matter, the use of only female interviewers, thorough interviewer training, advance letters

introducing the survey, and expensive, intensive nonresponse follow-ups[17].

Although the NCHS obtains data from the National Health Interview Survey (NHIS) on HIV-related knowledge and attitudes, the NSFG remains the NCHS's principal vehicle for collecting data on HIV-related behavior, including such sensitive topics as age at first intercourse, numbers and characteristics of sexual partners, and condom use.

Abortion reporting is critical in the NSFG Cycle 5 because 25% of all pregnancies, and half of all unintended pregnancies, end in abortion. Fertility surveys in the U.S. and other nations have obtained incomplete reporting of abortion. In the last 3 cycles of the NSFG, in 1976, 1982 and 1988, and in most other U.S. surveys, women reported less than half of the abortions they have actually had[18]. This incomplete reporting of abortion has several potential adverse effects: it makes impossible analyses of the determinants and consequences of abortion itself; it forces us to use ad hoc methods to produce estimates of pregnancy rates for the U.S.; it produces biased estimates of the failure rates of contraceptive methods[19]; and it forces us to study unintended births instead of unintended pregnancies.

The NSFG Pretest for Cycle 5, conducted in October-December 1993, was based on about 800 eligible women, of whom 500 completed interviews. The Pretest was an experiment, which was designed to test several alternative contexts for asking questions. The pretest had 3 main groups:

1) the first group was a standard Computer-Assisted Personal Interview in the respondent's home with no incentive.
2) in the second group, the interview was moved to a neutral site--a site outside the home--where spouses, children, or parents could not hear the respondent's answers. To reimburse women for the time and inconvenience of going to the neutral site, respondents were paid $40 in cash at the end of the interview.
3) when we considered these first two groups, we were concerned that we might

---

[17] Judkins, David P, W. Mosher, and S. Botman. (1991). National Survey of Family Growth: Design, Estimation, and Inference. Vital and Health Statistics, Series 2, No. 109.

[18] Jones, Elise and J. Forrest. (1992). Under-reporting of Abortion in Surveys of U.S. Women: 1976-1988. Demography 29 (1): 113-126.

[19] Jones, Elise and J. Forrest. (1992). Contraceptive Failure Rates Based on the 1988 NSFG. Family Planning Perspectives 24 (1): 12-19.

obtain higher response rates and data quality in the neutral site/$40 group than in the in-home no-incentive group, but we would not know whether the differences were due to the non-home site or to the $40 payment. We were also concerned that the costs of implementing a $40 payment and setting up non-home sites on a national scale might be prohibitive. Therefore, we added a third group--a $20 incentive for an interview in the home.

For half the respondents in group 1 and group 3, (the in-home interviews), we also tested a short questionnaire at the end of the interview, using Audio CASI (Computer Assisted Self-Interviewing with headphones, and the respondent entering her answers into the computer).

NSFG Results

Pretest response rates (as a percent of those located) were higher for incentive cases than for non-incentive cases: 81 percent for those who received a $20 incentive vs 73% for those who received no incentive. The percent who broke an appointment with the interviewer was one-third lower for those who received $20 than for those who received no incentive (24 vs 37%).

The number of hours that the average interviewer worked to get a completed case was about 2 hours less for incentive than for non-incentive cases (8.8 vs 10.9; note also that when the incentive increases to $40, hours per case drops more than 2 full hours - see Table 8). Since the time of interviewers costs more than $10 an hour for their wages plus benefits, if the interviewer can save 2 hours of effort per case by paying a $20 incentive, then the incentive pays for itself. That is precisely what happened in the NSFG Pretest: the incentives paid for themselves in the $20 group because respondents broke fewer appointments for interviews and made themselves available after fewer telephone calls and personal visits. Costs in the $40 non-home group were higher because of high costs to set up the neutral sites--obtaining permissions, renting office space, etc.

Reporting of the number of sexual partners was higher among respondents who received incentives, still lower than reports of comparable studies of men[20, 21]. In

[20] Billy, J.O.G., Tanfer, K., Grady, W.R., and Klepinger, D. (1993). The Sexual Behavior of Men in the United States. Family Planning Perspectives 25 (2): p. 52-60.

previous cycles of the NSFG, less than half of abortions were reported[22]. Thus, abortion reporting should be at least double what it was in the last Cycle. In the pretest, a $20 incentive and the use of Audio CASI doubled abortion reporting, compared to a no-incentive, no Audio CASI group. This more complete reporting of abortions is probably due to two factors: (1) women using the Audio CASI (headphone) questionnaire reported a higher percentage of their abortions, and (2) incentives produced better coverage of groups of women who have higher abortion rates--including black women and poor women.

Table 8. Response Rates, Broken Appointment Rates, and Interviewer Hours per Case from the NSFG Pretest for Cycle 5

Experimental Group	Response Rate (%)	Percent with Broken Appointment	Hours per Case
In-Home, No Incentive	73	37	10.9
In Home, $20	81	24	8.8
Non-Home, $40	80	31	6.4

Recommendations from the Pretest

The results from the NSFG pretest suggest that a $20 incentive plus Audio CASI (self-administered questionnaires over headphones) should be used in the NSFG main study. The incentive will increase response rates, particularly among minorities and low-income women, and reduce the cost of interviewer labor because respondents will cooperate more readily.

Can these results be generalized? The results on response rates, interviewer hours and costs, in the NSFG Pretest are quite similar to those in the National Adult Literacy

---

[21] Smith, T.W. (1991). Adult Sexual Behavior in 1989: Number of partners, frequency of intercourse and Risk of AIDS. Family Planning Perspectives 23 (3): p. 102-107.

[22] Op cit. Jones and Forrest, Demography, Feb. 1992.

Survey (NALS) Field Test, a survey with a much larger sample (n=2,000) than the NSFG Pretest. Like the NSFG, the NALS required considerable effort from respondents[23].

The NALS and NSFG Pretest results provide evidence that incentives may be most cost-effective when the interview is:

> (1) either long or a great deal of effort, or both; (2) sensitive either because it deals with private behaviors or may otherwise cause embarrassment (the NALS might cause such embarrassment among the adult illiterate, the NSFG because it includes questions on abortion and sexual behavior); and (3) part of a panel survey in which the response rate is critical to maintain the size of the panel over time.

All three of those conditions were common to both the NALS and the NSFG. The NSFG experience clearly demonstrates the success of incentives with hard-to-interview populations and sensitive topics. In addition, it clearly demonstrates the cost-effectiveness of the methodology in improving survey quality.

NATIONAL MEDICAL EXPENDITURE SURVEYS

The National Medical Expenditure Surveys (NMES) are designed to produce estimates of medical use, medical expenditures, sources of payment for medical care, and health insurance coverage. The surveys are sponsored by the Agency for Health Care Policy and Research (AHCPR). The household survey (HS) component of the NMES series yields estimates for persons in the civilian non-institutionalized population, while the survey of the institutionalized population (IPC) yields estimates for persons in nursing homes. Incentives have not been a design feature of the IPC surveys, therefore, the remarks that follow will focus on the use of incentives in the NMES HS surveys.

Respondent Incentives

The NMES series of household surveys includes the 1977 National Medical Care Expenditure Survey (NMCES), the 1987 National Medical Expenditure Survey (NMES2), and the 1996 National Medical Expenditure Survey (NMES3) presently in

---

[23] Berlin, M, L. Mohadjer, J. Waksberg, et al. (1993). An Experiment in Monetary Incentives, in American Statistical Association (editor), 1992 Proceedings of the Section on Survey Research Methods, pages 393-398.

the planning phase. These NMES studies share the following design features:

- the use of an initial screening interview to identify the households to be sampled for the study;
- the oversampling of poor people, blacks, Hispanics, and the elderly;
- a longitudinal or panel design in which sampled families are interviewed several times over a period of 14-16 months, in rounds of data collection that cover a year-long observation period;
- face-to-face interviews lasting on average between 2 and 2.5 hours in each round to be completed with a family respondent that provides information about him/herself and all other family members;
- the request that respondents prepare for the interviews by keeping a study calendar and saving records such as bills and insurance statements in order to improve the accuracy of their reporting of medical use and expenditures;
- special requests for information in addition to the interview itself, such as completion of self-administered forms;
- requests for signed permission forms from specific sample persons authorizing the collection of data from medical providers, employers and other health insurance providers to supplement and validate the data obtained from households.

In the context of the core NMES design summarized above, respondent incentives have been used primarily for the following reasons:

1) to motivate respondents to participate initially and in future interviews in order to minimize initial nonresponse and panel attrition.

2) to compensate respondents fairly for the burden associated with long interviews and the completion of additional survey tasks; and

3) to motivate respondents to keep records and provide fair compensation for the effort required to maintain the study calendar and save financial records over a long period of time.

The respondent incentives most frequently used in NMES surveys consist of cash

payments made by the interviewer at the end of each interview, typically in the form of a check. Checks provide a safe and convenient mode for interviewers to handle cash payments. In NMCES respondents were paid $5 at the end of each interview, beginning with Round 1. The amount was increased to $10 per interview in NMES2, and to a proposed $15 per completed interview in NMES3.

NMES respondents are required to sign a receipt acknowledging that money was received. In NMES2, the receipt was used as a vehicle to consolidate commitment to the role of respondent in the survey. The receipt the respondent was asked to sign included a statement indicating the willingness to accept responsibility for record-keeping in preparation for the next interview.

The use of incentives has no doubt contributed to the high response rates achieved in NMES studies, in spite of the burden that long interviews represent, and notwithstanding the oversampling of groups that in many surveys yield lower than average response rates, such as poor people and minorities. The overall response rate for the NMES2 household sample was 80.1 percent after four rounds of data collection.

Incentives are an important tool used by interviewers to convince reluctant respondents to participate in NMES surveys. Recent methodological research examined the characteristics of persons who had initially refused to be interviewed in any one of the NMES2 rounds of data collection[24]. The analysis revealed that reluctant respondents differed significantly from their cooperative counterparts with regard to the proportion of overall medical expenditures that different sources of payment covered for each group. The reduction of significant differentials with respect to health expenditures and insurance coverage, two core analytic concerns of NMES surveys, provide evidence of the beneficial use of incentives to guard off against potential nonresponse bias in national estimates.

In 1985, to aid in planning for NMES2, a feasibility study[25] was conducted to investigate a broad range of methodological issues. Among the issues examined was

---

[24] Cohen, S.B. and B.L. Carlson (1992). "An Analysis of the Characteristics of Reluctant Respondents in the National Medical Expenditure Survey". Proceedings of the Section on Social Statistics, American Statistical Association, in press.

[25] Mathiowetz, N.A. and Ward, E.P. (1987). Linking the National Medical Expenditure Survey with the national Health Interview Survey: An Analysis of Field Trials, Vital and Health Statistics, Series 2, No. 102, U.S. Government Printing Office. Washington, D.C.

the effect of several incentive protocols[26]. The feasibility study consisted of two rounds of data collection: 1) a personal interview 45-60 minutes long where the respondent was paid five dollars at the end of the interview, and 2) a second interview in person or by telephone where again the respondent was paid five dollars at the completion of the interview. Approximately two weeks prior to the second interview, self-administered questionnaires were mailed to all respondents. These questionnaires were designed to take approximately 30 minutes to complete and included some moderately threatening questions on health behavior and mental health status.

Reporting groups were divided into three treatment groups: 1) "Prepayment" - persons were sent a five dollar check with the questionnaire; 2) "Promised" - persons were told that they would be paid five dollars when the completed forms were returned; and 3) "No mention" - persons were not given any information on payment, but were paid five dollars upon questionnaire completion.

The results indicated that the prepaid incentive leads to a significant improvement in response rates. Seventy-three percent of those in the prepaid group completed the survey compared to 66% among those who were not told of the incentive. The prepaid incentive also worked better than the promised incentive, which resulted in a response rate of 60%. Item nonresponse rates were calculated for each completed self-administered questionnaire and were used as a general measure of data quality. The finding was that prepayment leads to lower item nonresponse. Ninety percent of those who were prepaid answered all of the questions in the 18 page questionnaire, compared to only 74% in the promised group and 87% in the no mention group.

The study concluded that prepaid incentives can result in higher response rates and more complete data with less need for follow-ups. These were achieved in this study at a very moderate increase in cost. However, the net added costs may be far less than the value of the incentive payments, since a substantial part of the incentive costs is offset by savings in the follow-up activities. This finding was in line with results from other mail surveys, and that mode of payment was adopted in NMES2 when self-administered forms were mailed to sample households.

Based on results from the most recent NMES feasibility study carried out in 1992, the schedule of cash incentives in NMES3 will be modified relative to earlier surveys. Instead of introducing record-keeping tasks and paying respondents for the first time at

---

[26] Berk, M.L., Mathiowetz, N.A., Ward, E.P. and White, A.A. (1987). The Effect of Prepaid and Promised Incentives: Results of a Controlled Experiment. Journal of Official Statistics, Vol. 3, No. 4, pp. 449-457.

the end of Round 1, the 1992 Feasibility Study presented the family respondent with the study calendar at the end of the Screener round and prepaid these respondents in anticipation of the time and effort that would be devoted to keeping records in order to prepare for the Round 1 interview. Payments for later rounds of the Feasibility study were also described as compensation for future effort. In the last round a token gift (a commemorative tile with the U.S. Public Health Service seal) was given to respondents instead of cash payment. Under the new plan, the gift is the only net increase in incentive costs compared to previous plans, and it appears that prepayment has advantages. The rates of NMES calendar use in Round 1 of the Feasibility Study are significantly higher than the rates achieved in NMES2 in the round immediately following the round when payment and instructions to keep records were first delivered[27].

The 1992 NMES Feasibility Study also tested successfully the use of incentives to motivate respondents to complete a complex data collection task that was time dependent. This involved the procurement of health insurance printed materials that included a description of the benefits associated with the health plan offered by the employer to each policyholder in the household[28]. In NMES2 the collection of comparable information was attempted from employers in the course of the health insurance provider survey, but the time lag between the end of the household survey and the start of that provider survey frequently made it impossible to locate the necessary information about the health plan in effect at the time the household was interviewed.

In the Feasibility Study, respondents were offered $15 per household to contact employers, either by mail or in person, and secure the necessary information. Interviewers gave respondents a request form that could be presented to the employer to facilitate the task. Payment was made when the health insurance materials were delivered in the next round, and the amount was not increased in the event that the family had more than one eligible policyholder. Health insurance booklet requests in the Feasibility Study were followed up in later rounds and, by the end of the study, a policy booklet had been retrieved for 75 percent of eligible plans, at a lower cost and

---

[27] Sanchez, M. E. (1993). "Enhancing Compliance with Record-keeping in a Household Survey". 1993 Proceedings of the Section on Survey Research Methods of the American Statistical Association , Vol. 2, p 1015-1020.

[28] Emmons, C.A., Curno, M., and Smith, K. (1993). Final Report on the Outcomes of the Procedure for Obtaining Health Insurance Policy Documents from Respondents in the NMES3 Household Survey Feasibility Study. Submitted by NORC and Westat, Inc. under contract requirements.

on a more timely basis than in NMES2.

The use of remuneration to improve survey scheduling, achieve self-response and record keeping, and establish a commitment for a longitudinal data collection process are shown in the NMES experience. The NMES experience also demonstrates the potential for improved data quality with remuneration.

Interviewer Incentives

The morale and motivation of interviewers and supervisors are important factors in response rate outcomes for sample surveys. While the impact of respondent incentives has been frequently discussed in the survey research literature, there is little evidence of systematic inquiry into the use of incentives for interviewers as a means of achieving high response rates.

Typically, studies resort to interviewer incentives in a haphazard and improvised fashion when production levels have tumbled and the study response rate is deemed unacceptably low. This plan of action may not be as desirable or as economical as the notion of setting up a planned and carefully crafted incentive plan for interviewers from the very start of the project. The experience in the 1992 NMES Feasibility Study with such a plan suggests the desirability of exploring further the manner in which interviewer incentives may be manipulated to achieve gains in field response rates and efficiency within acceptable budget limits.

The Feasibility Study included a plan for interviewer incentives in order to achieve high response rates within the schedule for data collection. The project staff and the contractor collaborated to come up with a plan that was acceptable to all. The field staff wanted a plan that would promote and reward team work as opposed to individual performance; thus, the team was defined as the cluster of interviewers working in each PSU.

Realistic response rate levels for two points in time during the round (a specified midpoint in the field period, and the end of the round) were defined for each PSU and communicated to interviewers at the beginning of each round. The incentives were cash payments (about $25 for each of the time points in a round) paid equally to all PSU interviewers provided the PSU had achieved the targeted response rate by the specified date. A very modest additional amount of money was paid for increases in the response rate beyond the specified minimal rate.

The bonus plan encouraged interviewers to talk to their fellow PSU interviewers to

coordinate the steady flow of work. A weekly memo informed interviewers of the progress in other areas and the ranking of their own PSU. With very few exceptions, the staff in each of the PSUs achieved the response rate goals consistently for a very modest investment.

The benefits of the incentives included: teamwork among the staff at the local PSU level; steady production which avoided the last minute dislocations and expenses typically associated with a late push to increase response rates; and availability of qualified staff towards the end of the study to handle difficult assignments in a planned fashion.

More systematic research on the use of interviewer bonuses and the performance of different bonus plans is desirable, as the strategy appears a cost-effective way of obtaining high response rates.

Remuneration can be a valuable tool for direct interviewer management as well as helping interviewers motivate respondents to participate.

Physicians are often surveyed to obtain a wide variety of medical information. The high frequency at which physicians are surveyed coupled with the problem of "gate keepers" probably contribute to the low response rates typically achieved in physician surveys. However, prepaid incentives seem especially effective for this population group[29, 30, 31, 32]. The following three experiences serve as examples.

---

[29] Berry, S. H. and Kanouse, D.E. (1987). Physicians response to a mailed survey: An experiment in timing of payment. Public Opinion Quarterly, 51, p. 102-114.

[30] Lockhart, D.C. (1991). Mailed surveys to physicians: The effect of incentives and length on the return rate. Journal of Pharmaceutical Marketing & Management, 6(1), p. 107-121.

[31] Mizes, J.S., Fleece, E.L., and Roos, C. (1984). Incentives for increasing return rates: Magnitude levels, response bias, and format. Public Opinion Quarterly, 48, p. 794-800.

[32] Berk, M.L., Edwards, S.E. and Gay, N.L. (1993). The use of a prepaid incentive to convert nonresponders on a survey of physicians. Evaluation & The Health Professions, Vol. 16, No. 2, pp.239-245.

NATIONAL SURVEY OF DIAGNOSTIC ALLERGY TESTING

In 1988 a remuneration experiment was conducted[33] on a subset of physicians selected to participate in the National Survey of Diagnostic Allergy Testing (sponsored by the Health Industry Manufacturers Association).

A sample of physicians was randomly divided into three experimental groups. The first group received a $10 incentive with the first mailing. Nonresponders to the initial mailing were sent a new questionnaire as well as a letter urging them to respond and mentioning the $10 incentive they had received earlier. The second group of physicians did not receive a monetary incentive with the initial mailing. On the second mailing of the questionnaire, however, they received another letter explaining the importance of the study as well as a $10 prepaid monetary incentive. No mention of payment was made to the third group on either the first or second mailing.

The results indicate the use of a prepaid incentive has a dramatic impact on the response to the initial mailing. Fifty-five percent of those physicians receiving a prepaid incentive responded to the initial mailing, compared to less than 20% who were not told about payment on the initial mailing. Overall, a 63% response rate was obtained for Group 1 physicians (prepaid incentive with the initial mailing), compared with only a 50% response rate for Group 2 (prepaid incentive on the first prompt), and a 40% rate for Group 3 (no incentive).

This study concluded that incentives should be used in cases in which its use is considered necessary to obtain adequate response rates. Also, while delaying the decision to use an incentive until the second wave of mailing enables the researcher to decide whether an adequate response rate is likely to be obtained without payment, the incentive is not nearly as effective when used in a follow-up mailing. Because obtaining high response rates on physician surveys is difficult, few researchers will be able to conclude at study onset that a high response rate can be obtained. The use of a prepaid monetary incentive enclosed with the initial questionnaire mailing, therefore, appears to be a cost-effective method for improving response rates on physician surveys.

NATIONAL HOSPITAL DISCHARGE SURVEY

Most of this paper has dealt with remuneration for individual respondents, but

---

[33] Op cit. Berk et al., Evaluation & the Health Professions, 1993

institutions and organizations are sometimes respondents for surveys, and the National Hospital Discharge Survey (NHDS) experience suggests that remuneration is essential in some institutional surveys.

The National Hospital Discharge Survey (NHDS), conducted by NCHS, was first fielded in 1964 following the completion of a feasibility study[34]. The NHDS is a continuing study designed to provide comprehensive general-purpose statistics on morbidity in patients discharged from the Nation's general and short-stay hospitals. The principal source of information for the survey is the medical record in the hospital. The data are obtained from probability samples of medical records abstracted in a sample of general hospitals. Hospitals are compensated for participation in the survey.

Some hospitals in the sample have automated records, and contract with an abstract service. Other hospitals abstract records manually as needed. There are two manual data collection procedures: a primary manual procedure in which hospital staff complete the abstracts, and an alternate procedure in which a Census Bureau representative completes the abstracts. Hospitals using the primary manual procedure receive an average of $2.40 per abstract submitted; those using the alternate manual procedure receive about $1.00 per abstract submitted. Data tapes of uniform abstracts covering all discharges for automated hospitals are purchased directly from abstract service organizations. The cost of these data ranges from $.003 to $.055 per discharged patient. These discharges are sampled for the survey. All hospitals participating in the NHDS are reimbursed $1.00 per record biannually for approximately 40 records that are reabstracted for quality control procedures.

A large part of the success of the NHDS depends on the willingness of the hospitals to perform substantial continuing activities. Once inducted into the survey, hospitals participate for an extended period of years. A substantial amount of work is involved, including sampling the discharge lists, pulling and refiling medical records, and abstracting approximately 20 records monthly. A feasibility study conducted in 1963-1964 found that most hospitals expected compensation for their effort.

Cost analysis supports the practice of remuneration, particularly for primary manual procedure hospitals. For example in Fiscal Year 1992, it cost the NHDS an additional average of $12.15 per record to have the Census Bureau sample and abstract the data

---

[34] Brown, A.M., Altman, I. and Thompson, D.J. (1966), Participation of Hospitals in the Pilot Study of the Hospital Discharge Survey, Vital and Health Statistics, Series 2, No. 19, U.S. Government Printing Office. Washington, D.C.

in the alternate procedure hospitals. If the Bureau of the Census performed this work in all the sample hospitals, the cost of the NHDS would be substantially increased.

The feasibility study found that "While some hospitals indicated they might be willing to collaborate in the survey without reimbursement, it was clear that most would expect some compensation for their contributions, especially where it was felt that overtime work or the employment of additional personnel might be required. The ... form shown to the [hospital] administrator ... was quite detailed and implied a fairly exhaustive review of the medical record. It contained questions on the characteristics of the patient and his *{sic}* hospitalization, including final diagnoses, operations, complications, laboratory tests, therapies, and the like. There were some differences about whether payment should be made to the hospital or to the personnel doing the work, but most administrators favored payment to the hospital."

The feasibility study recommended that "... a uniform policy be adopted for the compensation of hospitals and that fair payment, based on further examination of the true cost to the hospitals be made." A subsequent pilot study was used to confirm the acceptability of the survey procedures , including remuneration, and helped to calibrate the payment amounts.

The NHDS experience illustrates the testing and use of remuneration and their use to offset a financial disincentive, improve the acceptance of a government survey, provide evidence of the value of participation, and keep expenses down in a survey requiring the continuing participation of institutions.

Summary and Conclusions

The evidence summarized in this paper shows that remuneration for respondents can be an effective technique for raising response rates and data quality when otherwise good survey practices are not sufficient.

More and more in the last decade, policy makers and program administrators are demanding data that cannot be supplied with a standard survey -- the one-hour cross-sectional interview containing nonsensitive questions is no longer the norm. The health surveys reviewed here each have one or more features that do not fit the public's perception of what a standard survey is: some require long interviews (NSFG Pretest and NMES), others require the maintenance of records such as diaries (NMES) or event histories (NSFG); some are panel surveys with repeated interviews (NSFG, NHANES Follow-up, NMES); some have sensitive questions (NSFG Pretest, NHSS, and YRBS); others use non-home sites (NHANES and the NSFG Pretest); or ask for

medical tests (NHANES); or require information or testing that could be embarrassing (NALS, NHANES, and NHSS).

As interviews get longer and questionnaire content gets more difficult or intrusive, and hard to interview sub-populations are surveyed, the need to motivate respondent participation grows. We need to provide respondents with concrete evidence of our appreciation and the importance of their participation and willingness to provide accurate and complete information. The careful use of remuneration allows us to offer people more than the promise that policy makers including Congress will use the data to improve their lives. If an advance letter or first personal contact explains that they will be compensated, all of the practical evidence reviewed here suggests that completing a quality interview at a reasonable cost is more likely.

Although most of this paper has dealt with remuneration for individual respondents in households, it is important to note that institutions, medical professionals, and other organizations are frequently respondents for health surveys. The NHDS experience suggests that remuneration is important in some institutional surveys also.

The history of remuneration in health surveys as evidenced by the experience reviewed here is a successful one. Remuneration has stood the test of time, proving successful in controlled experiments, field trials, and long-term implementation. For relatively little cost, important improvements in response and data quality have been gained using remuneration methods.

THE USE OF INCENTIVES TO SURVEY "HARD-TO-REACH" RESPONDENTS: A BRIEF REVIEW OF EMPIRICAL RESEARCH AND CURRENT RESEARCH PRACTICE

Richard A. Kulka
National Opinion Research Center
University of Chicago

1. Introduction

Incentive payments to survey respondents have been used extensively for many years as a means of improving survey response rates (cf. Shuttleworth, 1931), and there is considerable research evidence supporting the value of compensation for increasing cooperation and improving the speed and quality of response in a broad range of data collection efforts (cf. Kulka, 1992; Willimack, Petrella, Beebe, and Welk, 1992). In particular, a large number of empirical studies concerned with increasing response to mail questionnaires consistently attests to the effectiveness of monetary incentives in increasing mail survey response rates (e.g., Armstrong, 1975; Church, 1993; Duncan, 1979; Fox, Crask, and Kim, 1988; Harvey, 1987; Heberlein and Baumgartner, 1978; Hopkins and Gullickson, 1992; Kanuck and Berenson, 1975; Linsky, 1975; Yammarino, Skinner, and Childers, 1991; Yu and Cooper, 1983).

Nevertheless, until fairly recently monetary incentives and other forms of respondent remuneration have not been used extensively in general survey practice, especially in studies under government sponsorship and large scale academic surveys--as opposed to their widespread and common use in commercial or market research. In recent years, however, it has become increasingly difficult to achieve response rates high enough to provide statistically valid results, and remuneration has become more common. In the United States, Federal statistical surveys cannot employ incentives without explicit authorization from the Office of Management and Budget (OMB), and OMB has generally prohibited the use of payments to respondents, except under circumstances where "substantial need" can be demonstrated. And, interpreting this rule on a case-by-case basis--as requests for the use of incentives from Federal agencies have become increasingly common--has made it more difficult for OMB to assure consistency in the application of these guidelines.

To assist OMB in developing appropriate principles and decision rules governing the use of respondent incentives by Federal agencies, in October, 1992, the Council of Professional Associations on Federal Statistics (COPAFS) convened a symposium of representatives of government, business, academic, and research organizations to consider the current state of experience, research, knowledge, and opinion regarding the use of such incentives (COPAFS, 1993). Although it was not expected that the symposium would provide definitive answers to the multitude of

questions surrounding the use of incentives, "OMB expected to obtain information that would help in preparing guidelines to foster greater consistency in reviewing future requests by Federal agencies to use incentives when conducting surveys" (p. 1).

Toward that goal, symposium participants discussed, in part, "the kinds of survey situations in which incentives have a high probability of being effective or necessary" (p. 8), articulating "a set of circumstances in which they thought OMB should seriously consider an agency's request to use incentives" (p. 9). Among these were a number of situations or circumstances that might be broadly conceived as under the general rubric of "hard-to-reach respondents." Specifically, the list included the use of incentives:

- To encourage *hard core refusals* to respond, especially in small subpopulations of interest where response rates are often quite low--low enough to raise serious questions about the quality of survey data for these subpopulations.

- When there is a significant likelihood that a *"gatekeeper"* will prevent the respondent from ever receiving the questionnaire or otherwise make it difficult to make contact with certain segments of the population to conduct an interview.

- When there is a special target population for whom our conventional means of motivation or encouragement will have little if any chance of working--i.e, where the positive forces to cooperate are quite low (e.g., prostitutes, the homeless, the disenfranchised).

- If the target population is a small group that is *often surveyed*, such that a particular respondent is likely to be sampled frequently for one survey or another (e.g., physicians, CEO's, university deans).

- If the population is a *control group* in a program evaluation or experiment in which it is imperative to achieve and maintain an adequate response rate in the control group sample if the integrity of the study is to be maintained.

More generally, one of five potential OMB incentive policies suggested for consideration in nonstandard survey situations was that incentives "be considered if the respondent incurred out-of-pocket cost; or if the survey was too intrusive; or the survey was aimed at a *hard-to-reach population* [emphasis added]" (COPAFS, 1993, p. 12). Overall, most of those present felt that:

> hard-to-reach really meant *hard-to-interview* [emphasis added]. This category could include those who are hard to encourage to cooperate, and therefore initially refuse. In such cases, incentives might be effective. [However], participants felt that incentives would not be effective for those who are hard to find.
>
> Participants also included in the hard-to-interview category those who are difficult to reach by mail, those who must be kept in a sample (such as members of a control group), and those disenfranchised from society. (p. 12).

Based on this broad conception, the focus of this paper is explicitly on the use of incentives to survey hard-to-reach respondents, in contrast, for example, to the use of respondent incentives as a reimbursement for out-of-pocket expenses, as a payment to respondents for their time and effort in participating in a survey, or to compensate respondents for carrying out survey tasks that entail unusual demands, i.e., those which are especially burdensome or intrusive or may put the respondent at risk. In principle, to the extent that "hard-to-reach" is viewed as synonymous with "hard-to-survey," this focus is quite broad, in that encouraging those who might otherwise be reluctant to cooperate with a given survey to indeed do so is, in essence, the basic intent of the vast majority of surveys that choose to provide remuneration to survey respondents as part of their design.

However, incentives provided to stimulate survey response are rarely given only to initial or hard-core refusals, and several important questions regarding the use of incentives to encourage response may be addressed by focusing on this particular use of remuneration (cf. COPAFS, 1993). These include:

- Are there indeed specific target populations who are routinely offered remuneration to participate in surveys by most or all survey organizations because they are regarded as especially difficult to survey?
- While respondent incentives may increase cooperation among initial refusals, are they really effective with hard-core refusals and the truly difficult or impossible to interview populations?
- Are incentives effective only for certain target populations or subpopulations or more effective for certain population subgroups than for others (i.e., are the effects of incentives different for different population subgroups)?
- Are incentives really effective in getting past "gatekeepers," either for certain professionals (e.g.,

physicians) or other difficult-to-survey subgroups of the general population?

- Are incentives indeed of little use in locating and interviewing hard-to-find cases?

- Should consideration be given to paying some, but not all respondents to a given survey?

- Should all respondents be paid the same incentive, or should consideration be given to different levels or types of remuneration for different respondents?

To seek possible answers to these and some other related questions, we conducted a focused review of the current research literature on the use of incentives, with particular attention to their use with hard-to-reach populations, broadly defined.

2. Current Research Practice

However, because we anticipated that empirical evidence bearing on many of these questions would likely be quite sparse, we also sought input from individuals at most of the government, business, academic, and research organizations represented at the COPAFS symposium in October 1992, along with a few others, to ascertain the current state of survey research practice with regard to the use of incentives to survey hard-to-reach populations. In addition to providing citations or references to any papers or publications related to this topic, each organization was asked to provide information on any recent surveys they had conducted with such populations, indicating when they had or had not used incentives, and a sense of their general organizational policies or conventions regarding the use of incentives under such circumstances. Prior to describing some of the evidence available from the research literature that bears on the questions raised above, it will be useful to summarize current practice in this area as described by these organizations.

Not surprising, the vast majority of these organizations routinely conduct surveys with hard-to-reach respondents under our broad definition. As noted by one organizational respondent, virtually every survey encounters and must deal with hard core refusals, but more specific categories of respondents designated by responding organizations as hard-to-reach are:

(1) the economically disadvantaged (e.g., lower income or lower socioeconomic status (SES) populations, welfare recipients or applicants, the homeless);

(2) the educationally disadvantaged (e.g., the less educated, high school dropouts, those with low literacy levels, the mentally retarded);

(3) minority populations (e.g., African Americans, Hispanics, disadvantaged minorities, impoverished urban minorities);

(4) adolescents, youth, and young adults (e.g., youth in general, minority youth, young black males, teen mothers, the young and mobile);

(5) drug users and those with special health problems (e.g., current or former drug users, drug abusers, cocaine users, diabetics, those with asthma);

(6) frequently surveyed professional or elite populations (e.g., physicians, nurses, CEO's, teachers, college and university faculty, both very small and very large farm operations); and

(7) transients and persons who wish not to be found for legal or other reasons (e.g., highly mobile and transient populations, runaway youth, absent parents owing child support, those defaulting on student loans).

Not only do these categories overlap a great deal, but also, in almost every case, incentives have been used in surveys of these populations to increase response rates by at least one organization, and quite often by many. For example, the use of (generally substantial) incentives in surveys of physicians is a standard practice in virtually all of these organizations. It is also the case, however, that recent surveys have been conducted with most of these subpopulations in which *no incentives* were used, including a few surveys of physicians.

Although most of the organizations queried feel that respondent incentives are generally effective in increasing response rates among these hard-to-reach target populations, very few controlled or randomized experiments have been conducted to demonstrate empirically the efficacy of incentives in improving response rates under such circumstances. Not surprisingly, the use of incentives in surveys conducted by or for Federal statistical agencies--which require OMB approval for providing respondent incentives--is somewhat more likely to be based on such empirical evidence than their use in surveys conducted by commercial, academic or private research firms under other auspices.

Even in the absence of such controlled experiments, several of these survey research professionals and firms believe that respondent incentives are an important overall tool in their arsenal for dealing with hard-to-interview survey populations, and

their experiences with such incentives bear on at least three of the basic issues raised earlier regarding the use of incentives with such populations. First, although many participants at the COPAFS (1993) symposium felt that incentives might not be effective in locating those who are hard-to-find, several of these organizations report experiences that suggest that paying respondents makes contacting and locating easier, less expensive and more effective, since contact individuals are more willing to convey messages and provide new address and telephone numbers for sample members when interviewers are able to mention that they have a monetary incentive for the respondent. Similarly, others cite experiences suggesting the efficacy of respondent incentives in "opening the door" or getting past "gatekeepers,"--i.e., in helping gain access to the respondent--because nurses, receptionists, relatives, friends and other "gatekeepers" are apparently more reluctant to restrict or deny access to a potential respondent when a monetary incentive is involved.

Third, although several of those responding to our inquiry expressed some ambivalence regarding this practice, a number of surveys have been conducted which provide incentives either to some but not all respondents, or different levels of remuneration to respondents in the same survey. In some cases, these different incentive levels (including no incentive) reflect different levels of burden, effort or risk for different respondents, but it is also not uncommon to offer incentives only to sample members for whom one is having difficulty getting them to respond--i.e., to do the survey without routine remuneration, but then use monetary incentives to try to convert refusals. Alternatively, in a survey providing incentives at a given level, interviewers might be permitted to offer increasingly larger amounts to convert increasingly hard-core refusals, to persuade extremely hard-to-convince cases to indeed cooperate.

Although these conditional incentive approaches can be quite cost effective, paying uncooperative sample members when cooperative respondents are not paid, or paying especially reluctant or difficult sample members more than those who cooperate more readily, violates our sense of fairness or equity. However, the selective or strategic use of remuneration to convert hard-core refusals and achieve higher or very high response rates is--though relatively rare and practiced with some reluctance--very much a part of current research practice with regard to the use of incentives to survey hard-to-reach respondents.

3. A Brief Review of the Literature

Having briefly summarized the current state of practice in this area, let us now explore what, if anything, the current research literature can tell us regarding the use of respondent incentives with difficult-to-survey populations or sample members.

3.1 The Use of Incentives in Surveys--A Summary

As a background for that analysis, it is important to consider momentarily what we know (or think that we know) about the use of respondent incentives in general (cf. Kulka, 1992, Willimack et al. 1992). As noted earlier, few today would question the general assertion that a monetary incentive enclosed with a mail questionnaire will serve to increase response rates. Hundreds of studies have been conducted, and review after review--both qualitative and quantitative--concludes that the importance of financial incentives is "second only" (perhaps) to the use of follow-up mailings or prompts in improving response rates (cf. Dillman, 1991). Moreover, the literature rather overwhelmingly supports the predominant effectiveness of prepaid as opposed to promised incentives (e.g., Armstrong, 1975; Berk, Mathiowetz, Ward and White, 1987; Berry and Kanouse, 1987; Blumberg, Fuller, and Hare, 1974; Church, 1993; Furse and Stewart, 1982; Hopkins and Gullickson, 1992; James and Bolstein, 1992; Kanuck and Berenson, 1975; Linsky, 1975; Peck and Dresch, 1981; Skinner, Ferrell, and Pride, 1984; Wotruba; 1966; Yu and Cooper, 1983). That is, incentives appear to be most effective in inducing survey response when they are paid in advance--at the time that the respondent's cooperation is initially solicited--rather than offered conditional on and paid subsequent to respondent cooperation, even when the promised or conditional incentive is greater than the amount prepaid (cf. Linsky, 1975; James and Bolstein, 1992).

The use of monetary incentives to increase response rates for telephone and personal interview surveys has received far less research attention, although the results of several studies are consistent with those derived from mail surveys. Overall, however, the conditions under which a monetary incentive will be effective or ineffective under these survey modes appear to be less general. Based on this more limited research literature, the greatest potential effectiveness of monetary incentives appears to be in surveys that place unusual demands upon the respondent, require continued cooperation over an extended period of time, or when the positive forces on respondents to cooperate are fairly low (cf. Cannell and Fowler, 1977).

In addition to a potential beneficial impact on response rates, the research literature suggests that incentives may have a beneficial impact on *data quality* as well. At least two theories suggest the opposite--a *detrimental* effect of remuneration on data quality:

(1) a concept based on "social desirability" theory (Cannell and Henson, 1974; Weiss, 1975) that suggests monetary inducements will increase the tendency of participants to try to please the interviewer by providing what the respondent believes is the desired or "correct" answer; and

(2) a "self-perception" model that argues that the introduction of financial incentives acts as an *external* motivator, thereby reducing the degree of internal motivation (i.e., the subject's interest or desire to participate), decreasing the degree or *quality* of compliance (i.e., quality of response), while increasing the *rate* of compliance or cooperation (cf. Hansen, 1980).

While relatively little empirical evidence has been found in support of either of these two models (see, however, Hansen, 1980; James and Bolstein, 1990), the preponderance of evidence reported to date (e.g., Berk et al., 1987; Cowan, 1977; Ferber and Sudman, 1974; Godwin, 1979; Goetz, Tyler, and Cook, 1984; Houston and Ford, 1976; James and Bolstein, 1990; Kerachsky and Mallar, 1981; McDaniel and Rao, 1980; Sudman and Ferber, 1974) is more consistent with a theory based on "social exchange," which posits that the offer of monetary incentives induces a greater commitment to the survey task among respondents, which in turn results in better data quality from survey respondents.

With regard to incentive size, the research literature is significantly less helpful, since the majority of studies have investigated the effects of incentives of $1 or less (e.g., Armstrong, 1975; Fox et al., 1988; James and Bolstein, 1992; Jobber and Saunders, 1988; Kanuck and Berenson, 1975; Linsky, 1975; Yammarino et al., 1991; Yu and Cooper, 1983), and few studies have successfully demonstrated the effectiveness of very large monetary incentives (e.g., Berry and Kanouse, 1987; Gunn and Rhodes, 1981; James and Bolstein, 1992). At both extremes, there is some evidence that increasing the size of monetary incentive will result in increases in survey response and/or quality (e.g, Armstrong, 1975; Findlay and Schaible, 1980; Fox et al., 1988; Furse and Stewart, 1982; Godwin, 1979; Gunn and Rhodes, 1981; Hubbard and Little, 1988a, 1988b; James and Bolstein, 1990, 1992), but there is also rather consistent evidence that this benefit may rather quickly reach "diminishing returns," whereby larger incentives no longer result in appreciable increases in survey response (e.g., Armstrong, 1974; Fox et al., 1988; Godwin, 1979; Hubbard and Little, 1988b; James and Bolstein, 1992; Mizes, Fleece, and Roos, 1984).

Overall, why payments are effective in improving response rates is not currently very well understood. Some believe that "payment works in increasing response rates . . . through motivating and supporting the *interviewer* [emphasis added] in his [or her] approach to the respondent' (Weinberg, 1973, p. 480), while others view "incentives . . . as impressing upon the survey *respondent* [emphasis added] the importance of the task" (Goetz et al. 1984:149; Berry and Kanouse, 1987). Still others have suggested that the predominant motivating power of an incentive is not its monetary value, but rather its symbolic, or "token" value (e.g., Linsky, 1975).

In turn, two basic classes of theories have been referenced to explain why incentives may increase survey participation (cf. Willimack et al., 1992). The first is based on the principle of "reciprocation" (Groves, Cialdini, and Couper, 1992):

> Every human society abides by a *norm of reciprocity* [emphasis added] that directs individuals to provide to others the general form of behavior that they have received from others (Gouldner, 1960). . . . [Based on] the reciprocity heuristic, . . . one should be more willing to comply with a request to the extent that the compliance constitutes the repayment of a perceived gift, favor, or concession. (p.480)

Thus, by providing an incentive as an unsolicited gift (e.g., a prepaid incentive), one invokes the norm of reciprocity among respondents, who can "reciprocate" by participating in the survey.

Closely related to the concept of a "norm of reciprocity" (and in the same basic class) are theories of "cognitive dissonance" and "social exchange." Under the former, it is postulated that the inclusion of a prepaid token incentive with a request for survey participation creates psychological dissonance, which is most easily resolved by consenting to in fact participate (cf. Furse and Stewart, 1982, 1984; Hackler and Bourgette, 1973).

As articulated by Dillman (1978), "social exchange" theory, which emphasizes the perceived costs and rewards of responding to a survey, suggests that, in order to maximize survey response, one must "minimize the costs for responding, maximize the rewards for doing so, and establish trust that those rewards will be delivered" (Dillman, 1978, p. 12). Rather than serving as a reward for survey participation, the use of an incentive serves as "a symbol of trust," a major factor necessary for social exchange to successfully occur. Consistent with this notion is research evidence showing that

> increasing the size of an incentive does not always increase response, and in fact may tend to decrease it, and that including it with the appeal [a prepaid incentive] is more effective than promising to send it on return of the questionnaire. The closer the monetary incentive comes to the value of the service performed, the more the transaction tends to move into the realm of economic exchange [in which money serves as a precise measure of the worth or value of one's actions] and the easier it becomes for people to refuse it. (Dillman, 1978, p. 16)

In general, smaller, prepaid incentives appear to invoke social exchange or the norm of reciprocity, while larger promised or conditional incentives are more likely to invoke *economic exchange*, which represents the second basic class of theories on how incentives serve to increase survey participation, i.e., by

literally paying respondents for the time and effort required to provide the information requested. Under this model, remuneration represents reimbursement for survey cooperation (cf. Cohen, Walden, and Ward, 1992), i.e., compensation to respondents for their time and effort in participating in the survey, rather than an "incentive," "gift," or "gesture of goodwill" too small to be considered equitable payment or exchange for respondents' time.

In this regard, Cannell and Henson (1974) emphasize that, since respondents rarely share the goals of a survey, they do not consider participation as a means of advancing their own, personal goals, and are thereby generally unmotivated to perform the survey task. An incentive, in the form of a payment, may serve to provide a personal goal which motivates participation in a survey. To be effective, the amount of money offered must be large enough to be worth working for--i.e., the value of the incentive should be commensurate with the task and/or time sacrificed by the respondent--but not extravagant, because, if individuals perceive that they are overcompensated, the effects on participation may actually be negative (Cannell and Henson, 1974; Groves, 1989).

Overall, the preponderance of research evidence appears to favor the concepts of social exchange or reciprocity as a basis for the effectiveness of incentives, although with the advent of increasingly more complex surveys and the use of larger incentives, a significant body of evidence consistent with the tenets of economic exchange is also accumulating, much of it in relation to surveys of apparently hard-to-reach or hard-to-interview populations, to which we now turn our full attention.

3.2 Hard-to-Reach Target Populations

Although at least seven different categories of respondents were identified by survey practitioners as "hard-to-reach," research on the effects of respondent incentives has focused directly on only a few of these--frequently surveyed professional and elite populations, adolescents and young adults, and the disadvantaged--and the relative coverage of even these three broad populations in that regard is quite uneven.

By "focusing directly" I mean that the subjects for experimentation with incentives are drawn largely or entirely from one of these target populations. A number of these studies include one or more of these subgroups as a component of the population surveyed, and such subgroups may or may not have responded differently to incentives than other groups represented in the sample, but the issue of the differential effectiveness of incentives among those hard-to-survey will be dealt with in a separate section.

Studies on the use of incentives with professional and elite populations--especially physicians--are legion. Physicians are

widely believed to be an especially difficult population from which to collect survey data (Sudman, 1985), since they "are frequently approached for surveys, the demands on their time are great, and their office staffs are vigilant in protecting them" (Berry and Kanouse, 1987, pp. 102-103). Physicians' reluctance to participate in surveys is a growing problem for researchers (cf. Berk, 1985), with the American Medical Association (AMA) warning that "physicians are becoming weary and wary of surveys" (Martin, 1984), while the response rates to even their own surveys have declined precipitously (Goodman and Jensen, 1981).

Both prepaid (Berk, Edwards, and Gay, 1993; Berry and Kanouse, 1987; Lockhart, 1991; Mizes, Fleece, and Roos, 1984) and promised (Gunn and Rhodes, 1981; Weber, Wycoff, and Adamson, 1982; Tambor, Chase, Faden, Geller, Hofman, and Holtzman, 1993) monetary incentives have been shown to significantly improve response rates in surveys of physicians, whether conducted by mail, telephone or in person. For example, Gunn and Rhodes (1981) conducted an experiment to determine the effectiveness of paying monetary incentives to physicians for their participation in a 20-30 minute telephone interview on attitudes toward influenza immunization. Physicians were systematically assigned to one of three subsamples designated to receive no incentive, $25, or $50. Study findings revealed a 58 percent response rate for the group offered no incentive, 69 percent for those promised $25, and 77 percent for those offered $50. In a personal interview survey conducted by Weber and his colleagues (1982) the same range of incentive conditions resulted in response rates of 38, 67, and 73 percent, respectively.

Berry and Kanouse (1987) compared the relative effectiveness of a prepaid and a promised incentive of $20 in a mail survey of physicians, obtaining a 78 percent response rate for the prepaid incentive group and a 66 percent rate for those paid only after they completed the survey. Mizes, Fleece, and Roos (1984) demonstrated the effectiveness of even a relatively small prepaid incentive in a brief mail survey of physicians, obtaining a response rate of 74 percent with either a $1 or $5 prepayment in comparison with 53 percent when no payment was provided. In a survey of 600 physicians from three specialty groups, Lockhart (1991) achieved a 57 percent response rate using a $20 prepaid incentive, compared with only 13 percent in a no incentive control group. More recently, Berk, Edwards and Gay (1993) confirmed the effectiveness of a prepaid incentive of $10 in a mail survey of physicians, achieving a response rate of 63 percent for those receiving the incentive with an initial mailing in comparison with 40 percent for the no incentive group.

The relative effectiveness of both large and small monetary incentives in improving response rates, speed, and/or quality among

other professional and elite populations has also been demonstrated empirically, including:

(a) international elites, ranging from university professors to cabinet ministers (Godwin, 1979);

(b) nurses (e.g., Kephart and Bressler, 1958);

(c) librarians (Hopkins, Hopkins, and Schon, 1988);

(d) various professionals subscribing to a magazine dealing with alcohol and drug use problems (Goodstadt, Chung, Kronitz, and Cook, 1977);

(e) owners of small construction subcontracting companies (James and Bolstein, 1992);

(f) community elites (Paolillo and Lorenzi, 1984);

(g) business executives (Erdos and Morgan, 1983; Robin and Walters, 1976); and

(h) farmers (Willimack, 1993).

However, some of these studies were poorly designed and/or obtained very low response rates even with incentives, and there are other studies where incentives used with professionals were either ineffective (e.g., Cook, Schoeps, and Kim, 1985; Abraham and Johnson, 1993) or resulted in poorer response rates or quality than when no incentives were provided (e.g., Hansen, 1980). For example, in the 1992 field test for a national survey of college and university faculty (Abraham and Johnson, 1993), three incentive conditions were used, including one monetary (a prepaid $2 bill); only the monetary incentive approached statistical significance in improving the response rate over no incentive (87 versus 79 percent), and this was not regarded as strong evidence of the efficacy of incentives with that particular professional population. In fact, the main survey, fielded without incentives, achieved an overall response rate of 87 percent (Abraham, 1994).

A second category of respondents which many nominate as hard-to-reach are adolescents, youth, and young adults. While there is less empirical evidence available with regard to these target populations, the results that are available are consistent with the assumption that incentives can be quite effective in stimulating survey cooperation among them. For example, a recent investigation conducted for NCHS by the Survey Research Center at the University of Michigan (Cannell and Camburn, 1991) studied the effects of respondent payments of $20 on the willingness of youth 12-20 to participate in the Youth Risk Behavior Surveillance System (YRBS) and on their motivation to answer YRBS questions as accurately and truthfully as possible. The results of this research indicated that paying respondents increased participation rates (from 79 to 90 percent), reduced parental consent refusal rates (thereby assisting in getting by an important "gatekeeper), aided

interviewers in converting refusals, increased respondents' perceptions of the importance of the survey, and may have improved the accuracy and honesty of responses.

The authors suggest that these youth represent the classic hard-to-interview respondent, in that the positive forces for them to respond and respond accurately to the YRBS are in fact quite low. "Young people had a low interest in this survey, and saw no compelling reason for responding honestly or being diligent in the task of answering the survey questions" (Cannell and Camburn, 1991, p. 1). Moreover,

> respondents could see no compelling reason to be diligent in answering survey questions. Participants in group interviews voiced skepticism about any benefits that may accrue from participating in surveys. Therefore, techniques typically used with adult respondents to encourage participation and more accurate reporting, such as appealing to their obligations as citizens [civic duty] or emphasizing the societal benefits likely to come from the survey, may be unproductive with youthful respondents. Therefore, some different ways of motivating respondent participation and accurate reporting are needed. (Cannell and Camburn, 1991, p. 11)

This is consistent, of course, with Cannell and Henson's (1974) earlier suggestion that this is precisely the role that a respondent incentive might play, i.e., providing a personal goal which motivates participation in a survey.

Similarly, students and former students have traditionally been difficult to survey. For example, "postsecondary students who have little motivation to participate in . . .research and have traditionally been difficult to survey include those who fail to complete the application process, those who are admitted but do not register for classes, dropouts, and alumni" (Zusman and Duby, 1987, p. 73). Thus, in a mail survey of undergraduate transfer students who subsequently withdrew, Zusman and Duby (1987) found that a prepaid incentive of $1 increased cooperation by nearly 20 percentage points. In an earlier study of the use of incentives with a postsecondary student survey, Huck and Gleason (1974) found that the response rate could be increased from 65 to 92 percent with the provision of an incentive. Peck and Dresch (1981) found that a prepaid $3 incentive with a 1½ hour mail survey of men and women three years after they completed high school yielded a response rate of 76 percent, compared with a 68 percent response rate for those promised a similar payment, and 54 percent among those who were offered no payment at all. Similarly, in a one-year follow-up mail survey of vocational-technical school graduates, Pucel, Nelson, and Wheeler (1971) found that the use of multiple nonmonetary incentives increased response rates by over 20 percentage points relative to a no incentive control group.

In contrast, in the 1992 Postcensal pretest for the NSF's 1993 National Survey of College Graduates, a mail survey with telephone and in person follow-ups, Mooney, Giesbrecht, and Shettle (1993) found that a $5 incentive with the initial mailing significantly increased response rates after two mailings by 11 percentage points, but, after mail and telephone follow-up, this difference was reduced to only 2 percent. In addition to increasing speed of response, however, those provided an incentive were significantly more willing to provide telephone numbers and names of contact persons, thereby reducing the effort required for future locating in this longitudinal study. In yet another incentive experiment conducted in connection with the National Assessment of Educational Progress (NAEP) household survey --in which young adults 26-35 were asked to complete a series of tests--Chromy and Horvitz (1978) demonstrated the effectiveness of a variable incentive procedure (no incentive for one package, $10 for two, $15 for three, four for $20) in increasing the overall response rate from 70.5 percent (with no incentive) to 83.3 percent, a rate which was maintained in subsequent years by adopting this procedure.

Perhaps the respondent categories most commonly mentioned as hard-to-reach or hard-to-interview are the educationally or economically disadvantaged and minorities. However, there is very little experimental evidence available attesting to the efficacy of respondent incentives based specifically on these target populations. One of the classic examples in fact overlaps our previous category, dealing as it does with disadvantaged youth. In that study (Kerachsky and Mallar, 1981), a $5 payment per interview was offered to a randomly selected portion of a national probability sample panel of 5,800 economically disadvantaged youth for three waves of interviewing, conducted in conjunction with an evaluation of the economic impact of the Job Corps program. "Youth in the age range of Corpsmembers (16 to 21) and with their economically disadvantaged backgrounds are generally very mobile and hard to locate [and interview]" (Kerachsky and Mallar, 1981, p. 263).

The researchers were able to verify the effectiveness of these respondent payments for improving both the quantity (search efficacy and interview completions) and quality (e.g., item nonresponse) of responses. After one or two interviews, the study's ability to locate potential respondents and obtain data from those who were located deteriorated in the absence of monetary incentives, but not when sample members were offered the $5 payment per interview. In addition, payment influenced the willingness of sample members to return postcards from advance letters, thereby reducing the cost of locating respondents for follow-up interviews. Significant reductions in item nonresponse associated with these payments were most pronounced at baseline and declined over time. More generally, the overall effectiveness of respondent incentives in increasing response rates and quality in studies that

overrepresent the poor and minority populations has also been demonstrated (e.g, Berk et al., 1987).

While studies of the effectiveness of incentives that focus directly on low income, minority, and disadvantaged populations are quite rare, those that address the more specific question of whether respondent incentives are *more* effective in improving cooperation among such target populations than the more affluent and advantaged are considerably more common. This question is part of a more general one--are incentives more effective for certain population subgroups than for others?--to which we now turn our attention.

3.3 Differential Effects of Incentives by Target Population

Restating our original question on this issue:
Are incentives effective only for certain target populations or subpopulations or are they more effective for certain population subgroups than for others (i.e., are the effects of incentives different for different population subgroups)?

At the most general level, a few of the quantitative reviews of the research literature on techniques designed to enhance mail survey response rates have addressed this issue. Yammarino and his colleagues (1991) found, for example, that associations between incentives and response rates "were statistically significant but derived from more than one population; that is, the relationships are situation specific and there is a need to examine potential moderators" (p. 627), other than year of publication and (a crude measure of the) type of sample (consumer versus institutional groups), the two moderators included in the study, which were not statistically significant. In contrast, neither Church (1993) nor Hopkins and Gullickson (1992) found significant differences in the effectiveness of incentives by target population (general, students, technical, business, and medical) or population type (professional, general, or semiprofessional), respectively.

However, a more detailed examination of respondent populations reveals a significant amount of variation in responsiveness to respondent incentives. For example, in an experiment recently reported by Willimack (1993) in connection with the 1992 Farm Costs and Returns Survey (FCRS), a prepaid nonmonetary incentive increased the overall response rate from 58 to 63 percent, but the incentive proved to be most effective in the smallest and largest classes of farm operations, increases of 17 and 12 percentage points, respectively.

Similarly, several studies of physicians have found significant differences by specialty in the effectiveness of incentives in increasing survey cooperation. Noting that different populations are more receptive to certain incentives than others, Erdos and Morgan (1983) reported that doctors in nine specialties

of medicine responded well to a 25¢ incentive, whereas psychiatrists required a $1 incentive. In the study cited earlier by Gunn and Rhodes (1981), response rates varied considerably across specialties: pediatricians and industrial physicians had high rates regardless of payment category ($0-$25-$50), but general and family physicians were very sensitive to payment--37 percent responded with no payment and 64 percent with $50. In the Lockhart (1991) study, the impact of incentives on general and family practitioners and physicians specializing in internal medicine was substantially greater than that observed among diabetes specialists, presumably because the topic of the survey--blood glucose monitoring--was especially salient to the latter. In contrast, in Berry and Kanouse's (1987) investigation of prepaid versus promised incentives, prepayment was effective with all nine specialties (statistically significant in four), except for oncologists, whose overall response rate was exceptionally high.

While it is commonly assumed that monetary incentives are more effective in increasing response rates among less educated, lower income respondents than among more educated, middle- or upper-income persons or households, research findings are in fact somewhat mixed regarding the effect of incentives on response rates for groups differing in socioeconomic status. For example, based on the early mail survey literature, Kanuk and Berenson (1975) cite several studies that failed to show that low income people were more likely to respond to monetary incentives than were people with higher incomes. In the 1971 National Health and Nutrition Examination Survey (NHANES I; Miller, Kennedy, and Bryant, 1972; Bryant, Kovar, and Miller, 1975), a $10 incentive increased examination cooperation rates relative to no incentive from 70 to 82 percent, but, contrary to expectations, the effect of the promised incentive did not increase as income level decreased. In a parallel remuneration experiment conducted in connection with the second survey (NHANES II) in this series (Findlay and Schaible, 1980), boosting the incentive from $10 to $20 increased the overall response rate from 74 to 79 percent, and the increased remuneration was more successful among whites than blacks, but there were no differences by income level. Goetz and his associates (1984) also found no difference in the effectiveness of incentives by education, race (cf. Dohrenwend, 1970), and income.

In contrast, in a nonexperimental comparison, Benus and Ackerman (1971) found that response rates for all major segments of the population were better in a panel where sample members were paid than a comparable panel where sample members were not paid, and they were disproportionately better among low income respondents. Similarly, using nonmonetary incentives in a mail survey, Nederhof (1983) found that incentives produced a disproportionately larger percentage of respondents from lower educational and occupational strata. Though modest, James and Bolstein (1990) found a similar pattern of relationships for level of education and income with various monetary incentives in their

mail survey of cable television subscribers. Ferber and Sudman (1974) also found that monetary incentives were indeed more effective in soliciting cooperation from lower income families than high income households, but such differential effects by education or income level have not been consistently observed in consumer expenditure studies (cf. Ferber, 1974; Cowan, 1977).

In a widely cited but poorly designed study, Gelb (1975) reported a significant difference in the response of lower-class and middle-class respondents to a conditional versus an immediate incentive to return a questionnaire, with middle-class recipients responding better to a prepaid and lower-class recipients to a promised incentive. More recently, Goyder (1990) reported a similar statistical interaction between socioeconomic status (SES) and size and type of incentive, whereby higher SES sample members were more likely to respond to no incentive or a prepaid $1 incentive, while lower SES sample members were more likely to respond to a post-paid $10 incentive than high SES sample members.

A more rigorous test of the differential effectiveness of monetary incentives by socioeconomic status was conducted in connection with a field test of 2,000 adults 16 and older for the National Adult Literacy Survey (NALS), a personal interview survey designed to measure one's ability to use printed and written material (Berlin, Mohadjer, Waksberg, Kolstad, Kirsch, Rock, and Yamamoto, 1992). Assessing the impact of incentives of $0, $20, and $35, a significant increase was found in response rates in the two incentive groups (81 and 83 percent, respectively) over the "no incentive" group (73 percent), but no significant difference by incentive level. Of particular importance to the current discussion, they found that incentives were most effective in improving response rates for people with low educational attainment and minority populations. When a monetary incentive was paid, significantly more black and Hispanic adults agreed to take both the background questionnaire and literacy test than when no incentive was offered, and similar results were observed for adults with lower levels of education. Similarly, in a recent pretest for Cycle V of the National Survey of Family Growth (NSFG), a survey of women of childbearing age with an oversample of black and Hispanic women, a $20 incentive resulted in an overall increase in cooperation of 8 percentage points, but the increase was disproportionately higher for black, Hispanic, and low income women (Duffer, Lessler, Weeks, and Moser, 1994). For example, among lower income women, the $20 incentive increased response rates by 12 percentage points, three times the level of increase (i.e., 4 percentage points) observed among higher income women. Similarly, the increases observed among black and Hispanic women were 10 and 26 percentage points, respectively, compared to only 2 percentage points for white and other women.

Thus, there is indeed some appreciable evidence that incentives are more likely to influence lower income, lower

socioeconomic subgroups and minority populations to participate in surveys. Overall, however, results regarding the effect of incentives on survey participation from groups differing by race/ethnicity or socioeconomic or status are currently still quite mixed.

3.4 The Conditional Use of Respondent Incentives

The foregoing discussion raises (but clearly doesn't answer) the question of whether all respondents should be paid the same incentive, or should consideration be given to different levels or types of remuneration for different respondents (cf. Peck and Dresch, 1981; Marrett, Kreiger, Dodds, and Hilditch, 1992). A related question raised earlier in our discussion of current practice is "should consideration be given to paying some but not all respondents to a given survey?" In fact, as noted earlier, a number of survey research organizations on some surveys use respondent incentives only for refusal conversion, usually only as a last resort to convert hard-core refusals, the truly hard-to-interview sample members. For example, in the General Social Survey (GSS), conducted by NORC under a grant from the National Science Foundation, "respondent fees" have routinely been used for this purpose; and, in 1989, up front respondent fees were also offered to a select group of respondents who resided in traditionally difficult urban primary sampling units (PSUs), a strategy that resulted in a lowering of the overall cost of the survey and a reduction in the field period, while still achieving the highest completion observed for this survey since 1985 (Law, 1989).

In a paper prepared for the COPAFS symposium in October 1992, Pendleton and Ginsberg (1992) explicitly raised this as an important research question (cf. Tucker, 1992):

> [Information is needed on] the most advantageous stage in the data collection process to offer incentives to minimize the cost and time involved in repeated follow up. A comparison of the cost of completed interviews when no incentives are offered, when offering incentives at first contact, and when offering incentives only when response rates are not found to be adequate in terms of cost of completed interview would be useful. (p. 8)

And, although none of them deal with the equity or fairness issues associated with this practice, a few studies have indeed explored the issue of the timing of when respondent incentives are offered, independent of the prepaid versus promised distinction, i.e., the "conditional" use of incentives to convert nonresponders.

For example, in a survey of college undergraduates, Huck and Gleason (1974) found that a follow-up mailing sending a quarter to nonrespondents rather than to everyone on the initial mailout list

cut costs in half, without significantly decreasing the response rate. In essence, the NAEP national assessment experiment with young adults reported by Chromy and Horvitz (1978) was a nonresponse follow-up study. After obtaining a disappointing response rate of only 44 percent with no incentive, the experiment was directed only at nonrespondents from the initial study. While nonrespondents receiving no incentive, but exposed to substantially improved field procedures, were increased to a participation rate of 71 percent, the response rates for those receiving one of three incentive conditions were boosted to 80 to 85 percent. Similarly, in the experiment with young adults surveyed by mail by Peck and Dresch (1981), a prepaid incentive of $3 in the first wave mailing resulted in a final response rate of 76 percent, compared to only 54 percent among those receiving no incentive; however, a condition in which nonrespondents to the first mailing who initially received no incentive were prepaid $3 with the second mailing ultimately achieved a response rate of 70 percent. In a household health survey of young adults in Switzerland, Perneger, Etter, and Rougemont (1993) experimented with two "incentives," a promise of 10 Swiss francs ($7 US) and a red reminder postcard mailed two days after the questionnaire. Initial response rates were 65 percent for those receiving both incentives, 57 percent for those offered the money only, 54 percent for recipients of the reminder card, and 48 percent for those who received neither. Follow-up mailings with incentives sent to all nonrespondents resulted in final response rates of 83, 84, 82, and 78 percent, respectively, attesting to the efficacy of these incentives as a productive nonresponse follow-up device.

In a mail survey of residents of New Zealand with three waves of mailing, a $1 incentive achieved a final response rate of about 70 percent, regardless of whether it was sent to all respondents with the initial mailing or included in the second mailing for nonrespondents to the first mailout (Brennan, Hoek, and Astridge, 1991). Similarly, in a mail survey of San Diego residents, a monetary incentive of $5 contingent on response (i.e., a promised incentive) to a second mailing of the questionnaire increased the response rate from initial "nonresponders" by 100 percent relative to controls who received no incentive, and by 75 percent over those who received $1 not conditional on response (Spry, Hovell, Sallis, Hofstetter, Elder, and Molgaard, 1989).

In a general population telephone survey on family health insurance recently conducted in 10 states for the Robert Wood Johnson Foundation by Mathematica Policy Research, Inc., promised incentives of $5 and $10 were compared with no incentive in three states to assess their effects on both response rates and data quality (Strouse and Hall, 1994). Although the $10 payment marginally increased screener (but not interview) response rates over no incentive prior to refusal conversion, after refusal conversion--which offered a $10 payment to nonresponders in both groups--cooperation rates for the $0 and $10 groups were virtually

equivalent. As a result, payments were retained only for refusal conversion efforts for the remainder of the survey.

Not surprisingly, the potential efficacy of using incentives only for nonresponse follow-up in surveys of physicians has also been examined, with mixed results. Recall that Berry and Kanouse (1987) achieved a 78 percent response rate with a $20 incentive included with the initial mailing, compared with only 66 percent for those who received a check only after the completed questionnaire was returned. However, a subsample of the original postpayment sample, sent a special follow-up mailing with a check, had a final response rate of 77 percent, indicating that prepayment was effective even if it was used late in the contact process.

In contrast, the recent study reported by Berk, Edwards, and Gay (1993) found that, while "some beneficial impact was found, delaying the incentive until the second round of mailing did not have the same effect as including an incentive with the initial mailing" (p. 241). Overall, a 63 percent response rate was obtained for those receiving a prepaid incentive with the initial mailing, compared with only a 50 percent response rate for those sent a prepaid incentive with a second mailing, and a 40 percent rate for those receiving no incentive.

4. Summary and Conclusion

As we noted in the introduction, the use of monetary incentives and other forms of respondent remuneration has become increasingly common in survey research, spreading rather steadily from the origin of this practice in commercial and market research to the increased use of remuneration in academic and government surveys. In any discussion of this trend among survey research professionals, examples of situations under which incentives are routinely touted as either necessary or highly desirable to achieve adequate response rates always include their use with hard-to-reach or hard-to-interview populations. Although this assertion could be regarded as essentially tautological, there is, in fact, a reasonable consensus on a broad conception of what constitutes "hard-to-reach" respondents, including: (a) hard core refusals; (b) sample members protected by gatekeepers; (c) frequently surveyed groups; (d) target populations or subpopulations among whom the traditional positive forces to cooperate are quite low; and, in general, (e) those who are hard to locate, gain access to, and interview for a variety of different reasons.

In summarizing what we know or think we know about the use of incentives with such populations, it would be most useful to do so in relation to the seven basic questions that we raised at the outset:

1. *Are there indeed specific target populations who are routinely offered remuneration to participate in surveys by most or all survey organizations because they are regarded as especially difficult to survey?*

The specific types of respondents designated as hard-to-reach by organizations that we contacted in preparing this paper include: (1) the economically disadvantaged; (2) the educationally disadvantaged; (3) minority populations; (4) adolescents, youth, and young adults; (5) drug users and those with special health problems; (6) frequently surveyed professional or elite populations; and (7) transients and persons who wish not to be found for legal or other reasons.

Of the these seven different categories of respondents, the research literature on the effects of respondent incentives has focused directly on only a few--frequently surveyed professional and elite populations, adolescents and young adults, and the disadvantaged. Overall, to varying degrees, the results of these studies suggest that respondent incentives can be quite effective in stimulating survey cooperation among each of these hard-to-reach populations, especially the first two categories.

2. *While respondent incentives may increase cooperation among initial refusals, are they really effective with hard-core refusals and the truly difficult or impossible to interview populations?*

That respondent incentives are quite effective in averting initial refusals or in converting them after the fact is fairly clear from the research literature, but none of the literature reviewed specifically addresses their efficacy with hard core refusals or those who are truly difficult or impossible to interview. Possible exceptions are the effectiveness of the NALS and NSFG Cycle V field experiments in achieving higher response rates among those subpopulations whose response rates are typically quite low, in spite of extensive follow-up and refusal conversion efforts. In addition, most survey research professionals believe that respondent incentives are an important overall tool in their arsenal for dealing with these difficult-to-interview populations in particular.

3. *Are incentives effective only for certain target populations or subpopulations or more effective for certain population subgroups than for others (i.e., are the effects of incentives different for different population subgroups)?*

While studies of the effectiveness of incentives that focus directly on low income, minority, and disadvantaged populations are quite rare, those that address the more specific question of whether respondent incentives are more effective in improving

cooperation among such target populations than the more affluent and advantaged are considerably more common. Overall, these studies provide some provocative evidence suggesting that incentives are indeed more likely to influence lower income, lower socioeconomic status and minority populations to participate in surveys than those more advantaged. Similarly, several studies of physicians have found significant differences by specialty in the effectiveness of incentives in increasing survey cooperation. More generally, there is a growing body of research evidence suggesting that incentives are more effective for certain populations or population subgroups than for others.

4. *Are incentives really effective in getting past "gatekeepers," either for certain professionals (e.g., physicians) or other difficult-to-survey subgroups of the general population?*

Evidence provided both by survey practitioners and the research literature suggests that respondent incentives can indeed be effective in getting past gatekeepers and gaining access to certain difficult to reach populations.

5. *Are incentives indeed of little use in locating and interviewing hard-to-find cases?*

A few studies and the experiences reported by several research organizations suggest that respondent incentives can indeed be effective in reducing the time and costs associated with locating, tracking, and interviewing highly mobile or otherwise difficult to locate populations.

6. *Should consideration be given to paying some, but not all respondents to a given survey?*

A number of survey research organizations on some surveys use respondent incentives only for refusal conversion, i.e., in general respondents are not paid, but incentives are offered as one tool in the refusal conversion process.

Although there are important questions to be answered related to the equity or fairness of this practice, most of the research literature bearing on its feasibility suggests that quite similar rates of response and response quality can be achieved in most surveys by delaying the use incentives to later stages of contact in the survey process.

7. *Should all respondents be paid the same incentive, or should consideration be given to different levels or types of remuneration for different respondents*

The research literature suggests that incentives (or incentives of a given level) are more effective for certain

population subgroups than for others, and that different respondent groups may be more sensitive to different levels or types (e.g., prepaid versus conditional) of remuneration than others. Thus, some have suggested that "an optimal survey design would probably use different incentives for different population groups" (Peck and Dresch, 1981, p. 256). However, while the observed variation across a number of studies is quite real, the evidence to date does not yet appear sufficient to justify the use of different incentives for different categories of sample members or survey respondents.

Overall, what does this overview suggest about how we define and approach hard-to-reach or hard-to-interview respondents and the relative or special efficacy of remuneration in surveys of these target populations or subpopulations? This question is best addressed perhaps by first approaching the problem from the other direction, i.e., by asking ourselves what motivates cooperative or "easy-to-reach" respondents to participate in surveys.

Fundamentally, it is important to realize that monetary incentives represent only one of many incentives or motivating factors available to survey researchers (cf. Groves, 1992). Survey practitioners already use a number of other types of incentives to encourage survey participation--including appeals to civic duty, the eventual use of the information provided to help people, etc.--and such appeals are generally quite effective. A sense of civic duty to respond to a government survey may be viewed as the cumulative result of the provision of goods and services by the polity to its citizens (Groves, 1992), and this "social contract" to provide information for the public good clearly appeals to many people. Thus, the techniques typically used by survey researchers to encourage participation include appealing to respondents' obligations as citizens and emphasizing the societal benefits likely to accrue from the survey.

In turn, these appeals seek to activate one of two basic types of respondent motive patterns (Cannell and Henson, 1974): (1) a perception that participation in the interview will enable one to achieve certain personal goals (i.e., by emphasizing the importance and purpose of the survey and attempting to link these to achieving some personal or societal goal); and (2) a habitual mode of response toward requests made by legitimate agencies or organizations in society, based on norms of good citizenship, politeness, acquiescence to requests for information, etc. Unfortunately, as Cannell and Henson (1974) note:

> these two types of motivations are not very effective in increasing respondent [cooperation or] activity. Respondents do not share the researcher's goals, or, if they do, they fail to see the interview as an effective way of achieving that goal. . . . Similarly, the respondent role may be seen as related to citizen responsibility, but this is not usually a

> sufficiently salient or strong motive to induce a high level of activity [or commitment] . . . (p. 313)

In effect, then, an incentive may be offered to respondents to create a personal goal which motivates participation where no other goal or motivation exists.

In this sense, hard-to-reach respondents are those for whom the positive forces to cooperate are quite low, where direct connections to personal goals cannot be readily established and appeals to civic responsibility or benefits to society are quite likely to fail. For example, economically and educationally disadvantaged populations--and others disenfranchised or alienated from the mainstream society--typically have no context for valuing research or their contribution to the research process and are less likely to be persuaded to participate in surveys on the basis of "social utility." It is thus generally unrealistic to expect them to give their time simply because of the potential policy implications of the data we ask them to provide. In short, one must recognize that there are indeed some subpopulations where the "social contract" has broken down, where the dominant mode of decision making is cost-benefit analysis (Groves, 1992) and where economic rather than social exchange processes are paramount.

It is at this intersection that we encounter many of those respondents who we designate as hard-to-reach, and where the use of remuneration takes on major significance as the incentive of choice, as those based on personal goals, obligations to society, and social utility become increasingly blurred or nonexistent. Although these segments of our society may well be growing, they still undoubtedly constitute a clear minority. Thus, a major challenge for those who choose to use respondent incentives to encourage participation among those who are hard-to-survey is to strike a pragmatic and appropriate balance between providing effective levels of remuneration to induce such groups to respond while simultaneously minimizing the risk of alienating those for whom the "social contract" is very much intact, among whom a sense of civic duty is alive and well, thereby continuing to provide a more than adequate basis for their participation in surveys they perceive to be of benefit to themselves, our society, and the nation.

REFERENCES

Abraham, S. Y. (1994, May). *Using pilot surveys to increase response rates: Findings from the 1992-93 National Study of Postsecondary Faculty.* Paper presented at the 49th Annual Conference of the American Association for Public Opinion Research, Danvers, MA.

Abraham, S. Y., and Johnson, R. A. (1993, May). *Effects of incentives in a multi-modal survey of elites: Findings from the National Study of Postsecondary Faculty.* Paper presented at the 48th Annual Conference of the American Association for Public Opinion Research, St. Charles, IL.

Armstrong, J. S. (1975). Monetary incentives in mail surveys. *Public Opinion Quarterly, 39,* 111-116.

Benus, J., and Ackerman, J. C. (1971). The problem of nonresponse in sample surveys. In J. B. Lansing, S. B. Withey, and A. C. Wolfe (Eds.), *Working papers on survey research in poverty areas* (pp. 26-59). Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan.

Berk, M. L. (1985). Interviewing physicians: The effect of improved response rate. *American Journal of Public Health, 75,* 1338-1340.

Berk, M. L., Edwards, W. S., and Gay, N. L. (1993). The use of a prepaid incentive to convert nonresponders on a survey of physicians. *Evaluation and the Health Professions, 16*(2), 239-244.

Berk, M. L., Mathiowetz, N. A., Ward, E. P., and White, A. A. (1987). The effect of prepaid and promised incentives: Results of a controlled experiment. *Journal of Official Statistics, 3,* 449-457.

Berlin, M., Mohadjer, L., Waksberg, J., Kolstad, A., Kirsch, I., Rock, D., and Yamamoto, K. (1992). An experiment in monetary incentives. *Proceedings of the Survey Research Methods Section,* American Statistical Association, 393-398.

Berry, S. H., and Kanouse, D. E. (1987). Physicians response to a mailed survey: An experiment in timing of payment. *Public Opinion Quarterly, 51,* 102-114.

Blumberg, H. H., Fuller, C., and Hare, A. P. (1974). Response rates in postal surveys. *Public Opinion Quarterly, 38,* 113-123.

Brennan, M., Hoek, J., and Astridge, C. (1991). The effect of monetary incentives on the response rate and effectiveness of a mail survey. *Journal of the Market Research Society, 33,* 229-241.

Bryant, E. E., Kovar, M. G., and Miller, H. (1975). A study of the effect of remuneration upon response in the Health and Nutrition Examination Survey, United States. *Vital and Health Statistics*, Series 2-No.67. DHEW Pub. No (HRA) 76-1341. Washington, D.C.: U.S. Government Printing Office.

Cannell, C.F., and Camburn, D. (1991, June). *The effect of financial incentives on respondent participation in the Youth Risk Behavior Supplement.* Unpublished manuscript, Survey Research Center, University of Michigan, Ann Arbor, MI.

Cannell, C. F., and Henson, R. (1974). Incentives, motives, and responses bias. *Annals of Economic and Social Measurement, 3*, 307-317.

Cannell, C. F., and Fowler, F. J. (1977). Interviewers and interviewing techniques. In U.S. National Center for Health Services Research. *Advances in health survey research methods: Proceedings of a national invitational conference, 1975.* NCHSR Research Proceedings Series, DHEW Publication No. (HRA) 77-3154. Rockville, MD: Health Resources Administration.

Chromy, J. R., and Horvitz, D. G. (1978). The use of monetary incentives in national assessment household surveys. *Journal of the American Statistical Association, 73*, 473-478.

Church, A. H. (1993). Estimating the effects of incentives on mail response rates: A meta-analysis. *Public Opinion Quarterly, 57*, 62-79.

Cohen, S. B., Walden, D. C., and Ward, E. P. (1992, October). *Providing payments to respondents: The NMES experience.* Paper prepared for the COPAFS/OMB Symposium on Providing Incentives to Survey Respondents, John F. Kennedy School of Government, Harvard University, Cambridge, MA.

Cook, J. R., Schoeps, N., and Kim, S. (1985). Program responses to mail surveys as a function of monetary incentives. *Psychological Reports, 57*, 366.

Council of Professional Associations on Federal Statistics (COPAFS). (1993, September). *Providing incentives to survey respondents: Final report* (Contract No. GS0092AEM0914). Washington, DC: Regulatory Information Service Center, General Services Administration.

Cowan, C. D. (1977). Incentive effects on amounts reported in an expenditure diary survey. *Proceedings of the Social Statistics Section*, American Statistical Association, 498-503.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method.* New York: John Wiley.

Dillman, D. A. (1991). The design and administration of mail surveys. *Annual Review of Sociology, 17,* 225-249.

Dohrenwend, B. S. (1970). An experimental study of payments to respondents. *Public Opinion Quarterly, 34,* 621-624.

Duncan, W. J. (1979). Mail questionnaires in survey research: A review of response inducement techniques. *Journal of Management, 5,* 39-55.

Duffer, A. P., Lessler, J. T., Weeks, M. F., and Mosher, W. D. (1994, May). *Effects of incentive payments on response rates and field costs in a pretest of a national CAPI survey.* Paper presented at the 49th Annual Conference of the American Association for Public Opinion Research, Danvers, MA.

Erdos, P. L., and Morgan, A. J. (1983). *Professional mail surveys*(rev. ed.). Malabar, FL: Robert E. Krieger Publishing Company.

Ferber, R. (1974). Consumer expenditure and price data: An overview. *Annals of Economic and Social Measurement, 3,* 299-305.

Ferber, R., and Sudman, S. (1974). Effects of compensation in consumer expenditure studies. *Annuals of Economic and Social Measurement, 3,* 319-331.

Findlay, J. S., and Schaible, W. L. (1980). A study of the effect of increased remuneration on response in a health and nutrition examination survey. *Proceedings of the Section on Survey Research Methods,* American Statistical Association, 590-594.

Fox, R. J., Crask, M. R., and Kim, J. (1988). Mail survey response rate: A meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly, 52,* 467-491.

Furse, D. H., and Stewart, D. W. (1982). Monetary incentives versus promised contribution to charity: New evidence on mail survey response. *Journal of Marketing Research, 19,* 375-380.

Furse, D. H., and Stewart, D. W. (1984). Manipulating dissonance to improve mail survey response. *Psychology and Marketing, 1* (2), 79-94.

Gelb, B. D. (1975). Incentives to increase survey returns: Social class considerations. *Journal of Marketing Research, 12,* 107-109.

Godwin, R. K. (1979). The consequences of large monetary incentives in mail surveys of elites. *Public Opinion Quarterly,* 43, 378-387.

Goetz, E. G., Tyler, T. R., and Cook, F. L. (1984). Promised incentives in media research: A look at data quality, sample representativeness, and response rate. *Journal of Marketing Research, 21*, 148-154.

Goodman, L. and Jensen, L. (1981). Economic surveys of medical practice: AMA's periodic survey of physicians, 1966-1978. In S. Sudman (Ed.), *Proceedings of the third biennial conference on health survey methods*. Rockville, MD: National Center for Health Services Research.

Goodstadt, M. S., Chung, L., Kronitz, R., and Cook, J. (1977). Mail survey response rates: Their manipulation and impact. *Journal of Marketing Research, 14*, 391-395.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 25, 161-178.

Goyder, J. C. (1990). *Cash incentives and socio-economic bias in surveys*. Unpublished manuscript, Department of Sociology, University of Waterloo, Ontario, Canada.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: John Wiley and Sons.

Groves, R. M. (1992, October). *Respondent incentives, minimum mean square error per unit cost, and the resident's responsibility to the polity*. Discussion notes prepared for the COPAFS/OMB Symposium on Providing Incentives to Survey Respondents, John F. Kennedy School of Government, Harvard University, Cambridge, MA.

Groves, R. M., Cialdini, R. B., and Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly, 56*, 475-495.

Gunn, W. J., and Rhodes, I. N. (1981). Physician response rates to a telephone survey: Effects of monetary incentive level. *Public Opinion Quarterly, 45*, 109-115.

Hackler, J. C., and Bourgette, P. (1973). Dollars, dissonance and survey returns. *Public Opinion Quarterly, 37*, 276-281.

Hansen, R. A. (1980). A self-perception interpretation of the effect of monetary and nonmonetary incentives on mail survey respondent behavior. *Journal of Marketing Research, 17*, 77-83.

Harvey, L. (1987). Factors affecting response rates to mailed questionnaires: A comprehensive literature review. *Journal of the Market Research Society, 29*, 341-353.

Heberlein, T. A., and Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative review of the published literature. *American Sociological Review, 43*, 447-462.

Hopkins, K. D., and Gullickson, A. R. (1992). Response rates in survey research: A meta-analysis of the effects of monetary gratuities. *Journal of Experimental Education, 61*(1), 52-62.

Hopkins, K. D., Hopkins, B. R., and Schon, I. (1988). Mail surveys of professional populations: The effects of monetary gratuities on return rates. *Journal of Experimental Education, 56*(3), 173-175.

Houston, M. J., and Ford, N. M. (1976). Broadening the scope of methodological research on mail surveys. *Journal of Marketing Research, 13*, 397-403.

Hubbard, R., and Little, E. L. (1988a). Promised contributions to charity and mail survey responses: Replication with extension. *Public Opinion Quarterly, 52*, 223-230.

Hubbard, R., and Little, E. L. (1988b). Cash prizes and mail survey response rates: A threshold analysis. *Journal of the Academy of Marketing Science, 16*, 42-44.

Huck, S. W., and Gleason, E. M. (1974). Using monetary inducements to increase response rates from mail surveys. *Journal of Applied Psychology, 59*, 222-225.

James, J. M., and Bolstein, R. (1990). The effect of monetary incentives and follow-up mailings on the response rate and response quality in mail surveys. *Public Opinion Quarterly, 54*, 346-361.

James, J. M., and Bolstein, R. (1992). Large monetary incentives and their effect on mail response rates. *Public Opinion Quarterly, 56*, 442-453.

Jobber, D., and Saunders, J. (1988). Modelling the effects of prepaid monetary incentives on mail-survey response. *Journal of the Operational Research Society, 39*(4), 365-372.

Kanuk, L., and Berenson, C. (1975). Mail surveys and response rates: A literature review. *Journal of Marketing Research, 12*, 440-453.

Kephart, W. M., and Bressler, M. (1958). Increasing the response to mail questionnaires. *Public Opinion Quarterly, 22*, 123-132.

Kerachsky, S. H., and Mallar, C. D. (1981). The effects of monetary payments on survey responses: Evidence from a longitudinal study of economically disadvantaged youths. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 258-263.

Kulka, R. A. (1992, October). *A brief review of the use of monetary incentives in federal statistical surveys.* Paper prepared for the COPAFS/OMB Symposium on Providing Incentives to Survey Respondents, John F. Kennedy School of Government, Harvard University, Cambridge, MA.

Law, J. W. (1989). *The impact of respondent fee payments on the 1989 General Social Survey.* Chicago: National Opinion Research Center, University of Chicago.

Linsky, A. S. (1975). Stimulating responses to mailed questionnaires: A review. *Public Opinion Quarterly*, 1975, *39*, 82-101.

Lockhart, D. C. (1991). Mailed surveys to physicians: The effect of incentives and length on the return rate. *Journal of Pharmaceutical Marketing and Management*, *6*(1), 107-121.

Marrett, L. D., Kreiger, N., Dobbs, L., and Hilditch, S. (1992). The effect on response rates of offering a small incentive with a mailed questionnaire. *Annals of Epidemiology*, *2*(5), 745-753.

Martin, B. (1984). *Don't survey physicians.* Chicago: American Medical Association.

McDaniel, S. W., and Rao, C. P. (1980). The effect of monetary inducement on mailed questionnaire response quality. *Journal of Marketing Research*, *17*, 265-268.

Miller. H. W., Kennedy, J., and Bryant, E. E. (1972). A study of the effect of remuneration upon response in a health and nutrition examination survey. *Proceedings of the Social Statistics Section*, American Statistical Association, 1972, 370-375.

Mizes, J. S., Fleece, E. L., and Roos, C. (1984). Incentives for increasing return rates: Magnitude levels, response bias, and format. *Public Opinion Quarterly*, *48*, 794-800.

Mooney, G. , Giesbrecht, L., and Shettle, C. (1993). *To pay or not to pay; That is the question.* Paper presented at the 48th Annual Conference of the American Association for Public Opinion Research, St. Charles, Illinois, May 20-23.

Nederhof, A. J. (1983). The effects of material incentives in mail surveys: Two studies. *Public Opinion Quarterly*, 47, 103-111.

Paolillo, J. G. P., and Lorenzi, P. (1984). Monetary incentives and mail questionnaire response rates. *Journal of Advertising, 13*(1), 46-48.

Peck, J. K., and Dresch, S. P. (1981). Financial incentives, survey response, and sample representativeness: Does money matter? *Review of Public Data Use, 9*, 245-266.

Pendleton, A., and Ginsburg, A. (1992, October). *Providing incentives to survey respondents in education program evaluations.* Paper prepared for the COPAFS/OMB Symposium on Providing Incentives to Survey Respondents, John F. Kennedy School of Government, Harvard University, Cambridge, MA.

Perneger, T. V., Etter, J-F., and Rougemont, A. (1993). Randomized trial use of a monetary incentive and a reminder card to increase the response rate to a mailed health survey. *American Journal of Epidemiology, 138*, 714-722.

Robin, D. P., and Walters, C. G. (1976). The effect on return rate of messages explaining monetary incentives in mail questionnaire studies. *Journal of Business Communications, 13*, 49-54.

Shuttleworth, F. K. (1931). A study of questionnaire technique. *Journal of Educational Psychology, 22*, 652-658.

Sudman, S. (1985). Mail surveys of reluctant professionals. *Evaluation Review, 9*, 349-360.

Sudman, S., and Ferber, R. (1974). A comparison of alternative procedures for collecting consumer expenditure data for frequently purchased products. *Journal of Marketing Research, 11*, 128-135.

Skinner, S. J., Ferrell, O. C., and Pride, W. M. (1984). Personal versus nonpersonal incentives in mail surveys: Immediate versus delayed inducements. *Journal of the Academy of Marketing Science, 12*, 106-114.

Spry, V. M., Hovell, M. F., Sallis, J. G., Hofstetter, C. R., Elder, J. P., and Molgaard, C. A. (1989). Recruiting survey respondents to mailed surveys: Controlled trials of incentives and prompts. *American Journal of Epidemiology, 130*(1), 166-172.

Strouse, R., and Hall, J. (1994, May). *Effects of payments on response rates and data quality for a general population telephone survey.* Paper presented at the 49th Annual Conference of the American Association for Public Opinion Research, Danvers, MA.

Tambor, E. S., Chase, G. A., Faden, R. R., Geller, G., Hofman, K. J., and Holtzman, N. A. (1993). Improving response rates through incentive and follow-up: The effects on a survey of physicians' knowledge of genetics. *American Journal of Public Health, 83,* 1599-1603.

Tucker, C. (1992, October). *Providing incentives to respondents in government surveys.* Paper prepared for the COPAFS/OMB Symposium on Providing Incentives to Survey Respondents, John F. Kennedy School of Government, Harvard University, Cambridge, MA.

Weber, S. J., Wycoff, M. L., and Adamson, D. R. (1982). *The impact of two clinical trials on physician knowledge and practice.* Arlington, VA: Market Facts, Inc.

Weinberg, E. (1973). The issue of payment to respondents--wise or unwise. *Public Opinion Quarterly, 37* (3), 478-480.

Weiss, C. H. (1975). Interviewing in evaluation research. In E.I. Struening and M. Guttentag (Eds.), *Handbook of evaluation research* (pp. 355-395). Beverly Hills, CA: Sage Publishing Company.

Willimack, D. K. (1993). Selected results of the incentive experiment on the 1992 Farm Costs and Returns Survey. *Proceedings of the International Conference on Establishment Surveys* (pp. 490-495). Alexandria, VA: American Statistical Association.

Willimack, D. K., Petrella, M., Beebe, T., and Welk, M. (1992, August). *The use of incentives in surveys: Annotated bibliography.* Survey Research Center, Institute for Social Research, University of Michigan.

Wotruba, T. R. (1966). Monetary inducements and mail questionnaire response. *Journal of Marketing Research, 3,* 398-400.

Yammarino, F. J., Skinner, S. J., and Childers, T. L. (1991). Understanding mail survey response behavior: A meta-analysis. *Public Opinion Quarterly, 55,* 613-639.

Yu, J., and Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research, 20,* 36-44.

Zusman, B. J., and Duby, P. (1987). An evaluation of the use of monetary incentives in postsecondary survey research. *Journal of Research and Development in Education, 20*(4), 73-78.

Summary of Resolvable Errors in Edit Output

Benzene

chckcode codable primary occupation (qprmy_cd) - housewife (aqid=03233);
freqs qisfx2 = 0 for 1 case
rtickicd look into the operative procedure codes on page 1 - should these codes be based on residual body site?
page 61, aqid=01187: can this be coded per earlier error resolution

TCE

mainedit flag155-multibirth no, 1 preg, 2 livebirths- aqid=06011
chckcode 1 case w/emp_cd=10- do not assign- aqid=08222
phone area codes- page 1 -obs 29 and 30: 910 should 919
duplicate phone #:
area code=815; qid=009020090-check last name-it appears it may be the same person who was a proxy with same first and middle names but different last name
addresses with more than 1 phone:
251 ...- fix 6th digit in 1st phone #
4329.. - put last digit in 3rd position for first observation

street page 69: blocks 5 and 6
page 88 blocks 4 and 5
page 126 - last 2 blocks

Transaction log

1) The transaction log should be log of CHANGES MADE to database - not changes INTENDED to be made - describe how this log is created with respect to the SAS program where corrections are programmed.

2) provide report of errors - changes not made- as provided earlier.

3) local variables (i.e. AQID=2795: LREGZIP) should never be changed

4) need consistent handling of #'s(i.e. AQID=0118: "#10"), apt, NO (1st vs. 2nd line of address)

5) re-sort this report: aqid, variable name, reverse chronological date and time

6) appears to be 2 qids for aqid=2923. Please look into this and also provide proc freq/list of aqid * qid.

7) NORC needs to do very careful QA/QC of transaction log prior to finalizing files and sending final deliverable - perhaps forward this to ATSDR for review electronically (scrambled) prior to review of other deliverables

specific problems:

AOID	Problem
0118	"Court" should be standardized
0192	"Calendar, CT" appears in final edit output, but in this log seems t be an old value that was changed to "Venice, CA". How can this be?
0802	Cancer condition (2/91) appears to have occurred BEFORE last intervie date
2692	Variable Name blank
2876	PK should be PARK
2923	*2 qids?? why so many changes - is this correct? need details on case
3954	condition (D/89) appears to have occurred BEFORE last interview date

DISCUSSION

Diane K. Willimack
National Agricultural Statistics Service

I would like to begin by congratulating and commending these authors on two thorough and well-presented papers. Kulka has written a theory section that sets the stage for incentive use in surveys and raises conceptual issues for consideration by survey designers. He goes on to provide a thorough review of the literature and a thoughtful evaluation of the implications for survey designers, particularly regarding "hard-to-reach" population subgroups.

Ezzati, White, Mosher, and Sanchez -- the NCHS authors plus one -- have provided a compilation of the findings from a series of experiments in which incentive use has been systematically studied on a wide variety of survey types and target populations -- from hospitals to households, from physicians to disadvantaged population subgroups. In addition they report the value to interviewers of incentive use: how respondent incentives enhance interviewers' confidence and their ability to do their job, as well as how interviewer incentives enhance their morale and motivation. This summary challenges us all to consider how to apply their findings to our own survey situations, so that we need not reinvent the wheel.

Allow me to share what I learned from these two papers. Kulka differentiates the social norm of reciprocity and social exchange theory from the theory of economic exchange as the conceptual basis for incentive use in surveys. Typically we describe incentives by their type -- monetary vs. nonmonetary -- and timing -- prepaid vs. promised. Let us focus on timing. Use of prepaid incentives is based on the social theories, while promised incentives are believed to invoke economic exchange in respondents.

These authors tend to use the terms "incentive" and "remuneration" interchangeably. Kulka also uses the word "compensation". I decided to look up these words in my Random House Collegiate Dictionary (1988). I know that our use of words as jargon need not have any relationship with the English language, but I thought this exercise might be instructive, as well as help me clarify my thoughts.

"Incentive," in my dictionary, is defined as "something that incites to action," and lists as synonyms, "stimulus, spur, incitement, encouragement." Let me interpret: encouragement to respond. "Remuneration" is "something that remunerates; reward; pay." "Remunerate" means "to recompense," which means "to repay or reward (someone), as for aid or service." Again I offer my interpretation: to recompense (repay) for aiding us by responding to our survey. Finally, "compensation" is "something given or received as an equivalent for services, etc.," with synonyms "recompense, remuneration, payment." It seems to me that "remuneration" and "compensation" are interchangeable with each other, but not with "incentive."

Let us return to theory. One of the problems with promised incentives is that, in many studies comparing them to prepaid incentives, the amount of money or the item was token in nature. Thus while the "promise" tends to invoke economic exchange in respondents, the token nature was not sufficient to influence respondents to engage in the economic contract. But remuneration does

invoke economic exchange, since it implies an amount of money sufficiently close to the value of the service of survey response rendered by the respondent.

Thus, if an incentive is token in nature, it must be prepaid, for it can only rely upon social norms in order to be successful. If a promised incentive is to be used, then, for it to be successful, it must be substantial enough to be considered remuneration or compensation in an economic exchange. Of course the latter must take into consideration the difficulty of the burden of the task of completing the survey. Perhaps I have just exhibited a firm grasp of the obvious, but this notion did not crystalize for me until I read these two papers side by side.

However, tokens appear to work in instances when signs of appreciation and good faith are meaningful to the respondent. For example, the National Agricultural Statistics Service (NASS) conducted an incentive experiment on the 1992 Farm Costs and Returns Survey, an annual voluntary survey of U.S. farm operators collecting detailed expenditure and income data in personal interviews lasting 90 minutes on average. Response rates were increased by nearly 5.5 percentage points in the group receiving a prepaid nonmonetary incentive. Of greater interest, though, was the finding that incentive use increased response rates by 17 points among the smallest farms (those with sales less than $20,000) and by nearly 12 points among the largest farms with sales of $500,000 or more. The token pocket portfolio and calculator that we gave them cannot have indicated an economic exchange to these large farms. Instead, it likely had appeal as a symbol of appreciation consistent with the repeated survey contact to which these two groups, in particular, are subjected (Willimack, 1993).

Farms are establishments, and NASS surveys collect establishment information from them. Although farms may be a special case because they exhibit many of the same characteristics as households, application of incentives or remuneration in establishment surveys is not trivial. The NCHS paper reports a survey of hospital records, in which hospitals are paid for their service of abstracting sampled records. This is clearly remuneration with basis in economic exchange. However this may not always be so clear cut for incentive use or remuneration in establishment surveys, particularly surveys of businesses.

Incentives are meant to "incite action," to influence the decision to participate in a survey and to motivate the respondent. But in an establishment survey, it is unclear who we are attempting to influence with incentive use. In an establishment, the person making the decision about survey participation may not be the desired respondent. The desired respondent is the person who is the most knowledgeable provider of the information being sought, usually the person who has access and understanding of any records to be used as a source for responding (Edwards and Cantor, 1991).

If a prepaid incentive is provided in an establishment survey, who gets it? Is it the boss, in order to "incite" or encourage a favorable decision about responding? Or is it the employee who actually completes the survey instrument, in order to motivate careful response? Or is it the business, as income or as payment for the service of completing the survey? It does seem less difficult to make this "leap" if it is remuneration being offered rather than an incentive. Remuneration is more clearly a payment to the business for the service of completing the questionnaire. But then the dollar amount must be reasonable relative to the burden of the task.

Let us consider further the dollar amount to be used as an incentive or as an offer of remuneration. I think money is tricky, unless the amount is clearly token. Money is a sensitive topic for many people, and it has all kinds of different connotations. No matter how much is offered, there will always be someone for whom that amount is not enough. Furthermore, since respondents use all the information available to them when responding to survey questions, it seems reasonable to suggest that the same is true in those initial moments of contact during which a survey participation decision is being made. A key piece of information is the dollar amount being offered. I think the dollar amount provides an indication to respondents about the potential difficulty of the survey task, and may arouse suspicion. The amounts of $20, $30, $50, even $100 in several of the health studies described by the NCHS authors may have provided a fair indication of the difficulty of the task. On the other hand, consider the James and Bolstein (1992) study cited by Kulka in which a promise of $50 failed to increase response rates in a survey of construction subcontractors. Similarly, in the HIV survey described in the NCHS paper, $175 offered to survey refusers increased response rates by only 4 percentage points, while $100 resulted in a 10 point increase. These inordinately large dollar amounts may have indicated a difficult survey task to the respective target populations, resulting in a sense, "Well, if you're offering that much money, then what you're asking me to do must be really hard, embarrassing, or uncomfortable. It must be something that I don't want to do." Meanwhile, the choice to accept or reject the offer -- to engage in the economic contract -- remains with the respondent.

The amount of money offered as an incentive or as remuneration is an important variable to the survey designer. Here we have reached the edge of a gaping hole in the literature on incentives. What is the trade-off between the incentive or remuneration amount and the difficulty of the survey task? Indeed, under which survey circumstances will a token incentive suffice, and when is remuneration needed? These papers seem to suggest that when the survey task is particularly intrusive, burdensome, lengthy, or longitudinal, then the promise of remuneration seems appropriate, if not necessary. But these factors must be evaluated relative to the expectations and perceptions of the members of the target population: Perhaps a personal interview lasting 1½ to 3 hours and requesting detailed expense and income data is considered by farm operators to be as personal and sensitive as the Survey of Family Growth asking women aged 15-44 about abortion and sexual practices. As discussants are required to say, "more research is needed."

Furthermore, participants at the 1992 COPAFS conference on incentives listed guidelines for OMB to consider when evaluating incentive use in Federal surveys (COPAFS, 1993). These include:

"To compensate a respondent if there is risk in participating."

"When there are unusual demands or intrusions on the respondent (e.g., lengthy interviews, keeping a diary, having a blood sample drawn, ...)."

"When sensitive questions are being asked."

"If there is a lengthy field period (e.g., a commitment over time for a panel survey)."

"If there is any out-of-pocket cost to the respondent ..."

"If the respondent is a small business or a nonprofit institution in a voluntary survey and the respondent perceives some cost and burden to participating."

Although these represent only a subset of COPAF's recommendations to OMB, it appears to me that they outline survey circumstances that favor remuneration, as I have defined it in this discussion. However, there may be survey situations in which an incentive, as I have differentiated it, will suffice. If OMB policy appears to favor remuneration over incentive use, then agencies will design surveys accordingly, with consequences for survey budgets, management, and respondent burden. OMB needs to write policy that recognizes and encourages appropriate use of both incentives and remuneration.

How do we evaluate the effectiveness of incentives or remuneration in surveys? In the early days, as Kulka points out in his literature review, it was as simple as "Were response rates increased by incentive use?" A resounding "YES" is supported by a large number of studies on various types of surveys. Soon attention turned to the quality of the data, adding the question, "Does incentive use improve the quality of the data through the respondent's increased attentiveness to the survey task, particularly reflected in reduced item nonresponse?", or, at least, "Can we be sure that data quality has not been reduced?" Again, both Kulka and the NCHS authors provide a great deal of evidence supporting improved data quality related to incentive use or remuneration.

Let me next turn attention to what I call potential "dividends" to be achieved from incentive use: reduction in components of total survey error. Both sets of authors provide a number of examples of increased survey participation among "hard-to-reach" or typically under-represented population subgroups. In addition, the NCHS authors have provided results that show systematic differences on key variables among groups whose participation was increased by remuneration. Thus, these papers present evidence that nonresponse bias in survey estimates may be reduced through use of incentives or remuneration.

Moreover, let me offer an additional "dividend" found in two incentive experiments in which I have been involved: enhanced ability to identify ineligible sample units, resulting in nonsampling error reduction. In an incentive experiment on the 1991 Detroit Area Study, a statistically significant increase was found in the rate of ineligibile sample addresses. This may have been due, in part, to more diligent postal returns of the small packages containing a prepaid pen incentive, along with more reliable interviewer confirmation, when an address was a vacant housing unit, a business, or an incorrect address (Willimack and others, 1994). A replication of this result is currently being tested in the 1994 Detroit Area Study.

In addition, in the incentive experiment conducted on the 1992 Farm Costs and Returns survey, a higher rate of sample units being screened out as non-farms, and therefore not eligible for the survey, was found to be statistically significant. Incentive recipients who had no agriculture may have been more attentive to the survey request and more determined to notify the interviewer of their non-farm status, rather than to refuse or be inaccessible based on a belief that the survey did not apply to them (Willimack, 1993).

Incentive studies of the 1990's appear to have added cost evaluations or cost-benefit analyses of incentive use or remuneration. The results of the pretest of the Survey of Family Growth are particularly compelling. They show that the $20 incentive/remuneration actually saved money over

no incentive, since these respondents required fewer contacts and were less likely to break appointments, resulting in reduced interviewing costs.

Cost evaluation is an area that still needs development. Achieved cost savings may not always be so clear. Instead, frequently we are left to judge whether observed benefits were sufficient to justify the additional costs of incentive/remuneration use. It is difficult to put a dollar value on increased response rates, improved data quality, and reduction in total survey error. No doubt we all consider these to be very valuable. But how much improvement is needed in order to justify the additional cost of incentives or remuneration? What production and efficiency measures should be monitored during data collection and post-survey review and processing? What comparisons with which other survey methods should be undertaken? We must be able to evaluate whether incentive use or remuneration is the best tool to pull out of our methodology tool box.

Again I congratulate the authors, and I thank them for their contributions to our understanding of incentive use in surveys.

REFERENCES

Council of Professional Associations on Federal Statistics (COPAFS) (1993), Providing Incentives to Survey Respondents: Final Report (Contract No. GS0092AEM0914), Washington, D.C.: Regulatory Information Service Center, General Services Administration.

Edwards, W. S., and Cantor, D. (1991), "Toward a Response Model in Establishment Surveys," Ch. 12 in Measurement Errors in Surveys, edited by P. P. Beimer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman, New York: John Wiley and Sons, Inc., pp. 211-233.

James, J. M., and Bolstein, R. (1992), "Large Monetary Incentives and Their Effect on Mail Survey Response Rates," Public Opinion Quarterly 56:442-453.

The Random House Collegiate Dictionary (1988), New York: Random House, Inc.

Willimack, D. K. (1993), "Selected Results of the Incentive Experiment on the 1992 Farm Costs and Returns Survey," Proceedings of the Interantional Conference on Establishment Surveys, Alexandria, VA: Amaerican Statistical Association, pp. 490-495.

Willimack, D. K., Schuman, H., Pennell, B. E., and Lepkowski, J. M. (1994),"The Effects of a Prepaid Nonmonetary Incentive on Response Rates and Response Quality in a Face-to-Face Survey," unpublished manuscript, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI.

DISCUSSION

W. Sherman Edwards
Westat, Inc.

The use of monetary incentives in surveys is an endlessly fascinating topic for those of us in the business of collecting and analyzing survey data. Regardless of the importance and policy relevance of any particular survey, regardless of the eminent history and usefulness of the statistics derived from a survey, at some point in data collection it gets down to the trenches -- getting the last reluctant respondent to cooperate, finding the last elusive subject, persuading the last indifferent citizen to answer what seem to him like senseless and repetitive questions. The use of monetary incentives to help in this unavoidable struggle raises interesting methodological questions and, even more provocatively, questions about our relationship with our respondents. As a society, we are strongly ambivalent about money; it is the most public of our life's trappings, and yet at the same time it is the most private. Every survey researcher knows that it is not questions about sexual behavior or the intimate details of health that elicit the highest item refusal rates -- it is questions about income and financial assets. Introducing money into any relationship that is not primarily an economic one complicates that relationship, as anyone who has loaned money to a relative or friend can attest.

My discussion is organized first around two specific methodological questions and then addresses the relationship issue. Although the literature on the effects of monetary incentives is extensive, I will suggest several areas that warrant considerable further study.

The first methodological question is the most basic, "Do monetary incentives improve response rates?" The two papers today provide a fairly unequivocal "Yes" response to this question for surveys where incentive experiments were conducted. The paper by Dick Kulka examines the use of incentives for hard-to-survey populations. This is a very comprehensive paper; it provides a thorough and very useful overview of the research literature on incentives, of different theoretical views of how incentives work, and of the methodological issues raised when considering monetary incentives. The paper then goes on to explore in more detail issues around incentives for hard-to-survey populations. Although this comment is not particularly relevant to my discussion, I can't help but note that I never before considered the commonalty among drug users, physicians, youth in general, CEOs, impoverished urban minorities, diabetics, unwed mothers, small farm operators, and those defaulting on student loans. (They are all considered hard to survey, and thus worthy of consideration for monetary incentives.)

The paper by Trena Ezzati and her colleagues reviews the experiences with monetary incentives of several surveys about health and the use of health care services conducted by the National Center for Health Statistics and the Agency for Health Care Policy and Research. This paper presents a more detailed look at a smaller number of studies, including some that have pushed at the edges of the envelope containing this issue.

Kulka surveyed participants in the 1993 COPAFS conference on incentives, and found their use widespread. I think it's fair to characterize the view of many survey firms towards

incentives as one of a number of tools that they can use to achieve satisfactorily high response rates, but one that is often used with some reluctance. Other tools in this kit, of course, include repeated callbacks to convert initially reluctant respondents, the use of specially designed persuasive materials that may be targeted to particular reasons for refusal, and the use of especially effective interviewers. How do we choose among these tools in designing our data collection strategies? The most sympathetic view is that we choose rationally, based on empirical evidence and careful consideration of all alternatives, choosing the most cost-effective procedures possible. A more cynical view might be that many surveys or survey organizations opt for incentives because they are easier to implement than other approaches. I suspect that both views have some truth. For mail surveys, a fair amount of thought and research have been devoted to optimal design strategies, most notably Dillman's Total Design Method (Dillman, 1978). The literature on telephone surveys is somewhat less well developed, and that on in-person still less. This is in part because the issues become more complex as the relationship between survey and respondent becomes more personal. Recent work by Groves and Couper (1994, e.g.) is notable in applying constructs and research techniques from several disciplines to the tool kit. One lesson from their work is that we still have a lot to learn about the tools, including when and how to use them.

An important part of the question of whether incentives work is whether they are cost-effective. Some of the research studies cited in the two papers included formal evaluations of cost-effectiveness. The recent NSFG evaluation, for example, found that the cost per case for the $20 incentive treatment was lower than for the non-incentive treatment, in addition to yielding higher response rates. Other studies have examined the cost of monetary incentives against other methods, such as repeated mailings, to achieve comparable response rates. Often, incentives seem to reduce the cost of surveys The cost-effectiveness of a monetary incentive is related to the mode of administration and the size of the incentive. The marginal cost of a $20 incentive for an in-person survey is clearly much lower than the cost of a similar incentive in a mail survey. Kulka mentions the possible effectiveness of an incentive in locating elusive respondents. In-person locating is one of the most expensive of data collection activities -- even a small improvement in locating effectiveness would probably be worth the cost of a monetary incentive judiciously mentioned to an informant.

The second methodological question is, "What effect do monetary incentives have on data quality?" Kulka notes two commonly expressed views from the research literature: the first that respondents view incentives as a kind of "social exchange," and thus work harder and provide better responses. The second is that incentives are an "external motivator," decreasing internal motivation and hence reducing the quality of responses. Here, the literature is somewhat less persuasive. Some studies, like the NSFG experiment and the seroprevalence survey pilot cited by Ezzati, take the "more is better" view, which is probably often appropriate, particularly when the results are as dramatic as in these studies. Few studies have combined incentive experiments with validation of data, which is often difficult or even impossible. In particular, how does data quality differ between those who would (or do) participate without an incentive and those for whom the incentive makes the difference? To examine this, we might use respondents' perceptions about the use of incentives as an explanatory variable for some measures of data quality.

The only study cited that explicitly explored respondents' attitudes was associated with the NHIS Youth Risk Behavior Supplement, as cited by Kulka. Kalton and his colleagues used follow-up cognitive interviews with youths responding to a field test of the survey to discover an apparently strong link between the incentive and the effort expended in answering questions. This seems an especially promising methodology for exploring the effects of incentives, particularly on "special" or hard-to-survey populations.

Monetary incentives are frequently used in diary surveys -- such use is even sanctioned by the Office of Management and Budget. Could a monetary incentive affect the behavior of interest by a diary keeper? At the recent AAPOR conference, Diane Woodard of Arbitron (1994) described a series of incentive experiments for respondents in their radio listenership survey. One treatment involved a sweepstakes with a $3,000 first prize in addition to the usual nominal monetary incentive. In this treatment, all radio ratings increased. The post hoc explanation of this phenomenon was that respondents had increased their radio listening (or at least their reporting) in the erroneous belief that this would increase their chances of winning the sweepstakes. Another explanation, of course, is that the sweepstakes respondents were simply reporting better, although I am inclined to accept Woodard's interpretation.

A strongly held view of many survey researchers is that all respondents should be subject to exactly the same stimulus, as nearly as this can be controlled by the survey design. Only in this way can we be confident of the reliability of survey responses. This view has come under increasing criticism from those who view the structured interview as too restrictive a vehicle for meaningful exchange of information. For the more traditional view, monetary incentives may throw a monkey wrench in the works. As Kulka notes, some surveys offer incentives only to "hard-core" refusals or to those who are expected to be difficult to locate or interview. Even when the same incentive is offered to all respondents, their reactions to being "paid" for their time undoubtedly vary considerably. In every survey I've been associated with that offered incentives, some respondents refused to accept them. Again, it would seem useful to ask respondents how they feel about the payment, and to examine how responses vary by these views. At the heart of the matter, do monetary incentives introduce more variation in respondent motivation and perception of the survey than otherwise exists?

Let me now turn to the issue of the relationship between the survey and the respondent. If we define this relationship in terms of the use of incentives, there seem to be three views of the nature of the relationship. The first, which I would associate with not using incentives, is that surveys are a social good, and that participation alone is sufficient reward. Clearly, government surveys are the most logical candidates for this high-minded view, and there is some evidence in support of it. The Census Bureau does not use incentives, and yet achieves the highest response rates in the industry, even setting aside their mandatory surveys.

Those of us in the private sector, however, tend to be fairly pragmatic, if still somewhat ambivalent, as Kulka reports, about the use of incentives. What claim does Westat or anyone else outside of the government have on respondents, even if we do often represent the government? (I note that all of the surveys described in the Ezzati et al paper were conducted by private contractors.) Even if we have ethical concerns about the fairness of paying respondents in some surveys but not others, or of paying the "hard-core" but not the compliant, are the

alternatives any more attractive? Is it ethically more appealing to ask again and again, or to fly in the charming interviewer from Indiana, than to offer a monetary incentive?

Another view of the relationship between survey and respondent is that of the "social exchange" that Kulka describes. I would suggest that many who conduct survey research are most comfortable with this view. I have heard many researchers, some in describing incentive experiments, say, "The interviewers feel better when they can give the respondent something." Interviewers in this case, as they so often are, are our proxies. Each of today's papers had one mention of the positive effect of respondent incentives on interviewer behavior. This is another area that warrants further study.

One of the groundbreaking studies described by Ezzati et al is the pilot test of the National Household HIV Seroprevalence Survey, conducted by RTI in Dallas. Here, the size of the incentives offered ($50, $100, $175) seem to go well beyond what would be called for in a social exchange view of the interview. Respondents were asked in return for a blood sample and to complete a brief sexual history questionnaire. However, the context of this survey was socially charged. Respondents might be fearful of the blood draw in general, but particularly because it was associated with an AIDS survey. One view of the size of the incentive was that it was an appropriately attractive exchange for asking respondents to overcome a whole set of fears.

The third view of monetary incentives in this relationship is that of a business transaction, or "economic exchange," as it is described in the Kulka paper -- we are paying the respondent for his or her time. The implication of this view is that we are participants in an information marketplace, subject to the principles of supply and demand. In government household surveys, even those conducted by contractors, most researchers shy away from this view, and even explicitly deny it. Even some incentive experiments wind up giving all respondents the same payment. The survey of allergists cited by Ezzati is an example -- regardless of the incentive treatment, all participating physicians were eventually paid the same amount.

One is really taking an economic view when one uses disproportionate payments for different survey respondents, including payment to initial nonresponders but not to compliers. Here, we pay according to how badly we want the information -- truly supply and demand. This practice is anathema to some researchers, who cite the ethics of fairness and their distaste for rewarding noncompliant behavior.

Where monetary incentives have sometimes become a business transaction is, not surprisingly, in surveys of businesses. The prediction about respondents coming to expect incentives has probably come to pass for physician surveys, although Kulka notes that some are still conducted successfully without incentives. Ezzati et al describe the experience of the National Hospital Discharge Survey, where hospitals are routinely paid piecework for completing survey forms. The National Medical Expenditure Survey's Medical Provider Survey also allowed payment for survey respondents, the amount to be negotiated, but only if requested by the medical provider. On the other hand, most Federal establishment surveys do not offer reimbursement on demand or incentives.

One argument put forward against monetary incentives is that respondents will come to expect payment, and response rates will be even harder to achieve at a reasonable cost. This is referred to in the summary of last year's COPAFS symposium as the "slippery slope" argument. In essence, this argument says that the use of incentives breeds the perception of a survey marketplace. It seems unlikely that this effect would occur among the general public solely from Federal surveys. Even with the current rate of surveying, most households are never or rarely selected for a government survey (excepting the decennial Census, of course). If we are in or to be in a survey marketplace in this sense, commercial surveys will have had more to do with that result, and OMB has no control over their behavior. The "slippery slope" is more likely to refer to the behavior of survey organizations if we come to rely too routinely on monetary incentives to achieve our response rate goals.

One noticeable feature of both of these well-prepared and very interesting papers is the number of experiments that have been conducted around incentives, particularly for Federal surveys. I suspect that as we in the statistical community continue to explore the complexities of our relationship with our survey respondents, we will continue to conduct incentive experiments at an unflagging pace. There is certainly much more to learn about the effects of incentives on data quality, and about alternative methods for influencing respondent compliance.

REFERENCES

Dillman, D.A. (1978), Mail and Telephone Surveys: The Total Design Method, New York: Wiley.

Groves, R. and Couper, M. (1994), "Understanding the Process of Survey Cooperation: Household-Interviewer Doorstep Interactions," paper presented at the 49th Annual Conference of the American Association of Public Opinion Research, Danvers, MA.

Woodard, D. (1994), "Raising Response Rates with Premiums, or Is $1 Worth Anything to Anyone Anymore?" paper presented at the 49th Annual Conference of the American Association of Public Opinion Research, Danvers, MA.

Session 7

COMPUTER ASSISTED SURVEY INFORMATION COLLECTION

REDESIGNING A QUESTIONNAIRE FOR COMPUTER-ASSISTED DATA COLLECTION: THE CURRENT POPULATION SURVEY EXPERIENCE

Cathryn Dippo and Anne Polivka
Bureau of Labor Statistics

Kathleen Creighton, Donna Kostanich, and Jennifer Rothgeb
Bureau of the Census

Over the last decade, there have been two new factors that have significantly influenced the design of survey data collection--the computer and the theories and methods of cognitive psychology. When, in 1986, staffs of the Bureau of Labor Statistics (BLS) and the Bureau of the Census initiated a process for redesigning the Current Population Survey (CPS), incorporating these two new factors was made a top priority. In the paper[1], the authors[2] illustrate how, by concentrating on the cognitive processes of respondents and interviewers, computer-assisted interviewing was used as a tool for reducing measurement error.

The following topics are covered in the paper: background material on questionnaire design and computer-assisted interviewing methodologies, development of the CPS questionnaire over the last 50 years and how redesigning the CPS questionnaire for the 21st century has brought together the two new methodologies, using the computer in evaluating alternative questionnaire designs, examples of the new CPS questionnaire's design features which aid the cognitive processes of the respondent and interviewer and are primarily dependent on the use of the computer, the effects of the new questionnaire and collection procedures on labor force estimates, and a discussion of issues for the future.

The Current Population Survey is a monthly survey of approximately 60,000 households. The CPS survey, conducted for BLS by the Bureau of the Census, is the primary source of information on the U. S. labor force. Each month BLS analyzes and publishes information from the CPS, such as the unemployment rate, demographic characteristics of individuals in the labor force, and the number of hours individuals work. The survey began in 1940 under the auspices of the Works Projects Administration and was called the Monthly Report of Unemployment. The current CPS questionnaire has remained essentially

---

[1] The presentation by Cathryn Dippo at the Seminar on New Directions in Statistical Methodology was based on a paper currently under review for publication in a refereed journal. Thus, only a brief synopsis is being published here, along with a detailed bibliography of papers related to the CPS redesign.

[2] The new CPS questionnaire is the result of a team effort which involved many staff members from both BLS and Census. Space does not allow us to recognize everyone. The other members of the BLS-Census Questionnaire Design and Overlap Analysis Steering Committees over the years were Chester Bowie, John Bregger, Shail Butani, Lawrence Cahoon, Kennon Copeland, Harvey Hamel, Elizabeth Martin, Michael McMahon, Thomas Scopp, Clyde Tucker, Ronald Tucker, and Alan Tupek.

unchanged since the last major revision in January 1967. With only minor exceptions, the concepts measured have remained constant since the late 1940's.

Over its 50+-year history, the CPS has continued to be a model for survey designers. It was the first national probability sample of households, and many of the statistical methods for sampling and estimation now considered common practice were originally researched and implemented in CPS. Two of the six research areas identified in 1986 related to data collection--computer-assisted interviewing and the questionnaire. A Questionnaire Design Task Force was established to identify the cognitive and conceptual problems in the existing questionnaire, to suggest possible solutions for identified problems, and to develop a research plan to design and test a new questionnaire, along with related survey procedures. A separate task force was established to investigate the potential uses of computer-assisted interviewing. When a final consolidated research plan was approved in 1988, a major premise of the plan was that all interviews would be conducted using a computer. Following a period of questionnaire development and extensive testing, Census began collecting all CPS data using a new fully-automated questionnaire in January 1994.

The data produced from the CPS are closely-watched by economic forecasters and policy analysts. Therefore, all changes had to be carefully researched prior to implementation. By concentrating on facilitating the cognitive processes used by respondents and interviewers, research on alternative measurement processes resulted in reduced nonsampling errors. By capitalizing on the power and versatility of the computer, new research tools were developed to provide the evidence needed to understand the effects of changes in data collection procedures. We hope that the approach used for developing the new measurement process for CPS will serve as a model for future survey redesign projects.

For details on the 8 years of research that went into redesigning the CPS, please consult the papers listed in the following bibliography.

1994 CPS REDESIGN BIBLIOGRAPHY

Bienias, J. L. (1988). Preliminary laboratory research on Current Population Survey questionnaire with employed persons. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 185-190.

Bienias, J. L., Dippo, C. S., and Palmisano, M. (Eds.) (1987). *Questionnaire design: Report on the 1987 BLS Advisory Conference*. Washington, DC: U.S. Bureau of Labor Statistics.

Boehm. L. E. (1989). Reliability of proxy response in the Current Population Survey. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 486-489.

Boehm, L. E. (1989, August). The relationship between confidence, knowledge, and performance in the Current Population Survey. Paper presented annual meeting of the Southeastern Psychological Association, Washington, DC.

Bowie, C., Cahoon, L., and Martin, E. (1993). Overhauling the Current Population Survey: Evaluating changes in the estimates. *Monthly Labor Review, 116(9)*, 29-33.

Bregger, J. (1977). Establishment of a new Employment Statistics Review Commission. *Monthly Labor Review, 100 (3)*, 14-20.

Bregger, J. (1984). The Current Population Survey: A historical view and BLS's role. *Monthly Labor Review, 107 (6)*, 8-14.

Bregger, J. (1994, May). *Raising the Nation's unemployment rate*. Paper presented at the Seminar on New Directions in Statistical Methodology, Bethesda, MD.

Bregger, J., and Dippo, C. S. (1993). Overhauling the Current Population Survey: Why is it necessary to change? *Monthly Labor Review, 116(9)*, 3-9.

Bushery, J. M., Weidman, L., Briley, S. M., and Ciochetto, S. (1987). Using CATI in the Current Population Survey. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 778-783.

Butani, S., Cahoon, L., Fay, R., and Kostanich, D. (1994, in press). Measurement of different design effects. *American Statistical Association Proceedings of the Section on Survey Research Methods.*

Butz, W. P., and Plewes, T. J. (1989). A Current Population Survey for the 21st century. *Bureau of the Census Proceedings of the Fifth Annual Research Conference*, 3-13.

Cohany, S. R., Polivka, A. E., and Rothgeb, J. M. (1994). Revisions in the Current Population Survey effective January 1994. *Employment and Earnings, 41-2*, 13-37.

Cahoon, L. S. (1990, 1991, 1992, 1993). *Results of the CPS CATI Phase-in analysis*. Series of internal memoranda (issued quarterly). Washington, DC: Bureau of the Census.

Cahoon, L. S., and Russell, G. (1991, October). *Effects of CATI on estimates from demographic surveys conducted by the Bureau of the Census*. Paper presented at the

Joint Census Advisory Committee Meeting of the American Marketing Association and the Committee on Population Statistics, Alexandria, VA.

Campanelli, P. C., and Esposito, J. L. (1992, May). *The use of behavioral coding to identify problematic questions on alternative questionnaires*. Paper presented at the annual American Association for Public Opinion Research Conference, St. Petersburg, FL.

Campanelli, P. C., Martin, E. A., and Creighton, K. P. (1989). Respondents' understanding of labor force concepts: Insights from debriefing studies. *Bureau of the Census Proceedings of the Fifth Annual Research Conference*, 361-374.

Campanelli, P. C., Martin, E. A., and Rothgeb, J. M. (1991). The use of respondent and interviewer debriefing studies as a way to study response error in survey data. *The Statistician, 40*, 253-264.

Campanelli, P. C., Rothgeb, J. M., and Martin, E. A. (1989). The role of respondent comprehension and interviewer knowledge in CPS labor force classification. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 425-430.

Cantwell, P. J., and Ernst, L. R. (1993). Short-term changes to the CPS composite estimator in January 1994. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 724-729.

Cantor, D., and Esposito, J. L. (1992). Evaluating interviewer style for collecting industry and occupation information. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 661-666.

Collins, C. L. (1975). *Comparison of month-to-month changes in industry and occupation codes with respondent's report of change: CPS job mobility study.* Response Research Staff Report No. 75-5. Washington, DC: U.S. Bureau of the Census.

Copeland K. R., and Rothgeb, J. M. (1990). Testing alternative questions for the Current Population Survey. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 63-71.

Couper, M. P., and Burt, G. (1993, in press). Interviewer reactions to computer-assisted personal interviewing (CAPI). *Bureau of the Census Proceedings of the Annual Research Conference*, 429-443.

Department of Labor. (1981). *Final report of the Secretary of Labor on the Recommendations of the National Commission on Employment and Unemployment Statistics*. Transmitted to Congress, October 26.

Dippo, C. S. (1993). Designing an instrument for computer-assisted data collection in the Current Population Survey. *Bulletin of the International Statistical Institute*, Contributed Papers 49th Session, 1, 373-374.

Dippo, C. S., Kostanich, D. L., and Polivka, A. E. (1994, in press). Effects of methodological change in the Current Population Survey. *American Statistical Association Proceedings of the Section on Survey Research Methods.*

Dippo, C. S., Polivka, A. E., Creighton, K. P., Kostanich, D. L., and Rothgeb, J. M. (1994). *Redesigning a questionnaire for computer-assisted data collection: The Current Population Survey experience*. Paper submitted for publication.

Edwards, W. S. (1991, May). *Measuring question comprehension and its effects on response error: An "hours worked" case study*. Paper presented at the annual meeting of the American Association for Public Opinion Research, Phoenix, AZ.

Edwards, W. S., Levine, R. and Cohany, S. (1989). Procedures for validating reports of hours worked and for classifying discrepancies between questionnaire reports and validations totals. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 496-501.

Esposito, J. L., Campanelli, P. C., Rothgeb, J. M., and Polivka, A. E. (1991). Determining which questions are best: Methodologies for evaluating survey questions. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 46-53.

Esposito, J. L., and Hess, J. C. (1992, May). *The use of interviewer debriefings to identify problematic questions on alternative questionnaires*. Paper presented at the annual American Association for Public Opinion Research Conference, St. Petersburg, FL.

Fisher, R. (1994, in press). Logistic regression analysis of CPS Overlap Survey split panel data. *American Statistical Association Proceedings of the Section on Survey Research Methods*.

Fisher, R., Thompson, J., Robison, E., and Welch, M. (1993). Variance estimation in the CPS Overlap Test. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 802-807.

Fracasso, M. P. (1989). Reliability and validity of response categories for open-ended questions in the Current Population Survey. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 481-485.

Fracasso, M. P. (1989, August). *A procedure for the development of an alternative set of category choices using the open-ended questions in the Current Population Survey*. Paper presented at the annual meeting of the Southeastern Psychological Association, Washington, DC.

Gaertner, G. H., Cantor, D., Gay, N. L., and Shank, S. E. (1989). Tests of alternative questions to elicit occupational information in the CPS. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 490-495.

Hamel, H. (1979). Two fifths of discouraged workers sought work during prior six month period. *Monthly Labor Review, 102*, 58-60.

Hess, J. (1993). *Results of the MLS Interviewer Knowledge Concepts Test*. Memorandum for the CPS Questionnaire Evaluation Steering Committee. Washington DC: Bureau of the Census.

Kojetin, B. A., and Cashman, E. R. (1994, May). *Prior interviewing and computer experience and evaluation of CAPI training*. Paper presented at the annual Field Technologies Conference, Danvers, MA.

Kominski, R., and Siegel, P. M. (1993). Measuring education in the Current Population Survey. *Monthly Labor Review, 116(9)*, 33-38.

Kostanich, D. L., and Cahoon, L. S. (1994). *Effect of design differences between the parallel survey and the new CPS*. CPS Bridge Team Technical Report 3, dated March 4.

Lent, J., Miller., S, and Cantwell, P. (1994, in press). Composite weights for the Current Population Survey. *American Statistical Association Proceedings of the Section on Survey Research Methods.*

Martin, E. A. (1987). Some conceptual problems in the Current Population Survey. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 420-424.

Martin, E. A., Campanelli, P. C., and Fay, R. E. (1991). An application of Rasch Analysis to questionnaire design: Using vignettes to study the meaning of "work" in the Current Population Survey. *The Statistician, 40*, 265-276.

Martin, E. A., and Polivka, A. E. (1992). The effect of questionnaire redesign on conceptual problems in the Current Population Survey. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 655-660.

Martin, E. A., and Polivka, A. E. (1994). *Diagnostics for redesigning survey questionnaires: Measuring work in the Current Population Survey*. Paper submitted for publication.

McKay, R. B. (1992). Cultural factors affecting within household coverage and proxy reporting in Hispanic (Salvadoran) households. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 614-618.

McKay, R. B. (1993). Undercoverage of Hispanics in household surveys. *Monthly Labor Review, 116(9), 38-42.*

McKay, R. B. (1994, in press). The Spanish translation of the redesigned CPS: Lessons learned. *American Statistical Association Proceedings of the Section on Survey Research Methods.*

McKay, R. B. (1994, in press). Touch-tone data entry for CPS sample expansion. *American Statistical Association Proceedings of the Section on Survey Research Methods.*

McKay, R. B., and LaVallee, A. (1993, May). *The Hispanic version of the redesigned CPS questionnaire: Applying sociolinguistic and survey research methods to translating survey questionnaires*. Paper presented at the annual meeting of the American Association for Public Opinion Research, Pheasant Run, IL.

McKay, R. B., Malik, A., and Tupek, A. (1992, May). *Potential use of touch-tone recognition technology in the Current Population Survey*. Paper presented at the annual Field Technology Conference, St. Petersburg, FL.

McKay, R. B., and Sangster, R. L. (1994, May). *Touch-tone data entry for household reporters*. Paper presented at the annual Field Technologies Conference, Danvers, MA.

Miller, S. M. (1994). *What would the unemployment rate have been had the redesigned Current Population Survey been in place from September 1992 to December 1993? : A measurement error analysis*. CPS Bridge Team Technical Report 1, dated March 3.

Miller, S. M. (1994, in press). Measurement error modeling of the CPS and parallel survey data. *American Statistical Association Proceedings of the Section on Survey Research Methods.*

Mullin, P. A., and Tonn, B. E. (1993). Factors associated with proxy knowledge of employment-related information. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 948-953.

National Commission on Employment and Unemployment Statistics. (1979). *Counting the Labor Force*. Washington, DC: Government Printing Office.

Palmisano, M. (1988). *Some results on the application of laboratory research techniques to survey methodology*. Paper presented at the meetings of the American Association for the Advancement of Science.

Palmisano, M. (1989, August). *Respondent understanding of key labor force concepts used in the CPS*. Paper presented at the annual meetings of the American Statistical Association, Washington, DC. (Not published in conference proceedings).

Polivka, A. E. (1994). *Comparisons of labor force estimates from the parallel survey and the CPS during 1993: Major labor force estimates*. CPS Overlap Analysis Team Technical Report 1, dated March 18.

Polivka, A. E., and Martin, E. A. (1992, May). *The use of vignettes in pretesting and selecting questions*. Paper presented at the annual meeting of the American Association for Public Opinion Research, St. Petersburg, FL.

Polivka, A. E., and Miller, S. M. (1994, December). *The CPS after the redesign: Refocusing the economic lens*. Paper to be presented at the Bureau of Economic Research conference.

Polivka, A. E., and Rothgeb, J. M. (1993). Overhauling the Current Population Survey: Redesigning the questionnaire. *Monthly Labor Review, 116(9)*, 10-28.

President's Committee to Appraise Employment and Unemployment Statistics. (1962). *Measuring employment and unemployment*. Washington, DC: U.S. Government Printing Office.

Rothgeb, J. M. (1982). *Summary Report of July Follow-up of the Unemployed*. Unpublished memorandum. Washington, DC: Bureau of the Census.

Rothgeb, J. M. (1994). *Revisions to the CPS questionnaire: Effects on data quality*. CPS Overlap Analysis Team Technical Report 2, dated April 6.

Rothgeb, J. M., and Cohany, S. R. (1992). The revised CPS questionnaire: Differences between the current and proposed questionnaires. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 649-654.

Rothgeb, J. M., and Hess, J. C. (1992, May). *The role of response distribution and item non-response analyses in evaluating alternative question wordings during the redesign of the Current Population Survey*. Paper presented at the annual meeting of the American Association for Public Opinion Research, St. Petersburg, FL.

Rothgeb, J. M., Polivka, A. E., and Creighton, K. P. (1992). *Measurement changes in the redesigned CPS questionnaire*. Paper presented at the Joint Committee Meeting of the Census Advisory Committee of the American Statistical Association and the Census Advisory Committee on Population Statistics.

Rothgeb, J. M., Polivka, A. E., Creighton, K. P., and Cohany, S. R. (1991). Development of the proposed revised Current Population Survey questionnaire. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 56-65.

Schwartz, N. (1987, June). *Cognitive aspects of labor surveys in a multi-national context*. Paper prepared for the OECD Working Party on Employment and Unemployment Statistics.

Shoemaker, H. H. (1993). Results from the Current Population Survey CATI Phase-In project. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 710-714.

Shoemaker, H. H., Bushery, J. M., and Cahoon, L. S. (1989). Evaluation of the use of CATI in the Current Population Survey. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 361-366.

Stein, R. L., and Levine, D. B. (1965). Research in labor force concepts. *American Statistical Association Proceedings of the Social Statistics Section*, 218-226.

Tanur, J. M., and Lee, J. (1992). Youth unemployment measured by the CPS. *American Statistical Association Proceedings of the Section of the Section on Survey Research Methods*, 667-671.

Tanur, J. M., Lee, J., and Shin, H. (1991). Intensity and conceptualizations of job search among youths. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 391-396.

Tanur, J. M., Lee, J., and Shin, H. (1992). Reports of youths' job searches in the CPS. *Bureau of the Census Proceedings of the Annual Research Conference*, 317-352.

Tanur, J. M., and Shin, H. (1990). Reporting job search among youths: Preliminary evidence from reinterviews. *Bureau of the Census Proceedings of the Annual Research Conference*, 746-776.

Tanur, J. M., and Shin, H. (1990, August). *Youths, reporting job search, and the Current Population Survey*. Paper presented at the annual meetings of the American Statistical Association, Anaheim, CA. (Not published in conference proceedings).

Thompson, J. (1994). *Mode effects analysis of major labor force estimates*. CPS Overlap Analysis Team Technical Report 3, dated April 14.

Thompson, J., and Fisher, R. (1994, in press). Two-sample McNemar Tests for complex survey data. *American Statistical Association Proceedings of the Section on Survey Research Methods*.

Thompson, J., and Parmer, R. (1994, in press). A comparison of nonparametric methods with parametric methods for the CPS CATI/CAPI mode effects analysis. *American Statistical Association Proceedings of the Section on Survey Research Methods*.

Tiller, R. (1994, in press). Predicting the National unemployment rate that the old CPS would have produced. *American Statistical Association Proceedings of the Section on Survey Research Methods*.

Tiller, R., and Welch, M. (1994). *Predicting the national unemployment rate that the "old' CPS would have produced*. CPS Bridge Team Technical Report 2, dated March 15. (See also addenda dated March 4 and April 1)

U.S. Bureau of Labor Statistics. (1986). *Report of the BLS-Census Bureau Questionnaire Design Task Force*. Staff report dated November. Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics.

U.S. Bureau of Labor Statistics. (1987). *Second Report of the BLS-Census Bureau Questionnaire Design Task Force*. Staff report dated February. Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics.

U.S. Bureau of Labor Statistics. (1988, June). *Response errors on labor force questions based on consultations with Current Population Survey interviewers in the United States*. Paper prepared for the OECD Working Party on Employment and Unemployment Statistics.

U.S. Bureau of Labor Statistics and National Center for Education Statistics. (1979). *Feasibility study for a project on improvement of occupational information*. Report of the National Occupational Information Coordinating Committee, U.S. Department of Commerce Contract No. AO-A01-788-00-1349. Washington, DC: U.S. Bureau of Labor Statistics and National Center for Education Statistics.

U.S. Bureau of the Census. (1988). *Final report on planning for a post-1990 redesign of the Current Population Survey*. Unpublished report.

Waksberg, J., and Pearl, R. B. (1965). New methodological research on labor force measurements. *American Statistical Association Proceedings of the Social Statistics Section*, 227-237.

Walsh T., et al. (1988). *Evaluation of CATI data quality and costs in the Current Population Survey*. Unpublished research report dated October 27. Washington DC: Bureau of the Census.

Westat, Inc., and AIR, Inc. (1989). *Research on hours of work questions in the Current Population Survey*. Final Report, Bureau of Labor Statistics Contract No. J-9-J-8-0083. Rockville, MD: Westat, Inc.

Westat, Inc., and AIR, Inc. (1989). *Research on industry and occupation questions in the Current Population Survey*. Final Report, Bureau of Labor Statistics Contract No. J-9-J-8-0083. Rockville, MD: Westat, Inc.

Westat, Inc., and AIR, Inc. (1992). *Summary of activities to evaluate the dependent interviewing procedure on the Current Population Survey*. Final Report, Bureau of Labor Statistics Contract No. J-9-J-8-0083. Rockville, MD: Westat, Inc.

AN OVERVIEW OF THE 1992 CENSUS OF AGRICULTURE CATI SYSTEM

Jeanette K. Mon
Bureau of the Census

I. Introduction

Computer assisted telephone interviewing, or CATI, is an enumeration methodology in which responses received by telephone are interactively entered, edited, and coded into data files. The CATI system adopted by the Bureau of the Census provides call scheduling management, contains manager search functions, and produces various monitoring/progress reports.

Uses of the CATI system can vary from collecting present indicators for research polling to accepting detailed enumeration for surveys and/or censuses. Responses can be used to produce skip patterns and are subjected to consistency and magnitude checks. In a typical situation, the interviewer reads the question displayed on the computer screen to the respondent and records the response by keying the appropriate entry; then, the computer performs checks (i.e., validity, comparative), stores the response, and proceeds to the next question. This process continues until all questions have been asked.

For the 1992 Census of Agriculture, the CATI system was used to perform follow-up action for specific nonrespondents. This paper is intended to provide an overview of the developmental and processing phases of this system as well as the handling of output resulting from the 1992 Census of Agriculture CATI System.

II. Background

In 1973, Census Bureau executives became interested in CATI after seeing a demonstration by a private research firm. After several years of research and consultation with other such firms, universities, and computer vendors, the Bureau tested CATI at the University of California's CATI site during the 1978 Current Population Survey. In the early 1980's, the Census Bureau established a CATI project. Hardware and software were acquired in order to construct the Bureau's own CATI system. During the project's first year, staff reviewed the design and capabilities of CATI systems at academic institutions and private firms. Working with the Berkeley and Michigan Survey Research centers, the staff prepared basic requirements for the Census CATI system. From 1982 through 1984, Census conducted CATI research and development surveys from a Telephone Bridge Facility set up in Suitland, Maryland headquarters. The first major testing of this system was conducted for followup of nonrespondents in the 1982 National Survey of Natural and Social Scientists and Engineers, and the 1982 Census of Agriculture.

The first Census Bureau telephone facility, the Hagerstown Telephone Center, (HTC), opened in January 1985. An additional telephone facility opened in Tucson, Arizona in early 1992.

The census of agriculture is required by law under Title 13, United States Code, section 142(a) and 191, which states that an agriculture census be taken in 1979, 1983, and in every fifth year after 1983, covering the prior year. As previously mentioned, Agriculture Division's initial use of the CATI system in the 1982 Census of Agriculture was part of a test to review the system as a viable method of data collection. Agriculture Division (AGR) selected approximately 10,000 delinquent large farm cases for enumeration using this system. The follow-up process for the remaining cases used a clerical unit of operators who called respondents and manually recorded data on an agriculture report form.

For the 1992 Census of Agriculture, a CATI system was developed primarily to address the "large farms" which had not responded to the mailed questionnaire as in 1982. The reasons for using the CATI system for followup were management efficiency, cost effectiveness, and availability of the operation (staff/hardware) from the decennial census. There were also other advantages such as eliminating the data keying step, promoting a paperless census, and using CATI's management capabilities for monitoring and scheduling cases.

III. Overview of the Agriculture CATI System

A. Agriculture Division's Use of CATI

In addition to large farm followup, the AGR CATI system was used for the Nonresponse Survey and low response county projects. The Nonresponse Survey involved contacting a sample of nonrespondents from the main agriculture census. Data collected for this survey was used to determine the proportion of nonresponse cases that are farms and to weight census totals to account for the nonresponse. In the low response county project, a sample of nonrespondents in counties that had not achieved a 75% response rate were selected for contact. The goal of this project was to ensure that published 1992 Census of Agriculture data were based on responses from at least 75% of each county's mailout cases. In the Nonresponse Survey a different (shorter) version of the interviewing instrument was used while the low response county project used the large farm follow-up instrument. Since the system used for large farm followup provided the groundwork for these two projects, this paper will deal only with AGR's experience with the large farm CATI system.

B. System Development

The development of the AGR CATI system was conducted by a Committee responsible for: writing system specifications; developing essential system components; administering tests; coordinating facility schedules; and providing training.

CATI Committee

The 1992 Census of Agriculture CATI Committee consisted of 15 knowledgeable individuals from five divisions within the Census Bureau that are familiar with the various parts of the system. Even though much effort went into the brainstorming, learning, and decisionmaking involved with the development phase, this group remained functional throughout the implementation process and, also assisted in the evaluation of the system. From the onset of the planning phase (August 15, 1991) to the close of the AGR CATI operations (September 30, 1993) several committee personnel changes occurred; however, in most instances, the strategic persons remained involved in this task. The Committee consisted of representatives from:

- AGR to specify the needs for data collection;
- Economic Programming Division (EPD) to facilitate the input and output data;
- Field Division (FLD) to implement the operation with Hagerstown and Tucson staffs;
- Demographic Surveys Division (DSD) to program the QISC interviewing instrument; and,
- Systems Support Division (SSD) to provide support of the CATI system.

For the first six months, the Committee held biweekly meetings to plan the development phase. During this period, the Committee prepared a flowchart and activity schedule; identified and assigned the required specifications; and, scheduled the project for testing and production at the CATI facilities. Attachments A and B are the processing flowchart and the Committee's activity schedule, respectively. After the planning phase, the Committee divided into workgroups to facilitate the development of their respective CATI tasks and reconvened monthly to assure the ongoing progress of each assignment.

During the remainder of 1992, AGR discussed aspects of the interviewing instrument such as sequence and wording of questions, availability and content of referral screens, and consistency checks between responses. This exchange of ideas

resulted in additional refinements in the instrument.

CATI Specifications

As required of all CATI projects, basic system specifications are necessary from the sponsor to provide the framework for system development. Such standard specifications were identified by FLD and SSD committee members and delegated to appropriate members for development.

Many of these specifications required AGR personnel to go through a "learning" period prior to development. In some instances, the concept and format for these documents as recognized by CATI system personnel were complex and involved considerable time for one to become knowledgeable enough to prepare the required documents.

Agriculture Division wrote several other CATI specifications and procedures to explain issues such as training, problem solving, and handling other details not addressed by the basic CATI system specifications. Attachment C is a list of the CATI specifications.

System Tests

Three tests were conducted to refine the system for production. These tests (September 28 & October 28, 1992 and January 14, 1993) were generally conducted in the same manner. Experienced interviewers at the Hagerstown CATI facility telephoned AGR personnel for enumeration. The "mock respondents" were comprised of AGR staff from several different areas of responsibility. Some were given scripts of varying situations (i.e., nonagriculture, refusal, complete) while others presented their own scenario. The tests checked for the following items:

- questions needing rewording and/or additional clarification;
- appropriate routing of interviewing screens from question to question;
- correct output coding for case types and responses (keycoded & nonkeycoded);
- appropriate transfer and installation of files;
- appropriate input file content;
- improvement ideas from interviewers and/or "mock" respondents; and,
- other aspects within the process requiring attention.

The tests also provided sample output files and status tables for review. The output files were passed on to EPD for subsequent reformat testing. The status tables were reviewed for format changes and/or programming errors.

Training

In addition to the usual CATI system training provided to interviewers, AGR supplemented the training with a 2-day intensive instruction course. This course included technical subject-matter learning, pronunciation of practical agricultural terms, a short session in unit conversion using a calculator, practice interviews, and other exercises needed for enumeration. Attachment D is the training schedule.

Training materials for the AGR CATI project were written in FLD's Training Branch based on information submitted by AGR. The materials included:

- Self Study Guide,
- Workbook for Training,
- Paired Practice Interviews Booklet,
- Final Review Exercises,
- Guide for Training CATI Interviewers, and
- Evaluation of Self Study and Classroom Training.

During the training sessions, a reference binder of general information was given to each participant. In view of the voluminous amount of technical detail involved with the AGR subject matter, interviewers were instructed to review this material and use it for assistance as needed. The contents of this binder included a report form guide, alphabetic crop listing, unit conversion chart, and glossary of terms.

Input File Preparation

Each state was processed as a separate file/survey. EPD was responsible for the creation of forty-nine state files to install at the CATI facilities. (Hawaii cases were called in Jeffersonville because of the unique nature of their products.) These files consisted of cases not received in the 1992 Census of Agriculture universe which were preidentified as a "large farm."

Each state file was then processed in directory assistance (DA) subunits at the CATI facilities to obtain missing telephone numbers and correct inconsistent area code/telephone number combinations. At the beginning of the CATI process, the cases which did not yield a "good" telephone number from the DA subunit were left in the state input file for appropriate output coding and to facilitate accounting of all cases in the output file. However, this procedure was modified during the first wave by assigning these cases the appropriate coding upon file installation and omitting them from the calling que.

In addition to correcting the state files for missing telephone numbers or inconsistent area code/telephone number combinations, the files were updated to reflect mail returns received after the creation of each preliminary state file. EPD produced a file of satisfied Census File Numbers or CFNs (i.e., mail-ins, other resolved). They updated this "alert" file daily for use in amending the respective state CATI file by removing satisfied cases from the calling que. This amendment process was conducted one day prior to interviewing and continued daily until state closeout.

C. CATI Processing

Large Farm CATI Schedule for 1993

In order for the entire large farm CATI follow-up process to run efficiently, AGR coordinated with EPD, FLD and SSD to develop a schedule that notified each division of their timely interaction within the process. Attachment E is the schedule showing state workloads and respective dates for each step of processing. The schedule was broken down into 5 waves with approximately 10 states in each wave. States were listed in priority order according to other AGR processing dates. FLD, together with AGR, divided the states between the two telephone centers (Tucson and Hagerstown).

AGR decided to have one state file installed at each site to test the system before installing all of wave 1--Delaware was installed at Hagerstown and Oregon was installed at Tucson. Interviewing began at Hagerstown and Tucson on February 22 and March 8, respectively. As interviewing progressed, FLD installed the remaining state files in both waves 1 and 2 to allow a backlog of available states to call. As the CATI sites ran low on available cases, CATI site managers notified FLD and AGR to approve installation of other states/waves. In May, for instance, Tucson requested more cases in the Pacific time zone to accommodate interviewers who worked late. Consequently, California and Alaska became wave 3A. Wave 4 was also divided into two waves--4 and 4A.

Approximately five weeks were allotted for the interviewing process. As the CATI interviewing progressed, it became necessary to extend some states' closeout dates so that the interviewing process for other states started later than originally scheduled. When this was done, the dates for EPD and Secondary Source Unit (SSU) processing were changed accordingly.

CATI interviewing stopped one day prior to CATI closeout to allow for instrument/output file manipulation. The output files were sent to EPD for reformatting which required three days before sending the files to SSU in Jeffersonville. The

SSU needed about 13 days to resolve these cases, i.e., determine whether or not they were in the scope of the agriculture census. All in-scope cases were then merged with general census processing while out-of-scope cases were coded as such and no further processing was needed.

This schedule was instrumental in keeping all aspects of the processing on track and meeting the goal to complete the entire CATI process by October 1. The CATI sites closed out all states by September 12; and, EPD and SSU processing were completed by September 22.

Field Division Support

During the CATI operations, feedback on problems related to the interviewing process, subject matter, or operational/system efficiency, was relayed from FLD to AGR by the Field Division liaison and resolved, in most cases, via the electronic mail system or telephone. For example, in the beginning of the operation, interviewers questioned whether crops grown in years prior to 1992 but sold in 1992 should be included as 1992 sales. The FLD liaison referred this concern to AGR. AGR personnel informed the FLD liaison that these sales should be included in 1992 sales totals. This was later reiterated in a "briefing note" or bulletin and distributed to supervisors at both CATI sites.

As with any problem posed to AGR, after finding a solution, AGR periodically prepared a briefing note to document/clarify changes to the interviewing or operating process. These changes were discussed at the pre-shift meetings at both CATI sites to keep the interviewers up-to-date on the CATI process. Attachment F is an example of a briefing note.

These notes were essential for transferring information between AGR and FLD at headquarters as well as supervisors and interviewers at the CATI sites. They were also helpful in accounting and documenting each problem's resolution.

Subject Matter Support

To keep members of AGR abreast of CATI status, a "CATI Newsletter" was developed and sent out about every six weeks from March 1992 to January 1993. The newsletter was written by AGR committee members and was distributed to Agriculture Division's Chief, Assistant Division Chiefs, and Branch Chiefs. These newsletters contained information such as status of specifications development; current CATI issues; and, schedules for CATI testing, training, and production. In response to these newsletters, division personnel were able to comment on the CATI process at hand.

Also, AGR analysts made several site visits to the telephone facilities throughout the AGR CATI operation--from initial testing to closeout--to observe the process, monitor interviews, answer technical questions, collect improvement ideas from facility personnel, as well as identify and resolve any instrument and/or processing problems. For example, prior to live interviewing, staff visited the facilities to monitor three tests and to conduct initial training on the Research Operation. Site visits were also scheduled at the start of interviewing at both the Hagerstown and Tucson sites. After this initial "getting started" stage, AGR scheduled site visits about once a month--usually at the start of a new wave of states or at closeout.

Monitoring

Monitoring was an important part of the CATI process. While site visits provided a method for AGR to monitor the flow of facility processing and handling of technical information, CATI management was responsible for monitoring the quality of the interviewing process. Also, AGR received system-generated reports that provided up-to-date workload status for monitoring CATI progress.

To monitor the quality of the interview, the CATI system contains a built-in network which allows supervisors or analysts to listen to an interview on the telephone while simultaneously viewing the computer screen to see how the information is recorded by the interviewer. The facilities maintain specific standards for such monitoring. During initial monitoring (the first three months interviewers are on the job), about 10% of an interviewer's interviewing time each month are monitored. After that, (systematic monitoring) supervisors monitor at least 2.5% of each interviewer's active interviewing time as well as any "special needs" monitoring. In half-hour monitoring sessions, supervisors were able to unobtrusively observe interviewers to identify their strengths and weaknesses. The supervisor would complete a monitoring report each time an interviewer was monitored and provide feedback to the interviewer.

FLD staff sent periodic monitoring reports to AGR via the electronic mail system. At each CATI site and for each wave of the survey, these reports showed: average number of interviewers; number of login hours; and number of monitoring sessions. In the early stages of interviewing, these reports indicated areas where monitoring was inadequate and notice was given to the facility to increase monitoring. At the end of the survey, a final report was received from FLD which included the overall monitoring rate. Attachment G is a copy of this final report.

Every day during processing, AGR received two system-generated reports reflecting the previous day's work for each active state--a Sample Status Report and an Interviewer Performance Report.

Attachment H is an example of a Sample Status Report. Each day AGR extracted and recorded pertinent totals from active state reports such as numbers of cases classified as in-scope, out-of-scope, and remaining active. Cumulative numbers were also extracted and recorded for the following case types: resolved; mail receipts; duplicates; secondary source; and, refused.

Attachment I is an example of an Interviewer Performance Report. This report showed each individual interviewer's call attempts categorized by outcome code and summarized all call attempts by interviewers excluding/including supervisors and managers. Each day AGR extracted from active state reports and recorded data for: number of interviewers used; login hours; and minutes of in-scope calls. At closeout, AGR received a cumulative interviewer's report for the entire period of the survey (i.e., Monday, March 22, 1993 - Sunday, May 2, 1993).

The extracted data from these reports was used in developing several spreadsheets, such as:

- CATI PROGRESS REPORT - A separate spreadsheet for each state (survey) showing the daily progress. For each state report, data is given for each active interviewing day plus state totals.

- LARGE FARM CATI WORKLOAD - A spreadsheet of all states at both telephone facilities by wave/state showing workload totals. The totals are shown for each telephone facility as well as for the U.S.

- CATI INTERVIEWER REPORT - A spreadsheet for each month showing active states (surveys) and giving the daily counts of interviewers working, login hours, and completed interviews for in-scope cases. These monthly reports helped to verify telephone company monthly charges.

Research Operation

When a respondent indicated that they received multiple forms under different names or CFNs for the same operation, the interview was handled as a "possible duplicate". The Research Operation was set up at both CATI facilities to review and/or verify the possible duplicate situations. This involved EPD support in acquiring access to various

agriculture databases for both CATI sites. In view of the security risk, CATI personnel were granted "read only" capability when accessing the AGR Research Menu network routines.

By accessing the "name & address" and "check-in" databases, trained CATI researchers were able to search for duplicate entries in the mail list and verify receipt of the duplicate report forms as needed. The training materials for this operation were developed by AGR analysts and required AGR personnel to conduct agriculture-specific training at both facilities due to FLD's unfamiliarity with this subject.

The benefits of this operation provided notable strides toward customer service as well as improved processing. In contrast to the spontaneous handling of possible duplicate situations in 1992, in 1987 these cases were referred to clerical reviewers after the conversation was ended; and, in cases where subsequent research indicated that an interview was still needed, the respondent was recontacted. The utility of the 1992 operation was invaluable since it resulted in the identification of about 12,000-18,000 duplicates on the AGR mail list and prevented callbacks to an estimated equal number of possible duplicate cases which were unverified by the researcher.

"Claims filed" cases, or situations in which the respondent claimed that they had already mailed their questionnaire, were not routed to the Research Operation since the alert file was updated daily and any mail returns were deleted from the calling que. In these cases, the interviewer prompted the respondent to complete the interview knowing that the form had not been received.

Other Production Processing

Several other details were handled during the CATI interviewing process. For example, specific procedures were established to process "send form" and "Title 13" requests. All other situations which were not predesignated in training were handled as "supervisor referrals."

"Send form" cases, in which the respondent would not agree to be interviewed by telephone but requested a form be sent for completion and mailing, were coded as such and systematically set for a callback 10 days later. The CATI supervisors would check the system daily to pick up all cases with send-form coding and refer them to AGR for form mailing. If the report form was received within 10 days, the CFN was automatically coded as resolved and deleted from the CATI calling que via the daily alert system. Otherwise, CATI interviewers would recontact the respondent to complete the

interview by telephone.

Some respondents requested a copy of the Census Law (Title 13 of the United States Code) that requires them to respond. Again, the interviewer was directed to relay this information to the supervisor who consequently notified AGR of these cases on a daily basis.

D. CATI Output/Results

The results of the CATI attempts were transmitted to Economic Programming Division (EPD) in the form of four files:

- o Answer file - interviewer coding including respondent data and/or interviewer remarks for resolved cases;
- o F7 file - interviewer remarks made during the interviewing process for resolved cases;
- o History file - "snapshot" of installed cases showing each time accessed; and,
- o Case Master file - system management information for each installed case.

Every interviewing day, an Answer and F7 file for each active state were transmitted to EPD for their subsequent reformatting. At each state's closeout, a cumulative version of all four files was transmitted to EPD. EPD was responsible for assuring the receipt of these files and subsequent processing for merging these files into the 1992 Census of Agriculture operations.

Daily Processing

The daily state answer files were the source of in-scope and out-of-scope records.

For in-scope records, besides the answers collected which referred to the farm unit, some data items were created based on interviewer responses to reflect:

- o CATI processing codes,
- o flag indicators of "zero" or "none" responses for specific items,
- o summing of valid duplicate items,
- o geographic changes, and,
- o section indicators.

In addition, these in-scope records were used to update various AGR databases. Whenever there were verbatim responses made during the interview (i.e., "other" crops/livestock not

listed, state or county changes, etc.), they were retrieved and loaded onto the "notes" database for analyst review. A check-in status code indicating "CATI satisfied" was assigned to each record in the "check-in" database. Corrections, if any, in the name, address, and/or telephone number fields were carried to the "name & address" database.

Out-of-scope records identified in the daily answer files were updated in the check-in database by assigning the appropriate status code.

Confirmed refusal cases were also included in the daily answer files to make these cases available for the next processing step (Secondary Source Unit), and, to alleviate storage space in the respective state's active case file. However, in all states, the confirmed refusal cases were processed by EPD from the cumulative state answer file.

The daily F7 files, which consisted of auxiliary notes made by the interviewer during the interview, were reformatted and loaded onto the notes database to provide analysts with supplemental information for in-scope cases.

Closeout Processing

The closeout answer file was cumulative and provided a single source of confirmed refusals and other unresolved cases requiring SSU processing. Cases routed to SSU were identified by specific final code. Attachment J is a list of these case types and the respective final codes.

Similar to the reformatting of in-scope records, interviewer comments, if any, and processing data items were created for each SSU case and forwarded to the Jeffersonville facility for assistance in determining the resolution.

Other Output Processing

The history file consisted of one record for each time a case was accessed. This file was transmitted to EPD at the closeout of each state survey. A separate history record was created for each case to show all the calls that were made as well as other actions where calls were not made (such as, when one "quits out" of a case before dialing). This file was routinely used by facility management to track the progression of a particular case. From this file, AGR manipulated information using SAS software to produce tables and graphs for management analysis. See Attachments K through N for examples of AGR charts produced from various history files. These state history files are in storage for potential studies at a later date.

A case master file was also transmitted to EPD at the closeout of each state survey. This file consisted of one record per case which contained essential information for that case for case-management purposes. Information in the case master file was always kept current--former values of variables were not retained. CATI facility personnel usually accessed this file as a convenient source for specific case details. For the purpose of analyzing the system, data from this file could be used for reviewing number of call attempts for specific outcomes, callback time preferences, and so on. Since AGR's plans for CATI evaluations are not complete at this time, these state case master files are being stored for possible later use.

IV. Summary of Agriculture CATI System Results

A. Final Case Accounting

Of the 152,815 cases installed for CATI resolution, 57,708 cases (37.8%) were enumerated as in the scope of the agriculture census while 15,148 (9.9%) were out-of-scope. Through research, 4,924 cases (3.2%) were determined to be duplicates of satisfied cases. After CATI file installation, 29,312 cases (19.2%) were omitted from the interviewing process as they were received in the mail. A total of 45,723 unresolved cases (29.9%), including 3,188 confirmed refusals, were routed to SSU for additional processing.

The results of the 1992 large farm CATI followup with comparative statistics from the 1987 nonCATI operation are shown in Attachment O, Table 1, Results of the 1992 and 1987 Delinquent Large Farm Followup.

When comparing the results of the 1987 clerical and 1992 CATI operations, one notices the higher in-scope rate (+5.2%) and slightly lower out-of-scope rate (-1.3%) in 1992. In both operations, a large number of cases were unresolved requiring SSU processing. These SSU cases (29.9% in 1992 and 31.7% in 1987) include respondent-contacted situations (i.e., insufficient partial interviews, refusals, callbacks) as well as noncontact cases (i.e., busy, no answer, never tried). Since the level of mail receipts is independent of operation type (nonCATI or CATI), these numbers only reflect the timely creation of the input call file prior to the telephoning process.

Another observation which is evident upon review is the difference in processing claims filed and possible duplicate situations. As mentioned previously, in 1992 these cases received spontaneous handling versus 1987's procedure to discontinue the telephone call, perform research, and, if

necessary, recontact the respondent. Claims filed cases resulted in prompting for an interview in 1992 (since the calling que was updated daily by the alert file). Possible duplicate cases were verified by the Research Operation and appropriate action was taken.

The different environments provided by the 1992 and 1987 systems make it difficult to conclude whether operation type is responsible for the resulting differences. Dissimilar management philosophies, physical locations/office set up, and experience level of interviewers are some factors which inhibit any attempt to denote collection methodology as the single reason for varying results.

B. Cost Analysis

The cost of the 1992 delinquent large farm followup amounted to almost $2.1 million or $13.56 per case. This amount included staffing for supervision/interviewing, system development (excluding AGR analysts), communications, equipment, and general administrative support.

Attachment P, Table 2, shows data for comparing the cost of the 1992 and 1987 delinquent large farm follow-up operations. When the 1987 operation cost is adjusted to show 1993 dollars, the 1992 CATI system cost savings is $2.17 or 13.8% less per case than the 1987 nonCATI operation. The overall cost of the 1992 CATI followup shows an approximate increase in expenses of $146,000 or 7.6% more than the 1987 nonCATI followup.

It would be misleading to use these statistics to make any conclusions concerning cost savings between the two operations. The expenses shown in Table 2 are actual charges made to the respective projects. Costs for general staff time (AGR and other), questionnaire design, use of previously procured hardware, and communication expenses covered by blanket costs are some of the factors which need to be addressed in making a system comparison. In addition, there are operational differences (i.e., quality control, supervisor/interviewer ratio) and post-interview processing (i.e., data entry needs, file reformatting) which should be addressed to gauge the benefits of each operation's yield versus cost.

V. Evaluation of the Agriculture CATI System

Reactions to the CATI system are necessary for the proper assessment of the operation. They provide valuable input towards building future CATI systems. To develop an improved efficient data collection system, AGR requested feedback from various persons

involved with the 1992 CATI operation.

A. Feedback from CATI Committee/Site Staff

After interviewing concluded, the CATI Committee was asked to submit any ideas or comments on the entire CATI developmental and operational process. In general, the suggestions consisted of need for: 1) more lead time for system development; 2) additional in-depth testing; 3) periodic retraining for interviewers during the process; 4) tracking of file transmissions; and, 5) prompt downloading of closed out state files.

In conjunction with AGR, FLD held five debriefing sessions at both CATI sites. These open forum meetings were conducted by the FLD liaison with AGR personnel present to answer questions and monitor delivery of the presentation.

To gather feedback for all three AGR CATI projects (i.e., large farm, nonresponse, low response counties), general questions as well as separate questions pertaining to the specific project were asked. The time set for these sessions was such that the majority of the states/surveys were closed out or approaching close-out. Large farm interviewing officially ended on September 12. Nonresponse and low response county projects closed out on August 20 and September 30 respectively.

The three debriefing sessions at the Tucson CATI facility were held on August 31 (10:00 AM and 2:00 PM) and September 1 (10:00 AM). Nine people attended the first morning session and eight people attended each of the other two sessions. In Hagerstown, two sessions were held on September 9 with nine attendees at 10:00 AM and seven attendees at 2:00 PM.

All personnel involved with the CATI operation (interviewers, supervisors, directory assistance callers, researchers, on-site analysts) were given an opportunity to complete the handout of questions (Attachment Q). However, a sample of these persons were invited to attend the debriefing sessions for extended discussions on the AGR CATI operation.

Overall, the sessions yielded many constructive suggestions for improving the system. The majority of problems cited were repeated in each of the sessions. A detailed summary of the responses and comments were consolidated into the 1992 Census of Agriculture Manual for reference. Such documentation will be useful to AGR in prospective CATI systems.

B. Evaluation Documentation

As traditionally conducted for the majority of census processing systems, AGR staff is currently documenting the CATI process in an evaluation paper. This paper will include: 1) system development and processing details; 2) CATI merge processing; and, 3) 1992 CATI versus 1987 nonCATI data review. Aspects of CATI recognized by CATI Committee members, CATI site staff, and AGR subject matter analysts as requiring "enhancements" or additional address are also noted in this synopsis for improvement of future AGR CATI systems.

Other CATI-associated studies in AGR are underway as well. A 1997 Census of Agriculture planning team has been formed to establish criteria for a CATI system for the next census based on review of CATI versus nonCATI data in the 1992 census. For example, this team has solicited feedback from AGR subject matter analysts to isolate problematic data items in CATI cases. Subsequently, the team will create and evaluate tallies of these suspect items for CATI and nonCATI records to verify their source.

C. CASIC Presentation

After the AGR CATI operations concluded in September 1993, AGR presented a brief overview to the Computer Assisted Survey Information Collection (CASIC) group. This group at the Census Bureau is dedicated to the automation of data collection through computerized technologies. In view of the fact that CATI is undergoing "redesign" in the Bureau to produce a centralized computer assisted data entry system across three sites (present facilities: Hagerstown and Tucson; new third site: Jeffersonville, IN.), and that the 1992 Census of Agriculture was the largest single project to use the CATI system, CASIC requested feedback from AGR to identify system strengths and weaknesses. Many of the comments submitted to CASIC are also included in this paper.

VI. Suggestions for Future Agriculture CATI Systems

Based on the success of the 1992 AGR CATI System, especially the processing phase, it is my opinion that CATI can be an efficient tool for follow-up data collection. Considering the 1992 incurred cost, beneficial automated features, paperless reporting, and installation of the Research Operation, this system is a viable method of enumeration. However, AGR's large scale encounter with this processing technology has lent itself to many improvement ideas.

As a result of this experience with CATI, the following are suggestions for enhancing prospective AGR CATI systems:

A. Developmental Recommendations

- Learn the CATI system to become aware of its potential. Sponsors should be knowledgeable of the capabilities to maximize its use and deal with system shortcomings properly for other processing coordination.

- Allow adequate time for testing. If possible, mock respondents not associated with the survey should be used in the tests.

B. Processing Recommendations

- Search for continual improvement ideas during processing. For example, review of case management files could possibly show significant "trends" toward specific callback times. In this case, interviewer scheduling should be adjusted accordingly.

- Acknowledge retraining needs. Even though AGR training was perceived to be complete, several interviewers indicated in the debriefing sessions that they had additional questions and/or situations requiring reverification after initial interviewing.

- Acquire on-site support staff for communications problems. Operations can be severely hampered awaiting personnel from headquarters to address problems.

C. Systems Recommendations

- Allow for customized output answer files. Files with standardized CATI output pose inefficient storage/handling situations when only a portion of the data is needed.

- Allow for customized progress reporting. System generated reports are primarily used by facility management and are difficult for sponsors to use.

- Develop interviewer paths for various situations. Allow for regional crops/livestock, basic information collection for reluctant respondents, and easy-skip paths for respondents indicating involvement with few commodities.

- Permit sponsor "read only" access to CATI management system. Since full access can be a security problem "read only" capabilities would allow the sponsor to make spontaneous status checks, track specific case

responses, etc. without FLD intervention.

- Automate system file transmissions. Routing of input and output files was manually monitored requiring much attention for operations consisting of numerous files/surveys.

- Explore hardware/software for CATI enhancements such as automated dialing to prevent misdialing, instrument software to allow for word processing capabilities, etc.

VII. Conclusion

Due to the use of the CATI system, improvements in this census' followup of delinquent large farms are apparent in the production phase regarding more interactive features and in eliminating subsequent processes.

Recognizing the long-term goals of the Bureau to use the latest technologies for more efficient and productive systems, CATI has offered several improved features over the 1987 Census of Agriculture clerical follow-up unit. Most notable of these are the systematic call scheduler, management research capabilities, and automated status report generation. In addition, other attributes yielded by CATI usage are savings from eliminating a separate data keying phase, omitting the need for paper questionnaires and their handling/storage, and providing better customer service through spontaneous research of possible duplicate cases requiring fewer recontacts.

Comparing the CATI to nonCATI follow-up operations is difficult. As mentioned in Section IV, part B, Cost Analysis, incurred costs are an unfair indicator of cost savings between 1992 CATI and 1987 nonCATI since savings and production efficiencies are not considered. Also, to conclude that any data disparity is a result of the collection methodology would be a misinterpretation in view of operational differences. Because of these factors which make comparison of CATI versus nonCATI systems very complex, AGR will be cautious in drawing conclusions--other than the realized progress achieved in process automation and better service to the respondent.

Aside from the deficiencies in comparing CATI and nonCATI operations for conclusive statements, AGR plans to review CATI-originated data with nonCATI data in the 1992 Census of Agriculture by examining the level of edited and/or imputed statistics. Rather than conclude that the operation source is the cause of any differences, the 1997 Census of Agriculture CASIC planning team will use these comparisons to detect dissimilarities in the data capture (i.e., wording and/or sequence of questions).

In summary, the 1992 Census of Agriculture CATI System was successful in terms of completing the follow-up phase in an efficient and timely manner. The automated features including the customized Research Operation provided many enhancements over the 1987 clerical operation. AGR hopes to build upon their experience with this initial system and use CATI in the 1997 Census.

REFERENCES

Ferrari, Pamela W. "An Evaluation of Computer-Assisted Telephone Interviewing Used During the 1982 Census of Agriculture," Research Paper, Agriculture Division, January 1986.

Mon, Jeanette K. and King, Elizabeth A. "Evaluation of the 1992 Census of Agriculture CATI System," in 1992 Census of Agriculture Manual, Project Manual MC, Chapter 27, Subchapter A, Document 2, Draft version, April 1994.

LIST OF ATTACHMENTS

A. Agriculture Division Large Farm CATI Flowchart

B. CATI Committee Schedule

C. 1992 Agriculture Census CATI Specifications

D. Schedule of Training

E. Large Farm CATI Schedule for 1993

F. Large Farm Followup Briefing Note (sample)

G. Monitoring Report as of 9/25/93

H. Sample Status Report (sample)

I. Interviewer Performance Report (sample)

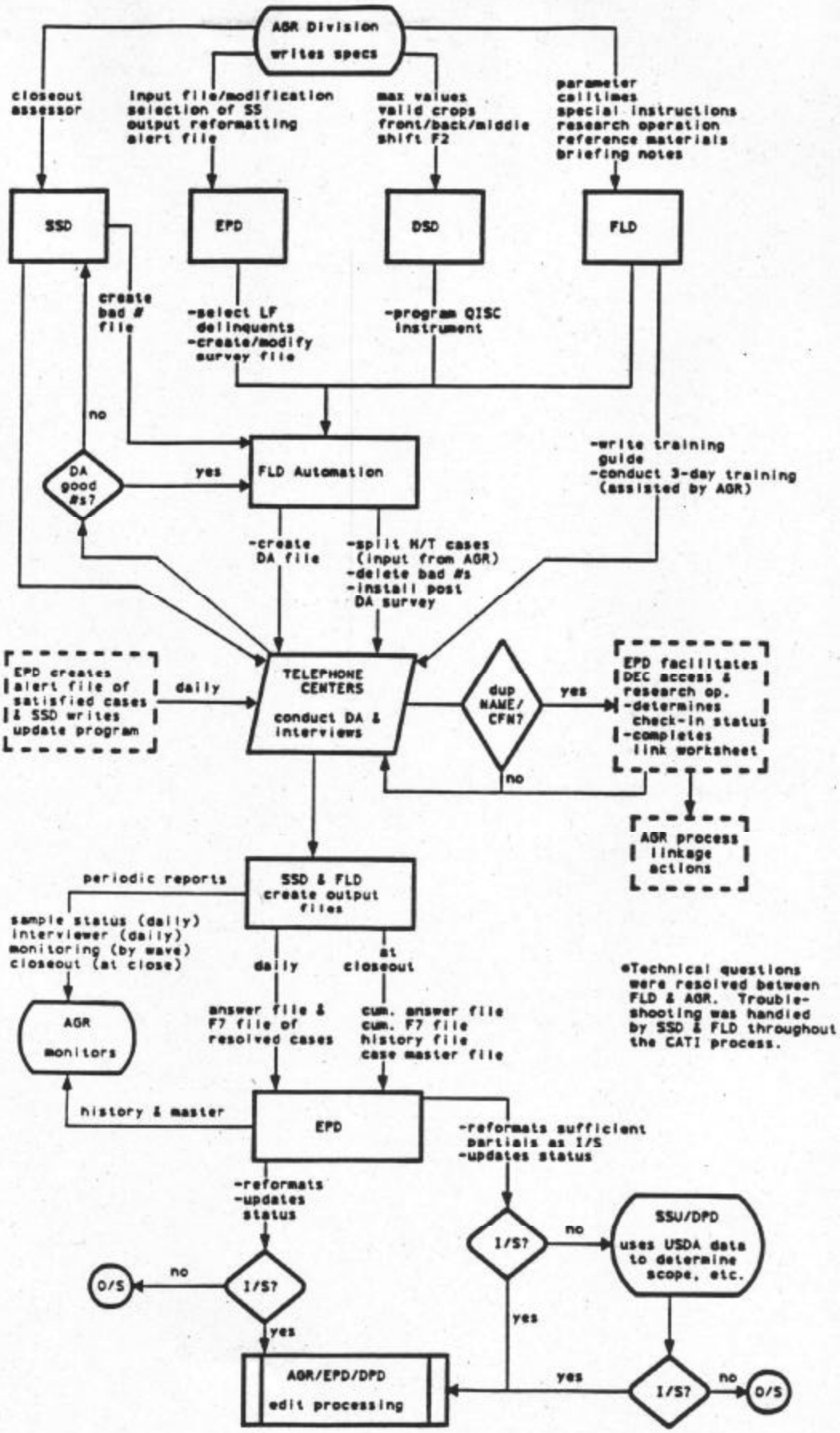
J. List of SSU Designated Case Type by Final Code

K-N. Agriculture charts/graphs from history record data

O. Table 1, Results of the 1992 & 1987 Delinquent Large Farm Followup

P. Table 2, Cost Comparison of the 1992 and 1987 Large Farm Followup

Q. Telephone Center Debriefing Handout


AGR Division writes specs
closeout assessor
input file/modification selection of SS output reformatting alert file
max values valid crops front/back/middle shift F2
parameter calltimes special instructions research operation reference materials briefing notes
SSD
EPD
DSD
FLD
create bad # file
-select LF delinquents -create/modify survey file
-program QISC instrument
no
DA good #s?
yes
FLD Automation
-write training guide -conduct 3-day training (assisted by AGR)
-create DA file
-split H/T cases (input from AGR) -delete bad #s -install post DA survey
EPD creates alert file of satisfied cases & SSD writes update program
daily
TELEPHONE CENTERS conduct DA & interviews
dup NAME/ CFN?
yes
EPD facilitates DEC access & research op. -determines check-in status -completes link worksheet
no
AGR process linkage actions
periodic reports
SSD & FLD create output files
sample status (daily) interviewer (daily) monitoring (by wave) closeout (at close)
daily
at closeout
AGR monitors
answer file & F7 file of resolved cases
cum. answer file cum. F7 file history file case master file
•Technical questions were resolved between FLD & AGR. Trouble-shooting was handled by SSD & FLD throughout the CATI process.
history & master
EPD
-reformats sufficient partials as I/S -updates status
-reformats -updates status
I/S?
no
SSU/DPD uses USDA data to determine scope, etc.
O/S
no
I/S?
yes
yes
AGR/EPD/DPD edit processing
yes
I/S?
no
O/S

Revised 3/2/92

CATI COMMITTEE SCHEDULE

Activity	Start Date	Complete Date	Responsible Person/Div
1. QISC Specs	9/91	3/92	Battle/King Mon
Modifications	11/91	8/92	DSD/AGR
2. Testing			
a. QISC	3/92	8/92	Battle/King/ Mon/DSD
b. Transmission to/from EPD	6/92	1/93	FLD/SSD/AGR/ EPD
c. Complete system (Pretest)	8/92	8/92	AGR/FLD/SSD/ DSD
d. State test	2/93	2/93	AGR/FLD/SSD/ EPD/DSD
3. Other Specifications	2/92	12/92	AGR/FLD/SSD/ EPD/DSD
4. Training Package/ Personnel	6/92	2/93	AGR/FLD
5. Implementation of System	3/93	8/93	FLD/AGR/SSD/ EPD
6. Evaluation	9/93	5/94	Committee

Attachment C

1992 AGRICULTURE CENSUS CATI SPECIFICATIONS

Manual Number	Title (explanation)
92EAG-A-MC-10-E-04	CATI Specs for Large Delinquent Telephone Followup (QISC programming specs for middle of interview--sections 1-31 of report form--designating outcome codes)
92EAG-A-MC-10-E-06	CATI Input File Specs (components and layout of the input file)
92EAG-A-MC-10-E-08	Valid Crop Listing for CATI QISC Specs (listing valid crops for each state)
92EAG-A-MC-10-E-10	CATI Assessor Specs (assignment of agendum and final codes which designate the next action)
92EAG-A-MC-10-E-12	CATI Front/Back for Delinquent Large Farm Followup (QISC programming specs for introduction and closing portions of the interview designating outcome codes)
92EAG-A-MC-10-E-13	CATI Test File Specs (input file specs for testing)
92EAG-A-MC-10-E-14	CATI Closeout Specs (assignment of final, outcome and agendum codes which designate whether sufficient partial or secondary source processing)
92EAG-A-MC-10-E-15	CATI Parameter Specs (a consolidation of several specs for a file used to pass individual survey info to the CATI system and specifies other file management requirements)
92EAG-A-MC-10-E-19	CATI Responses to Commonly Asked Questions (shift F2 option available to interviewers during interview)
92EAG-A-MC-10-E-22	Research Operation for Large Farm CATI (instructions for accessing AGR database to verify "claims filed" or "duplicate" situations)
92EAG-A-MC-10-E-23	Special Instructions for CATI Supervisors (supervisor duties regarding the research operation and certain agendums, i.e., refusals, language barrier, etc.)

92EAG-A-MC-10-E-24	CATI Output Reformatting Specs (info concerning the raw CATI answer file and needed reformatting/data manipulation)
92EAG-A-MC-10-E-30	CATI Calltimes Specs (setting times and days for making calls)
92EAG-A-MC-10-E-31	Large Farm CATI Schedule for 1993 (dates set for each step of processing)
92EAG-A-MC-10-E-38	Selection of Secondary Source Cases (specs for creating SSU file)
92EAG-A-MC-10-E-40	Change in CATI Interviewing Process for Acres (K787)
92EAG-A-MC-10-E-41	CATI Workbook for Training (interviewer training)
92EAG-A-MC-10-E-42	CATI Self Study for Interviewers (part of interviewer training)
92EAG-A-MC-10-E-43	CATI Final Review Exercise (part of interviewer training)
92EAG-A-MC-10-E-44	Large Farm CATI Briefing Notes (notes to supervisors/interviewers instructing how to handle particular situations)
92EAG-A-MC-10-E-45	Large Farm CATI Newsletters (information to keep AGR staff abreast of what was happening regarding AGR CATI)

Schedule of Training

Day 1

Chapter	Topic	Estimated Length	Total Elapsed Time
Chapter A	Introduction to the 1992 Census of Agriculture	1/2 hr.	1/2 hr
Chapter B.	Using the Agriculture Census Reference Materials	1 hr.	1 1/2 hrs
	Break	1/4 hr.	1 3/4 hrs
Chapter C	Farm Operations	1 1/2 hrs.	3 1/4 hrs
	Break	1/4 hr.	3 1/2 hrs
Chapter D	Walk-through Training Interview	3/4 hrs.	4 1/4 hrs
Chapter E	Concepts and Procedures	3/4 hrs.	5 hrs
Chapter F	Two Paired Practice Interviews	1 1/4 hrs.	6 1/4 hrs

Day 2

Chapter G	Another Walk-through Interview	1 1/4 hrs.	1 1/4 hrs
	Break	1/4 hr.	1 1/2 hrs

Day 2 (continued)

Chapter	Topic	Estimated Length	Total Elapsed Time
Chapter H	Additional Concepts and Procedures	3/4 hr.	2 1/4 hrs.
	Break	1/4 hr.	2 1/2 hrs.
Chapter I	Two Paired Practice Interviews	1 1/4 hrs.	3 3/4 hrs.
	Break	1/4 hr.	4 hrs.
Chapter J	Final Review Exercise	3/4 hr.	4 3/4 hrs.

Attachment E

Revised 10/5/93

LARGE FARM CATI SCHEDULE FOR 1993

WAVE	STATE	ESTIMATED MAXIMUM WORKLOAD	ACTUAL WORKLOAD	ACTUAL DATES FOR: CREATE FILE	DIR'Y ABST.	BEGIN INTVIEW	END INTVIEW	CATI CLOSEOUT	EPD PROCESS	SECOND. SOURCE	DATA KEY'D CLOSEOUT
1-H	DE (51)	726	1,147	2/4	2/8	2/22	3/27	3/28	5/4	5/7	5/19
1-H	IN (32)	4,001	3,882	2/25	2/26	3/2	4/17	4/18	5/5	5/7	5/18
1-T	OR (82)	3,482	3,866	2/25	3/1	3/8	4/10	4/11	5/21	5/24	6/3
1-T	WI (35)	3,152	3,026	3/4	3/8	3/16	4/17	4/18	5/14	5/14	5/28
1-T	WA (91)	3,528	3,528	3/4	3/8	3/23	5/1	5/2	5/27	5/27	6/10
1-T	IA (42)	5,966	5,796	3/4	3/8	3/23	5/1	5/2	5/18	5/19	6/1
1-T	MO (43)	4,789	5,251	3/4	3/8	3/30	5/13	5/14	5/25	5/27	6/9
1-H	OH (31)	3,562	3,475	3/4	3/8	3/16	5/1	5/2	5/26	6/7	6/15
1-H	MD (52)	1,376	1,577	3/4	3/8	3/16	5/1	5/2	6/9	6/10	6/24
1-H	WV (55)	825	857	3/4	3/8	3/16	4/17	4/18	6/10	6/11	6/25
2-H	VA (54)	3,018	3,007	3/4	3/8	3/22	5/1	5/2	6/11	6/15	6/28
2-T	KS (47)	3,394	3,282	3/4	3/8	4/2	5/8	5/9	6/16	6/21	7/7
2-T	IL (33)	4,158	3,830	3/4	3/8	4/6	5/8	5/9	6/25	6/28	7/14
2-H	MI (34)	3,069	3,282	3/4	3/8	3/22	5/1	5/2	7/6	7/6	7/26
2-T	WY (83)	2,352	2,606	3/4	3/8	4/7	5/8	5/9	7/16	7/21	8/5
2-H	ME (11)	718	830	3/4	3/8	3/29	5/1	5/2	7/11	7/14	8/2
2-H	NH (12)	234	243	3/4	3/8	3/29	4/17	4/18	7/11	7/14	8/2
2-H	VT (13)	1,020	1,217	3/4	3/8	3/29	5/1	5/2	7/12	7/15	8/2
2-H	CT (16)	427	534	3/4	3/8	3/29	5/1	5/2	7/12	7/15	8/2
2-H	MA (14)	597	685	3/4	3/8	3/29	5/1	5/2	7/17	7/20	8/2
2-H	RI (15)	84	113	3/4	3/8	3/29	4/17	4/18	7/17	7/20	8/2
3-T	CO (84)	2,695	2,249	4/8	4/12	5/3	6/10	6/11	7/20	7/23	8/13
3-T	ID (82)	2,790	2,406	4/8	4/12	4/22	6/1	6/2	7/24	7/27	8/18
3-H	NJ (22)	1,800	1,805	4/8	4/12	4/22	5/22	5/23	7/24	7/27	8/18
3-H	PA (23)	6,084	4,940	4/8	4/12	4/22	6/5	6/6	7/31	8/3	8/20
3-H	NY (21)	4,550	4,053	4/8	4/12	4/22	6/5	6/6	7/25	7/28	8/20
3-H	KY (61)	3,672	3,472	4/8	4/12	4/22	6/5	6/6	7/27	7/30	8/24
3-H	TN (62)	3,414	3,407	4/8	4/12	4/22	6/5	6/6	7/29	8/3	8/30
3-T	UT (87)	2,404	2,391	4/8	4/12	5/1	6/1	6/2	7/30	8/5	8/6
3-T	NE (46)	5,661	5,064	4/8	4/12	4/19	6/14	6/15	8/2	8/5	8/3
3-T	MN (41)	8,247	7,515	4/8	4/12	4/21	6/24	6/25	8/3	8/12	8/20
3-T	MT (81)	4,480	3,950	4/8	4/12	5/10	6/24	6/25	8/4	8/13	8/24
3A-T	CA (93)	8,803	7,627	5/5	5/10	6/4	7/17	7/18	8/5	8/10	8/17
3A-T	AK (94)	825	789	5/5	5/10	6/4	7/17	7/18	8/7	8/10	12/10
4-H	TX (74)	9,311	8,022	6/3	6/4	6/15	6/14	6/15	8/17	8/25	10/5
4-H	NC (56)	4,018	3,105	6/3	6/4	6/15	7/17	7/18	8/13	8/17	10/22
4-H	FL (59)	5,755	5,396	6/3	6/4	6/15	7/17	7/18	8/25	9/2	11/24
4-T	NV (88)	800	525	6/3	6/15	6/21	7/17	7/18	8/4	8/12	10/5
4A-H	ND (44)	4,006	3,419	6/3	6/11	6/15	7/31	8/1	8/16	8/19	11/3
4A-H	SD (45)	3,787	3,480	6/3	6/11	6/15	7/31	8/1	8/18	8/31	11/3
4A-T	AR (71)	6,374	4,870	6/3	6/15	6/22	7/17	7/18	8/19	9/1	11/5
4A-T	AZ (86)	1,432	1,141	6/3	6/10	6/22	7/17	7/18	8/23	9/1	11/19
4A-T	NM (85)	3,040	2,578	6/3	6/10	6/22	7/17	7/18	8/24	9/7	12/1
5-T	AL (63)	3,805	2,717	7/12	7/14	7/20	8/11	8/12	8/14	8/20	12/3
5-H	MS (64)	2,867	2,241	7/12	7/26	8/2	8/11	8/12	8/13	8/15	12/2
5-T	OK (73)	5,770	4,822	7/12	7/14	7/20	8/11	8/12	8/16	8/22	12/3
5-T	LA (72)	3,213	2,727	7/12	7/14	7/19	8/11	8/12	8/13	8/15	12/6
5-H	SC (57)	1,934	1,717	7/12	7/26	8/2	8/11	8/12	8/14	8/17	12/21
5-H	GA (58)	4,841	4,291	7/12	7/26	8/2	8/11	8/12	8/18	8/21	12/5
	TOTAL	166,778	152,815								

H = Hagerstown

T = Tucson

*Each state was updated for mail receipts the night before interviewing began

1992 Census of Agriculture CATI System
Large Farm Followup Briefing Note #5 - March 16, 1993

Thank you for your continued efforts in our data collection phase! Please note the following points:

- If the respondent indicates he/she has grapes, please read "grape vines" instead of "grape trees" as shown on the screen. (This will be corrected ASAP.)

- If respondent indicates he has peppermint, select "mint for oil" in section 2 or 7. They are the same.

- Only call the Research Operator when there is an actual duplicate situation (duplicate forms received for one operation with DIFFERENT names and/or CFNs). Remember to read the second sentence on the >multiforms< screen and the >claimsfile< screen before accepting a response.

- When "READ LISTING" appears on the screen, read the entire list.

  When "READ, IF NECESSARY" appears, you may use information previously given in the interview to decide if you need to read the entire list. For instance, in S9 the letter "P" appears next to the crops produced, so you want to probe for those particular items and then ask "Any others?" (it may not be necessary to read the entire list). If in doubt, read the entire list.

- The Shift F1 option brings up the >info-ref< screen which gives you the name and full address of the respondent. This option is useful for the interviewer in filling out the top portion of the worksheet when research is unavailable.

  After getting into the middle of the interview (beginning with S1), the Shift F1 >info-ref< screen will also give you the option "C" to change the respondent.

- Any CFNs reported as duplicates on the >check< screen will be displayed on the >research< screen. If the CFN is displayed, MOVE THE CURSOR by pressing enter to get to where you indicate whether the dup was found. DO NOT type in anything else on the line where the cursor first appears, unless there was no CFN displayed and you need to enter a CFN found by the Research Operator.

- On the >hello-2< screen, if you enter "11" (deceased & sold farm) you go to the >intro-b< screen where you should enter "8" (sold farm). The next question asks if operated during any part of 1992--if yes, the interview continues; if no, the interview ends.

Below find the requested Ag monitoring data as of September 25:

Hagerstown:

	Ave. # Intvrs.	Login Hours	Mon. sessions	Mon. rate
Wave 1	88	5757	311	5.4
Wave 2	73	4524	236	5.2
Wave 3	146	9550	269	2.8
Wave 4 & 5	73	14424*	375	2.6
Wave 5	198	3538	204	5.8

* these login hours include an underestimated amount of 14.7 for state 74

Tucson:

	Ave. # Intvrs.	Login Hours	Mon. sessions	Mon. rate
Wave 1	107	9283	144	1.55
Wave 2	122	4176	71	1.7
Wave 3	153	12281	440	3.6
Wave 4	141	9094	335	3.7
Wave 5	93	7200	243	3.4

TTC has included AG Model Drop in their counts.

ACRLF57 CREATED: 09/06/93 01:52:05

SAMPLE STATUS REPORT

FOR: SUNDAY SEPTEMBER 5, 1993

REMAINING ACTIVE CASES BY TIME ZONE

		TOTAL	EASTERN	CENTRAL	ROCKY	PACIFIC	YUKON	HAWAII
NO TELEPHONE NUMBER	(1)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
NEW CASE	(0,2)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
UNREACHED CASE	(3)	1 0.02	1 0.02	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
DIRECTORY ASSISTANCE NEEDED	(4)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
CONTACT, RESP NOT REACHED	(6)	4 0.07	3 0.06	1 0.14	0 0.00	0 0.00	0 0.00	0 0.00
CONTACT, RESP REACHED	(7)	11 0.20	10 0.21	1 0.14	0 0.00	0 0.00	0 0.00	0 0.00
AWAITING MAILBACK	(8)	9 0.16	8 0.17	1 0.14	0 0.00	0 0.00	0 0.00	0 0.00
REFUSAL/NO PROGRESS	(9)	5 0.09	5 0.10	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
PARTIAL INTERVIEW	(14)	25 0.45	21 0.44	4 0.57	0 0.00	0 0.00	0 0.00	0 0.00
HOSTILE BREAK/OFF REFUSAL	(15)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
OTHER	(5,10-13,16-20)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
TOTAL/MARK	(0-20)	55 1.00	48 1.00	7 1.00	0 0.00	0 0.00	0 0.00	0 0.00

TODAY'S RESOLVED CASES BY TIME ZONE

		TOTAL	EASTERN	CENTRAL	ROCKY	PACIFIC	YUKON	HAWAII
COMPLETE (ALL ITEMS)	(1)	5 0.83	4 0.80	1 1.00	0 0.00	0 0.00	0 0.00	0 0.00
PARTIAL TYPE 1	(2)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
PARTIAL TYPE 2	(3)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
PARTIAL TYPE 3	(4)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
PARTIAL TYPE 4	(5)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
PARTIAL TYPE 5	(6)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
MAILED FORM/SAME CFN	(7)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
MAILED FORM/DUP CFN	(8)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
OTHER DELETION	(9)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
DECEASED/NO REFERRAL	(10)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
LANDLORD ONLY	(12)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
NONAGRICULTURE	(13)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
NONAGRICULTURE = 10 AC	(14)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
SOLD FARM	(15)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
SOLE OPERATOR MOVED/NO REFERRAL	(17)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
SOLE OPERATOR ILL/NO REFERRAL	(18)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
OTHER NONINTERVIEW	(19)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
UNCONVERTIBLE LANGUAGE BARRIER	(21)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
UNPUBLISHED NUMBER	(22)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
NO LISTING (LEARNED FROM DA)	(23)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
NO INITIAL NUMBER SUPPLIED	(24)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
CONFIRMED REFUSAL	(25)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
SEARCH CUTOFF	(26)	1 0.17	1 0.20	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
CLAIMS FILED/UNCONFIRMED	(35)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
MAIL RETURN UPDATE	(104)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
CLOSEOUT CUTOFF	(115-127)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
OTHER	(11,16,28,27-34,36-103,105-114, 128-200)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
TOTAL/FINAL	(1-200)	6 1.00	5 1.00	1 1.00	0 0.00	0 0.00	0 0.00	0 0.00

CUMULATIVE RESOLVED CASES BY TIME ZONE

	TOTAL	EASTERN	CENTRAL	ROCKY	PACIFIC	YUKON	HAWAII

COMPLETE (ALL ITEMS)	(1)	818	0.49	781	0.54	37	0.18	0	0.00	0	0.00	0	0.00	0	0.00
PARTIAL TYPE 1	(2)	9	0.01	9	0.01	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PARTIAL TYPE 2	(3)	2	0.00	2	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PARTIAL TYPE 3	(4)	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PARTIAL TYPE 4	(5)	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
PARTIAL TYPE 5	(6)	103	0.06	97	0.07	6	0.03	0	0.00	0	0.00	0	0.00	0	0.00
MAILED FORM/SAME CFN	(7)	65	0.04	61	0.04	4	0.02	0	0.00	0	0.00	0	0.00	0	0.00
MAILED FORM/DUP CFN	(8)	79	0.05	78	0.05	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00
OTHER DELETION	(9)	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
DECEASED/NO REFERRAL	(10)	3	0.00	3	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
LANDLORD ONLY	(12)	30	0.02	26	0.02	4	0.02	0	0.00	0	0.00	0	0.00	0	0.00
NONAGRICULTURE	(13)	73	0.04	71	0.05	2	0.01	0	0.00	0	0.00	0	0.00	0	0.00
NONAGRICULTURE = 10 AC	(14)	75	0.05	75	0.05	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
SOLD FARM	(15)	17	0.01	16	0.01	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00
SOLE OPERATOR MOVED/NO REFERRAL	(17)	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
SOLE OPERATOR ILL/NO REFERRAL	(18)	7	0.00	7	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
OTHER NONINTERVIEW	(19)	11	0.01	11	0.01	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
UNCONVERTIBLE LANGUAGE BARRIER	(21)	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
UNPUBLISHED NUMBER	(22)	18	0.01	17	0.01	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00
NO LISTING (LEARNED FROM DA)	(23)	261	0.16	110	0.08	151	0.72	0	0.00	0	0.00	0	0.00	0	0.00
NO INITIAL NUMBER SUPPLIED	(24)	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
CONFIRMED REFUSAL	(25)	17	0.01	16	0.01	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00
SEARCH CUTOFF	(26)	23	0.01	22	0.02	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00
CLAIMS FILED/UNCONFIRMED	(35)	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
MAIL RETURN UPDATE	(104)	51	0.03	50	0.03	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00
CLOSEOUT CUTOFF	(115-127)	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
OTHER(11,16,20,27-34,36-103,105-114, 128-200)		0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TOTAL/FINAL	(1-200)	1662	1.00	1452	1.00	210	1.00	0	0.00	0	0.00	0	0.00	0	0.00

AGRLF57 CREATED: 09/06/93 01:52:05

SAMPLE STATUS REPORT

FOR: SUNDAY SEPTEMBER 5, 1993

CUMULATIVE RESOLVED CASES SPECIFIC BY FINAL

		TOTAL	EASTERN	CENTRAL	ROCKY	PACIFIC	YUKON	HAWAII
COMPLETION SERIES	(1-6)	932 0.56	889 0.61	43 0.20	0 0.00	0 0.00	0 0.00	0 0.00
FORM REC/OTH DEL	(7-9)	144 0.09	139 0.10	5 0.02	0 0.00	0 0.00	0 0.00	0 0.00
CONTACT MADE/NO DATA	(10,12-15,17-19)	216 0.13	209 0.14	7 0.03	0 0.00	0 0.00	0 0.00	0 0.00
UNREACHABLE	(21-24)	279 0.17	127 0.09	152 0.72	0 0.00	0 0.00	0 0.00	0 0.00
CONFIRMED REFUSAL	(25)	17 0.01	16 0.01	1 0.00	0 0.00	0 0.00	0 0.00	0 0.00
SEARCH CUTOFF	(26)	23 0.01	22 0.02	1 0.00	0 0.00	0 0.00	0 0.00	0 0.00
UNCONFIRMED CLAIMS FILED	(35)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
MAIL RETURN UPDATE	(104)	51 0.03	50 0.03	1 0.00	0 0.00	0 0.00	0 0.00	0 0.00
CB SCHEDULED/SUFF PARTIAL	(115)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
PREREFUSAL/HOS BREAKOFF	(116)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
NEEDS RESEARCH WORK	(117)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
CB SCHEDULED/PARTIAL INT	(118)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
TEMP UNAVAILABLE	(119)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
UNCOMPLETED CALL	(120)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
LANGUAGE BARRIER	(121)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
UNCONFIRMED CLAIMS FILED (HOLD)	(122)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
WILL FILE/SEND FORM	(123)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
NEW CONT/CONF NUMBER	(124)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
NEW CONT/UNCONF NUMBER	(125)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
ANSWER SERV/MACH. LEFT MESSAGE	(126)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
NEVER TRIED	(127)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
OTHER	(11,16,20,27-34,36-103,105-114,128-200)	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00
TOTAL/FINAL	(1-200)	1662 1.00	1452 1.00	210 1.00	0 0.00	0 0.00	0 0.00	0 0.00

Attachment H (cont'd)

INTERVIEWER PERFORMANCE REPORT

September 6, 1993 01:52:42

Sunday, September 5, 1993

Survey: AGRLF97

TOTAL ALL INTERVIEWERS
(Excluding supervisors/managers)

OUTCOME OF CALL ATTEMPTS		COUNT OF CASES	PERCENT OF TOTAL CASES	TOTAL TIME (MINUTES)	PERCENT OF TOTAL TIME	MEAN TIME (MINUTES)
A. Complete interviews	(01)	5	0.08	128	0.54	25.60
B. Partial interviews	(02-06)	0	0.00	0	0.00	0.00
C. SUBTOTAL, COMPLETES AND PARTIALS	(01-06)	5	0.08	128	0.54	25.60
D. Other call progress	(40)	0	0.00	0	0.00	0.00
E. Out of scope	(07-16)	0	0.00	0	0.00	0.00
F. HH refusals and hostile breakoffs	(30-34)	1	0.02	8	0.03	8.00
G. Other noncooperation	(35-39)	2	0.03	2	0.01	1.00
H. Other noninterviews	(17-29)	0	0.00	0	0.00	0.00
I. No contact	(80-89)	24	0.41	33	0.14	1.38
J. Other outcomes	(41-79)	27	0.46	67	0.28	2.48
K. TOTALS	(01-89)	59	1.00	238	1.00	4.03

L. Total login time (minutes).....	369	
M. Total login time (hours).....	6.15	
N. Call progress/login hours ((C+D)/M).....	0.81	
O. Ratio of household refusals and hostile breakoffs to call progress (F/(C+D)).....	0.20	
P. Call progress rate ((C+D+E+F+G+H+J)/K).....	0.59	
Q. First refusal rate.....	0.13	
R. Total refusal/hostile breakoff conversion attempts with contact...		0
RESULTS OF CONVERSION ATTEMPT:		
S. Complete or Partial interview..		0
T. Out-of-Scope.....		0
U. Second refusal.....		0
V. Other noninterview.....		0
W. Refusal/Hostile breakoff conversion rate ((S+T+V)/R).....		0.00

INTERVIEWER PERFORMANCE REPORT

September 6, 1993 01:52:42

TOTAL ALL STAFF
(Including supervisors)

Sunday, September 5, 1993

Survey: AGRLF57

OUTCOME OF CALL ATTEMPTS		COUNT OF CASES	PERCENT OF TOTAL CASES	TOTAL TIME (MINUTES)	PERCENT OF TOTAL TIME	MEAN TIME (MINUTES)
A. Complete interviews	(01)	5	0.08	128	0.54	25.60
B. Partial interviews	(02-06)	0	0.00	0	0.00	0.00
C. SUBTOTAL, COMPLETES AND PARTIALS	(01-06)	5	0.08	128	0.54	25.60
D. Other call progress	(40)	0	0.00	0	0.00	0.00
E. Out of scope	(07-16)	0	0.00	0	0.00	0.00
F. HH refusals and hostile breakoffs	(30-34)	1	0.02	8	0.03	8.00
G. Other noncooperation	(35-39)	2	0.03	2	0.01	1.00
H. Other noninterviews	(17-29)	0	0.00	0	0.00	0.00
I. No contact	(80-89)	24	0.41	33	0.14	1.40
J. Other outcomes	(41-79)	27	0.46	67	0.28	2.48
K. TOTALS	(01-89)	59	1.00	238	1.00	4.04

L. Total login time (minutes).....	369
M. Total login time (hours).....	6.15
N. Call progress/login hours ((C+D)/M).....	0.81
O. Ratio of household refusals and hostile breakoffs to call progress (F/(C+D)).....	0.20
P. Call progress rate ((C+D+E+F+G+H+J)/K).....	0.59
Q. First refusal rate.....	0.13

R. Total refusal/hostile breakoff conversion attempts with contact...	0
RESULTS OF CONVERSION ATTEMPT:	
S. Complete or Partial interview..	0
T. Out-of-Scope.....	0
U. Second refusal.....	0
V. Other noninterview.....	0
W. Refusal/Hostile breakoff conversion rate ((S+T+V)/R).....	0.00

Attachment I (cont'd)

INTERVIEWER/FACILITY SUMMARY PAGE — September 6, 1993 01:52:42

Sunday, September 5, 1993 — Survey: AGRLF57

INTERVIEWER CODE	LOGIN TIME (MINUTES)	LOGIN TIME (HOURS)	COMPLETED INTERVIEWS	COMPLETES PER LOGIN HOURS	CALL PROGRESS PER LOGIN HOUR	RATIO OF HH REF TO CALL PROGRESS	REFUSAL CONVERSION RATE
FACILITY	369	6.15	5	0.81	0.81	0.20	0.00
ALL INTERVIEWERS	369	6.15	5	0.81	0.81	0.20	0.00
IM38	369	6.15	5	0.81	0.81	0.20	0.00

NUMBER OF INTERVIEWERS LISTED IN SUMMARY : 1

NUMBER OF SUPERVISORS/MANAGERS LISTED IN SUMMARY : 0

TOTAL NUMBER OF ALL USERS LISTED IN SUMMARY : 1

Attachment I (cont'd)

Attachment J

LIST OF SSU DESIGNATED CASE TYPE BY FINAL CODE

	TYPE OF CASE	FINAL CODE
1.	Insufficient partial	06
2.	Unconvertible language barrier	21
3.	Unpublished number	22
4.	No listing of telephone number	23
5.	No initial number supplied	24
6.	Confirmed refusal	25
7.	Search cutoff	26
8.	Callback scheduled, sufficient partial	115
9.	Prerefusal/hostile breakoff	116
10.	Needs research work	117
11.	Callback scheduled, insufficient partial	118
12.	Temporarily unavailable	119
13.	Uncompleted call, no contact on callback	120
14.	Language barrier	121
15.	Unconfirmed claims filed	122
16.	Will file, send form	123
17.	Never contacted, confirmed number	124
18.	Never contacted, unconfirmed number	125
19.	Answering service/machine, left message	126
20.	Never tried	127

1992 Census of Agriculture
Length of Call for Completed CATI Cases

OHIO (31)

MINUTES	1-15	16-30	31-45	46-60	>60	TOTAL
NUMBER OF CASES	118	719	348	115	7	1,307
	9.0%	55.0%	26.6%	8.8%	0.5%	

NORTH CAROLINA (56)

MINUTES	1-15	16-30	31-45	46-60	>60	TOTAL
NUMBER OF CASES	148	803	447	73	1	1,472
	10.1%	54.6%	30.4%	5.0%	0.1%	

NEW YORK (21)

MINUTES	1-15	16-30	31-45	46-60	>60	TOTAL
NUMBER OF CASES	135	1,048	451	100	9	1,743
	7.7%	60.1%	25.9%	5.7%	0.5%	

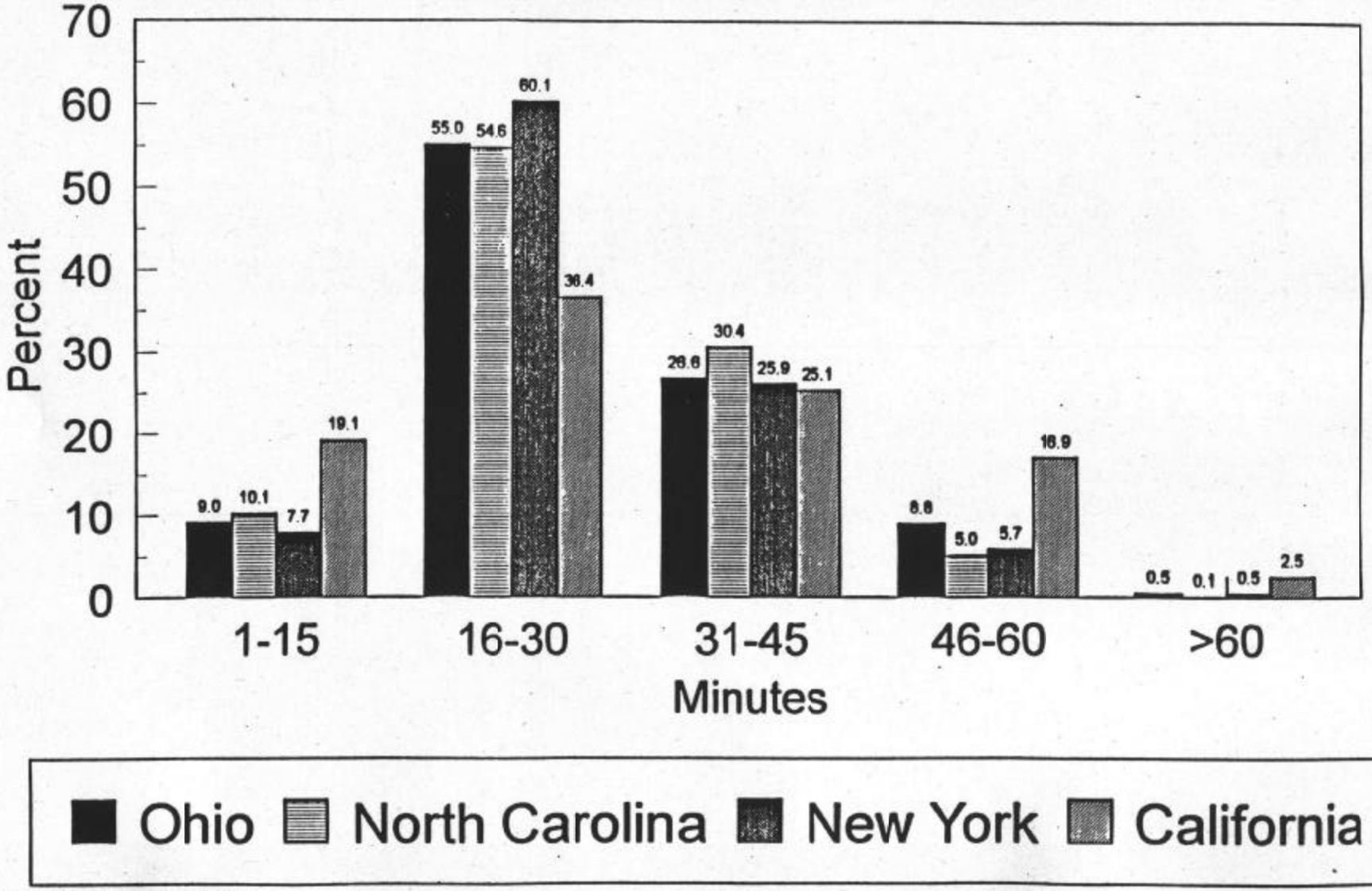
CALIFORNIA (93)

MINUTES	1-15	16-30	31-45	46-60	>60	TOTAL
NUMBER OF CASES	204	389	268	180	27	1,068
	19.1%	36.4%	25.1%	16.9%	2.5%	

AG CENSUS USA

1992 Census Of Agriculture CATI System

% of Completed Cases by Length of Call



Source: US Bureau of the Census
Agriculture Division

(Rev. 6/94)

1992 Census of Agriculture
Respondent Time for Completed CATI Cases

OHIO (31)

6:00 am to 6:59 am	7:00 am to 7:59 am	8:00 am to 8:59 am	9:00 am to 9:59 am	10:00 am to 10:59 am	11:00 am to 11:59 am	12:00 to 12:59 pm	1:00 pm to 1:59 pm	2:00 pm to 2:59 pm	3:00 pm to 3:59 pm	4:00 pm to 4:59 pm	5:00 pm to 5:59 pm	6:00 pm to 6:59 pm	7:00 pm to 7:59 pm	8:00 pm to 8:59 pm	9:00 pm to 5:59 am
2	106	103	86	81	64	98	95	86	90	95	98	97	92	97	17
0.2%	8.1%	7.9%	6.6%	6.2%	4.9%	7.5%	7.3%	6.6%	6.9%	7.3%	7.5%	7.4%	7.0%	7.4%	1.3%

TOTAL = 1,307

NORTH CAROLINA (56)

6:00 am to 6:59 am	7:00 am to 7:59 am	8:00 am to 8:59 am	9:00 am to 9:59 am	10:00 am to 10:59 am	11:00 am to 11:59 am	12:00 to 12:59 pm	1:00 pm to 1:59 pm	2:00 pm to 2:59 pm	3:00 pm to 3:59 pm	4:00 pm to 4:59 pm	5:00 pm to 5:59 pm	6:00 pm to 6:59 pm	7:00 pm to 7:59 pm	8:00 pm to 8:59 pm	9:00 pm to 5:59 am
0	180	115	71	63	60	113	94	78	66	84	81	113	109	123	72
0	12.7%	8.1%	5.0%	4.4%	4.2%	7.9%	6.6%	5.5%	4.6%	5.9%	5.7%	7.9%	7.7%	8.6%	5.1%

TOTAL = 1,422

NEW YORK (21)

6:00 am to 6:59 am	7:00 am to 7:59 am	8:00 am to 8:59 am	9:00 am to 9:59 am	10:00 am to 10:59 am	11:00 am to 11:59 am	12:00 to 12:59 pm	1:00 pm to 1:59 pm	2:00 pm to 2:59 pm	3:00 pm to 3:59 pm	4:00 pm to 4:59 pm	5:00 pm to 5:59 pm	6:00 pm to 6:59 pm	7:00 pm to 7:59 pm	8:00 pm to 8:59 pm	9:00 pm to 5:59 am
0	175	139	160	164	175	135	132	95	103	139	92	94	117	108	14
0.0%	9.5%	7.5%	8.7%	8.9%	9.5%	7.3%	7.2%	5.2%	5.6%	7.5%	5.0%	5.1%	6.4%	5.9%	0.8%

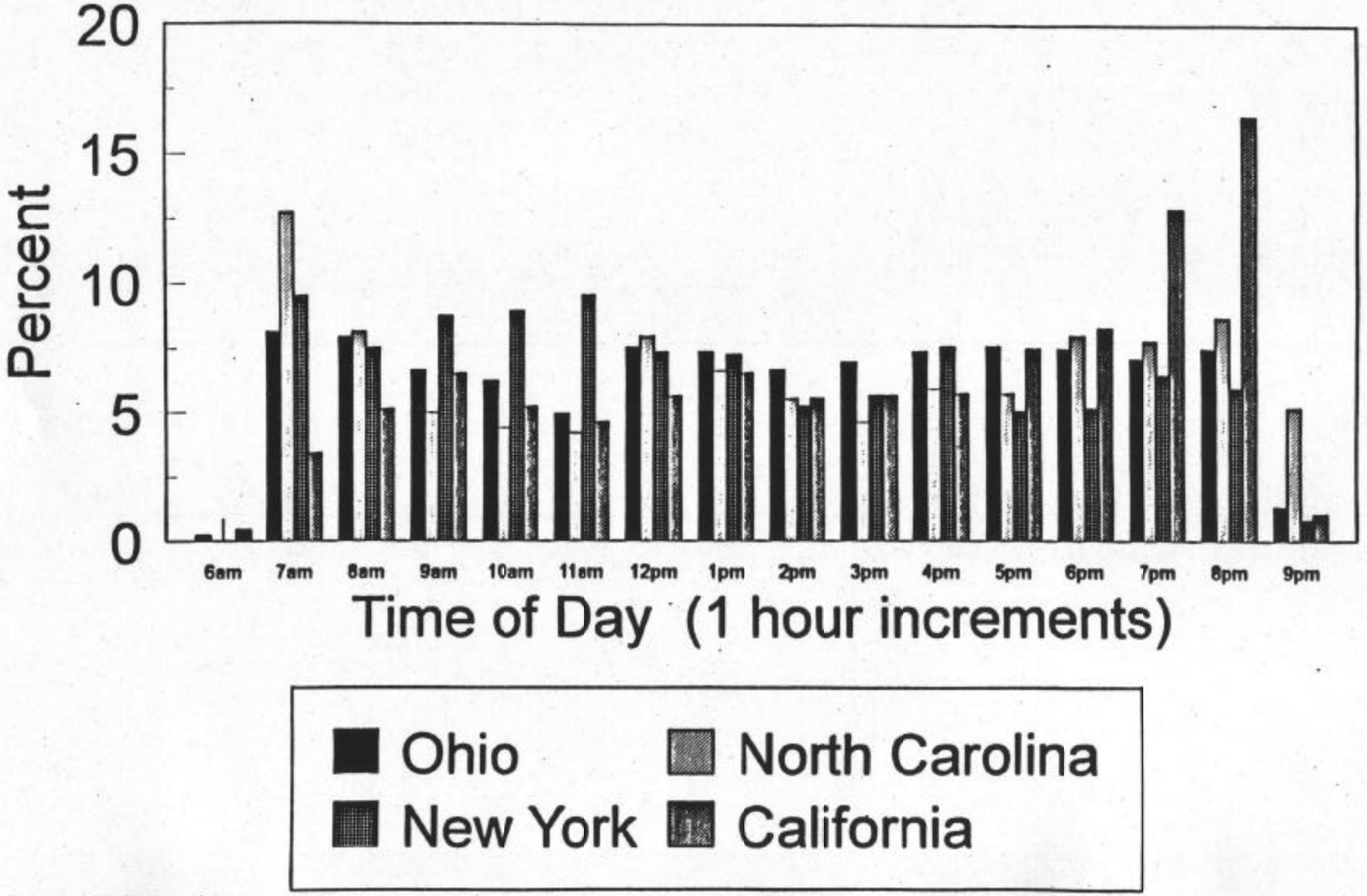
Total = 1,842

CALIFORNIA (93)

6:00 am to 6:59 am	7:00 am to 7:59 am	8:00 am to 8:59 am	9:00 am to 9:59 am	10:00 am to 10:59 am	11:00 am to 11:59 am	12:00 to 12:59 pm	1:00 pm to 1:59 pm	2:00 pm to 2:59 pm	3:00 pm to 3:59 pm	4:00 pm to 4:59 pm	5:00 pm to 5:59 pm	6:00 pm to 6:59 pm	7:00 pm to 7:59 pm	8:00 pm to 8:59 pm	9:00 pm to 5:59 am
8	72	109	140	112	99	121	140	118	119	123	159	175	274	352	21
0.4%	3.4%	5.1%	6.5%	5.2%	4.6%	5.6%	6.5%	5.5%	5.6%	5.7%	7.4%	8.2%	12.8%	16.4%	1.0%

1992 Census of Agriculture CATI System

% of Completed Cases by Respondent Time of Day



Source: U.S. Bureau of the Census
Agriculture Division

Attachment N
(Rev. 6/94)

Table 1

Results of the 1992 and 1987 Delinquent Large Farm Followup

CATEGORY	1992	1987
In-scope cases	57,708 (37.8%)	39,893 (32.6%)
Out-of-scope cases	15,148 (9.9%)	13,677 (11.2%)
Claims filed cases	NA[1]	20,030 (16.4%)
Duplicates	4,924 (3.2%)	NA[2]
Mail receipts	29,312 (19.2%)	10,000 (8.2%)
Secondary source	45,723 (29.9%)	38,800 (31.7%)
Total Workload	152,815	122,400

[1]Number of claims filed cases was not recorded in 1992. In these situations, if the respondent indicated mailing took place within the last week, a callback was systematically set for 7 days later. If mailing took place prior to the last week, the respondent was prompted to proceed with the interview. Any resulting refusals were sent to SSU.

[2]Number of duplicates was not recorded in 1987.

Table 2

Cost Comparison of the 1992 and 1987 Large Farm Followup

CATEGORY	1992	1987
Facility supervisory staff	$889,266	
Interviewing staff	$264,797	$626,000[1]
Benefits, other applied costs	$315,327	$323,000
Communications	$500,048	$122,000
Equipment, micro-computers	$49,982	$20,000
Facility administrative support, training, travel	$2,761	$158,000
Data Keying		$106,500
Other	$50,619	$210,795
TOTAL COST	$2,072,800	$1,566,295, $1,926,543[2]

FINAL COST PER CASE:	
1992	1987
$13.56	$12.79
	$15.73[2]

[1]Includes supervisory staff

[2]inflated by 23% to reflect 1993 dollars

1992 CENSUS OF AGRICULTURE CATI SYSTEM

TELEPHONE CENTER DEBRIEFING HANDOUT

Name:_____

Position: (check all that apply)

- ☐ Interviewer
- ☐ Research Operator
- ☐ Supervisor
- ☐ Directory Assistance

Date started working on Ag CATI:

_____

AGENDA

I. Introduction

II. Debriefing on Training

III. Debriefing on the Interviewing Instrument

IV. Debriefing on Special Processing

V. Other

I. INTRODUCTION

II. DEBRIEFING ON TRAINING

Large Farm:

1. Did the training prepare you for conducting interviews?

2. Were AG concepts and definitions sufficiently covered during training?

3. Were the reference materials you received sufficient? Were they necessary? Did you use them?

4. Were the practice interviews helpful? Were there enough?

5. Do you have any ideas for improving training in the future?

6. Is there anything you would like to see covered in the training that was not included?

7. Were there topics in training that needed more or less time? If so, what were they?

Nonresponse:

1. Did you think the self-study was adequate for nonresponse training? If not, what other types of training would you suggest?

2. Did you feel that the nonresponse self-study prepared you for conducting nonresponse survey interviews adequately? Did you need more or less time?

3. Did the 4 practice interviews provide enough practice?

Low Response:

1. Was the purpose of the Low Response County follow-up surveys made clear to you?

2. Was the introductory briefing of this survey sufficient for your interviewing? If not, what other types of training/information would be helpful?

III. DEBRIEFING ON THE INTERVIEWING INSTRUMENT

Front of Instrument:

1. In the >review< screen, information about the operation (1987 acreage, value of sales, type of organization, and so on) was available for the interviewer.

   Did you find this information useful?

2. Did you find that the "front" screens accommodated for most situations?

3. Were the screens for "claims filed" or "multiforms" situations sufficient?

Large Farm/Low Response Middle of Instrument:

1. Are there any general screen changes that you would recommend for the next census' instrument?

2. Included in the instrument were "menu" screens in which selection was made from a listing.

   Were these screens difficult to collect responses? Any problems with these screens?

3. There were screens for verification of responses.

   Were there any problems with using these screens?

4. There were screens with indication of prior responses.

   Were these hints helpful?

5. For some "menu" screens, the "other" option was available for selection 3 times.

   Was this sufficient? Did you need more than 3 "other" selections at times?

6. Many of the basic screens presented a question followed by an "If Necessary" statement which offered more detail. These were to be used in case the respondent needed more explanation of the basic question.

   Were these "If Necessary" statements helpful? Did you need to read these often, sometimes, or rarely?

7. Were there any differences between the Low Response and Large Farm surveys? (Differences such as, amount of respondent cooperation, kinds of respondent reactions, wording in the instrument that did not apply, unique problems, etc.)

8. There were screens in which more than one response was requested.

   Are these types of screens preferable to asking for a single item per screen?

9. There were screens available for "help or further explanation."

   Did you use these screens often, sometimes, or rarely?

10. The "F7" key was available for additional note taking. Did you use this often, sometimes, or rarely?

11. Did you find the "Shift F2" Q & A function helpful? Did you use this often, sometimes, or rarely? Were other Q & A needed? If so, give some examples.

12. Did you find it necessary to use the "F1" key to back up often? Did this function work well when needed?

Nonresponse Middle of Instrument:

1. Were the skip patterns logical? If not, what would you suggest?

2. Was the instrument adequate in terms of collecting all pertinent informatin to areas covered?

3. Were there any specific types of agricultural production for which the instrument did not adequately provide questions and/or answers?

4. Were there any questions which were not clear to the respondents? What questions were they, and how could the question(s) be reworded?

5. Was the instrument sufficient in questioning "small" and/or borderline farms? If not, explain.

Back of Instrument:

1. Were the callback screens sufficient for making a callback?

2. Was the >inotes< screen sufficient?

3. Was there any confusion as to whether a callback was a "soft" or "hard" appointment? If so, explain.

IV. DEBRIEFING ON SPECIAL PROCESSING

1. Did you find it difficult to connect with a research operator? Was it often, seldom, or rarely that a research line was busy or unavailable?

2. Were the roles of the respondent, researcher, and interviewer clear whenever research was being conducted?

3. Was the procedure clear as to what was to be done if a fax number instead of a telephone number was identified?

4. Did you find it helpful to have a calculator for the data collection process?

5. Were the periodic "Briefing Notes" informative?

V. OTHER

QUESTIONS FOR SUPERVISORS

- Were you given adequate training for your job as supervisor? If no, what type of training would have been helpful?

- Do you have any other comments in regards to your supervisory function?

QUESTIONS FOR RESEARCH OPERATORS

- Were you given adequate training for your job as research operator? If no, what type of training would have been helpful?

- What was the most serious problem encountered?

- Do you have any suggestions for improving the computerized research operation or the research process?

- Do you have any other comments in regards to your research operator function?

QUESTIONS FOR DIRECTORY ASSISTANCE CALLERS

- Do you have any suggestions for improving your job in calling DA for telephone numbers? If yes, what are they?

- Do have any comments in regards to your job in calling DA for telephone numbers? If yes, what are they?

TABLE OF CONTENTS

		Page
I.	Introduction	1
II.	Background	1
III.	Overview of the Agriculture CATI System	2
	A. Agriculture Division's Use of CATI	2
	B. System Development	3
	C. CATI Processing	6
	D. CATI Output/Results	11
IV.	Summary of Agriculture CATI System Results	13
	A. Final Case Accounting	13
	B. Cost Analysis	14
V.	Evaluation of the Agriculture CATI System	14
	A. Feedback from CATI Committee/Site Staff	15
	B. Evaluation Documentation	16
	C. CASIC Presentation	16
VI.	Suggestions for Future Agriculture CATI Systems	16
	A. Developmental Recommendations	17
	B. Processing Recommendations	17
	C. Systems Recommendations	17
VII.	Conclusion	18
	References	19
	List of Attachments	20

DISCUSSION: WHAT CAN CAI LEARN FROM HCI?

Mick P. Couper
Joint Program in Survey Methodology

1. Introduction

This discussion will not focus on the papers presented in this session. These are two good papers that demonstrate the feasibility of computer-assisted interviewing (CAI) for large-scale data collection, both CAPI (computer-assisted personal interviewing) and CATI (computer-assisted telephone interviewing). Instead, this discussion will take to heart the "New Directions" part of the title of this seminar, and attempt to take a glimpse at the future of CAI or, more broadly, CASIC. In doing so, I will focus on certain aspects of computer-assisted interviewing, and particularly the question of the user interface.

Sixty years ago, writing about the industrial revolution, Lewis Mumford (1934: 6) wrote: "So far we have embraced the machine without fully understanding it...". I believe the same may be said of CAI at its present stage of development. We know that CAI "works", as these two papers ably demonstrate, and we are enthusiastically advocating the application of computer technology to virtually all areas of survey data collection. The CASIC movement is well-established in government, academic and private survey research organizations. But how much do we understand this new method of data collection, and its impact on the data collected, on the people who provide it, and on the people who collect and process it?

To quote further from Mumford, "In order to reconquer the machine and subdue it to human purposes, one must first understand it and assimilate it" (1934: 6). This is a view of the machine as a tool in the hands of the user. Rather than making the human conform to the machine, attention should be turned to the needs of the user. There are many areas of research in CASIC that remain unexplored (see Couper, Groves and Kosary (1989), Groves and Nicholls (1986) and Baker (1992) for some examples), and much work that needs to be done to optimize the use of such systems. I will focus on only one of these areas, namely the question of usability.

Whereas feasibility addresses the question "Can it be done?", usability focuses rather on "How best can it be done?". "Best" in this case should be defined in part from the perspective of the users of the system. It is in this area that I believe we have a great deal to learn from the field of human computer interaction research or HCI. Marchionini and Sibert (1991) define HCI research as being "concerned with the design of interfaces that allow easy and efficient use of computer systems." Hix and Hartson (1993) offer a less formal definition of usability: "If your computer were

a person, how long 'til you punch it in the nose?".

The Dippo et al. paper in this session talks at length about the "new methodology" of using cognitive psychological theories and methods in questionnaire development. A similar opportunity presents itself with the application of HCI research methods and findings to CAI. There is much we can learn about usability, both in terms of findings from existing research in other domains, and in terms of methods for usability testing and evaluation of user interfaces.

By usability is meant simply that the focus of our attention turns from the system to the user. This means person-centered design rather than system-centered design. At present, the capabilities and limitations of the hardware and software we use are driving the design of CAI systems. We are making the user adapt to idiosyncracies of the system rather than the other way around. The notion of "user-friendliness" or the subjective reaction of the user to the system, is only one component of usability. Shneiderman (1992: 18) defines usability in terms of the following five measurable components:

(a) time to learn
(b) speed of performance
(c) rate of errors by users
(d) retention over time
(e) subjective satisfaction

In the spirit of continuous quality improvement, if we focus on measurable aspects of usability, we will be able to demonstrate concrete improvements in the design, development and implementation of computer-assisted survey instruments. In doing so, we should achieve measurable gains in data quality, defined by Kalton (keynote address, this conference) to include not only accuracy, but also timeliness, cost effectiveness, relevance and accessibility.

2. Types of Users of CAI Systems

In promoting a user-centered view in CAI, we need to define who the users are. I have four sets of people in mind: (a) the programmer or instrument designer, (b) the interviewer, (c) the manager or supervisor, and (d) the end-user or analyst. To this list could be added a fifth set of users, the respondents[1]. Each of these users faces a different set of usability and other issues. Some of these concerns, and areas for further research and development, are as follows:

---

[1]I thank Judith Lessler for reminding me of this group of users.

2.1. Programmer/author/instrument designer

Part of the ambivalence about what to call the people who create the CAI instrument reflects uncertainty regarding the combination of people and skills needs for this work. Should non-programmers be able to create and test a CAI instrument? What is the role of questionnaire designers versus computer programmers in instrument design? What do we mean by "programmers"? Developing an understanding of the process of instrument design will facilitate identification of the optimal skill combinations required to enhance the product of the CAI design process. Some exploratory work has been done in CAI on the development of new software tools or application of existing tools to facilitate the instrument development process (see for example Baker, 1988; Balestrino, Fortunato and Montagna, 1992; Dibbs and Hale, 1993; Pierzchala, 1993). Another area that needs further development is in the tools for testing and debugging CAI instruments (see Connett, Mockovak and Uglow, 1994).

In part, the mix of skills required for these tasks may be dictated by the design of the CAI systems used. Some CAI systems may use more natural language interfaces, while others use cryptic code in authoring specifications. Should we be expecting this group of users to adapt to the complexities of the systems being used? A human factors perspective would argue not. The first generation CAI software was relatively unsophisticated in its design interface, but we should expect more of future CAI systems.

2.2. Interviewer

In some senses this is the most critical group of users for a number of reasons. The large number of interviewers and the diversity of their computer skills and knowledge relative to other groups of users, the minimal training they receive on the computer hardware and software they will use with little close supervision, and the potential they have to impact the data collection process (in terms of both costs and errors), all make it imperative to design systems to maximize interviewer efficiency and minimize errors. We will return to this group later.

2.3. Manager and/or supervisor

This set of users requires detailed information on the process of data collection, including cost and production data. While CAI can provide vast amounts of timely information that were previously not available in paper-and-pencil data collection, we have yet to devise methods to manage the information flow in ways that would facilitate the work of these users. This group has the potential for information overload unless such tools are provided.

2.4. End-user, analyst

The production of analytic data sets in a form that analysts can readily use, and in a way that they can understand how the data were collected is an important area of usability. This includes both the data itself and the metadata (codebook, variable labels, information on skips and edits, etc.). Is the analyst provided with a hard copy version of the questionnaire to review, or are all users expected to load the interviewing software on their system to look at a particular set of questions they may wish to analyze? Some of the information that may be needed by this group of users includes: (a) where did this question appear (what questions came before or after)? (b) which respondents were asked this question (skip patterns)? (c) what edit, range and consistency checks were built into this question? and (d) how was this variable created (recode, combination of multiple questions, etc.)?

These needs speak to the integration of the survey data collection process with the production of useful data sets. This view acknowledges that CAI systems are more than just a set of interviewing tools, and are (or should be) a fully integrated system of data collection, management and data preparation (see Kreighton, Matchett, and Landman, 1994).

2.5. Respondent

In interviewer-administered surveys respondents may have little direct contact with the computer, other than through the interviewer. However, in a variety of self-administered surveys using CAI (such as CASI, CSAQ, TDE, VRE, etc.; see Weeks (1992) for a review), respondents may interface with the system directly. These respondents may have had little or no training on the use of the system, may have limited prior experience with computers, and may not be highly motivated to participate in the survey. Thus, in addition to concerns about interacting with the computer, they may be uncertain about the nature of the interview task itself. For these reasons, the design of the interface is especially critical for this group of users.

All of these sets of users (and there are others) vary in their information needs, the tools they need to access or use this information, their computer skill levels and/or knowledge of the particular system being used, and so on. Much of our energy seems to have gone into the task of getting working CAI instruments up and running. We have expected the various groups of users to adapt to the idiosyncracies and shortcomings of the systems we currently have at our disposal. A more user-oriented approach would be to systematically determine what the needs of each set of users are and understand the nature of their work, then design systems that specifically meet those needs or that facilitate the completion of their tasks.

One of the critical lessons from HCI or human factors research is the importance of involving users in the design and evaluation process (see Galitz, 1993; Gould and Lewis, 1983; Norman, 1983). This cannot be emphasized enough. Users are a valuable resource that we have tended to ignore or pay only lip service to in CAI. We often proceed from the assumption that we know best, and design systems with little regard for those who will attempt to use them. As Powell (1990: 31) notes caustically, "Dumbo the elephant used his ears to fly. Use yours to listen to the users."

In the remainder of this discussion, I will focus my remarks on the second group of users, the interviewers. This is not because the other users are less important, but rather that some of the problems they face may be relatively intractable in the short run, whereas measurable improvements can be made to the instruments used by interviewers will relatively little investment.

3. Design Principles for CAI

Thus far, I have talked in the abstract about the need to pay attention to the human-computer interface, and of the importance of designing for usability. Note that usability is more than simply screen design, it is the entire system as experienced by the users. As Jagodzinski (cited in Davis and Bostrom, 1992) notes, to most users (and this would certainly include interviewers), the interface <u>is</u> the system. Usability considerations cannot be separated from other aspects of system design and development (see Gould and Lewis, 1983; Gould, 1988).

Because usability or "user friendliness" can be a quite nebulous concept, let me offer a set of design guidelines for CAI systems. These are adapted from a variety of sources in the HCI literature, including Hix and Hartson (1993), Galitz (1993), Mayhew (1992), Norman (1983), Powell (1990), Ravden and Johnson (1989), and Shneiderman (1992). Some of these are empirically-based principles, others are more prescriptive. Nonetheless, these are a set of desirable qualities of computer systems generally agreed upon in the field of HCI that may be applicable to CAI. This list may serve as a starting point to focus our attention on some of the issues that need to be addressed in terms of enhancing the usability of the systems we use for CASIC. A well-designed CAI system should exhibit the following qualities:

4.1. Functionality

The system should meet the needs and requirements of users when carrying out the tasks (Ravden and Johnson, 1989). Note that this is functionality from a usability perspective. It is not what the designer thinks the users should do, but rather what the user needs to do in order to complete the task correctly and efficiently. Furthermore, it is not enough that the <u>system</u> can do

X; the critical question is whether the user can do X with the system.

Functionality is a necessary but not sufficient condition for any CAI system. Shneiderman (1992: 10) notes that if the functionality of a system is inadequate, it does not matter how well the human interface is designed. The remaining guidelines or principles are essentially ways in which system functionality can be presented to the user to facilitate successful completion of the task.

4.2. Consistency

This refers to the look and feel of the system. At its simplest level, consistency refers to the placement of items on screens, including the use of fonts, upper or lower case, color, highlighting, etc., to distinguish between questions, interviewer instructions, response options, and so on. However, consistency should also include input modes, mapping of function keys and movement and navigation around the instrument.

Consistency can be viewed at a number of different levels:

(a) Consistency within a particular survey instrument. There are probably few who would disagree with this in principle, but I have seen a number of production CAI instruments where this is not achieved in practice.

(b) Consistency between the instrument and other interviewer tools (case management, transmission software, e-mail, etc.). We give interviewers a variety of tools to use, often without taking much effort to integrate them in a consistent fashion. Do the function keys assigned to operations in case management, for example, have consistent effects when used in the survey instrument? Many of the case management systems used by survey organizations (see Nicholls and Kindel, 1993) are written in-house, usually with little consideration of the CAI interface with which they will be used.

(c) Consistency across different surveys instruments within a particular organization. This is an area where organizational standards or guidelines in the authoring of CAI instruments would be beneficial (see Hunter, 1993). It appears that many programmers or authors have a particular style, which may be internally consistent within the instrument (or module) they develop, but differs from other survey instruments interviewers have used. An extension of this is interviewers actually using different CAI systems for different surveys.

(d) Consistency across organizations. Although I am not advocating that this be done, we ought to acknowledge that interviewers may work for multiple organizations using different hardware and software systems. There are no universally accepted design guidelines for CAI systems, and this may impact on the transferability of knowledge.

The first of these levels may be the most easy to implement, but

the other levels are no less important to consider in CAI.

Another component of consistency is predictability. System actions should be expected within the context of actions that are performed by the user (Galitz, 1993). In other words, if the interviewer does X, the system should always do Y. Thus, there is not only consistency in terms of what the user sees and does, but also consistency in terms of what the system does in response to user inputs.

Although consistency is probably the most universally endorsed principle, there are those who caution against its rigid application without consideration of other design principles. Grudin (1989) shows examples of how blindly following the maxim of consistency to the exclusion of other interface considerations can lead to poor usability design decisions (see also Reason, 1990).

4.3. Informative feedback

For every user action there should be some system feedback (Shneiderman, 1992: 73). This may take the form of immediate execution, change in state or value, correction message, confirmation message or in-progress message (see Ravden and Johnson, 1989: 56). System feedback is especially critical when system time is slow. Such feedback should be clear, concise and intelligible to the user.

4.4. Transparency

The system should permit the user's attention to be focussed entirely on the task being performed, without concern for the mechanics of the system (see Galitz, 1993). The computer is ideally suited for automating routine functions, and these should not detract from those activities requiring human attention. In CAI these may include time stamps, range and consistency checks, read-write operations, and other system functions. The interviewer's focus should be on the interviewing task, rather than on the operation of the CAI system.

However, there may be times when it is necessary for the user to see what the system is doing. One example in CAI may concern skip patterns. Usually these would be transparent to the interviewer, but there may be times when s/he needs to make judgements about an appropriate response to a root question. Without knowing the logic of the skip and the outcome of a particular choice, the interviewer cannot make an informed judgement as to the appropriate response.

4.5. Explicitness

Whereas the actions that the system performs without human intervention should be transparent to the user, the steps that the

user needs to take should be obvious. Norman (1988) uses the notion of affordances in evaluating the design of everyday objects. Essentially, affordances are properties of objects that suggest what sort of operations and manipulations can be done. For example, the design of a door handle suggests the operation to be performed, by affording pulling, pushing, turning or twisting actions. Affordances can be similarly applied to the human-computer interface: the computer screen should make the required user actions explicit or self-evident. Norman (1983) also cautions against the "tyranny of the blank screen" in DOS, where the "c:\" prompt provides the user with no clue as to what operation need to be performed. Many CAI systems assume that the user knows what to do in a particular situation, sometimes without providing any hints as to the expected action or guidance on where to find help to complete the task. Well-designed systems should make both the semantics (what can be done) and the syntax (how to do it) of the system explicit (Mayhew, 1992).

4.6. Comprehensibility

Systems should be understandable to users. Jargon, idiosyncratic language and abbreviations should be avoided. Norman (1988: 179) suggests ways to violate this guideline: "Be arbitrary. Use nonobvious command names or actions. Use arbitrary mappings between the intended action and what must actually be done." Where possible, natural language and real-world analogies should be used (Hix and Hartson, 1993). Ravden and Johnson (1989: 32) note that "The way the system looks and works should be compatible with user conventions and expectations." For example, using the Page Up, Page Down, and arrow keys for movement may make more sense than using function keys. The layout of dates, telephone numbers, etc. in the CAI system should match users' expectations or common conventions for the presentation of such information.

4.7. Tolerance

The system should be tolerant of human capacity to make errors. Galitz (1993: 26) writes: "The fear of making a mistake and not being able to recover from it is a primary contributor to a fear of dealing with computers." System design should recognize that errors will be made, and should include appropriate error prevention, detection and correction facilities (see Reason, 1990). Efforts should be made to prevent serious errors while facilitating easy recovery from more common errors. The more potentially disastrous an action, the more difficult it should be to perform. Thus, backing up to change a previous answer in CAI should be easier to do than suspending an interview in midstream. Actions should be easily reversible.

4.8. Efficiency

The system should be designed to minimize effort and maximize

efficiency on the part of the user. System response time is only one aspect of efficiency. As Mayhew (1992: 508) notes, overall task time is a function of both system and user response time. In other words, task time = system response time + system display rate + user scan/read time + user think time + user response time + time making errors + time recovering from errors. By improving the speed of the system (without attending to interface issues to reduce user time or errors), only the system side of efficiency (response time and display rate) will be addressed, without affecting other components of overall task time. All aspects of good interface design should facilitate the overall efficiency of operation. For example, user response time can be optimized by avoiding complex sequences of actions for common operations. User scan/read time can be reduced through effective screen design. Galitz (1993: 48) notes that system responsiveness should match the speed and flow of human thought processes, and offers some specific guidelines for various types of operations (see also Shneiderman, 1992: 284-297).

4.9. Supportiveness

This is closely related to the principles of explicitness and comprehensibility. Tolerance of errors and facilities for easy recovery from errors is another characteristic of a supportive system. The limited cognitive capacities of users should be recognized and accommodated. This can be done by reducing the amount of memorization of commands, codes, syntax and rules required by users (Brown, 1988: 97). Reliance on recognition rather than recall will help reduce cognitive burden for the user. Norman (1991: 6) writes, "It is typically the case that for systems with 40 plus commands, only about 7 commands show any frequency of use". Complex sets of commands and those that are rarely used are less likely to be remembered. Supportive systems provide online help and make it readily accessible to the user. If one needs to consult a manual to find out how to get online help, something is gravely wrong with the system.

4.10. Optimal Complexity

The early dictums on design (on both screen and paper) called for keeping things simple and maximizing the use of blank space. This view has given way to a recognition from a growing body of research (see Tullis, 1983; Coll and Wingertsman, 1990; Staggers, 1993) that users' preference for complexity exhibits an inverted U shape. Users both prefer and perform better with a moderate amount of complexity, rather than too simple or too complex. Galitz (1993: 35) notes that complexity should be commensurate with the capabilities of the system users. Complexity refers not only to the amount or density of information of the screen, but to all aspects of screen design. Hix and Hartson (1993: 49) recommend organizing the screen to manage complexity.

5. Lessons to be Learned from HCI

These general design guidelines for CAI systems should be seen as a set of goals for improving CAI design, rather than principles set in stone. There may be other characteristics of effective systems not mentioned here. There is also a recognition that compromises among these qualities or guidelines may be necessary to achieve optimal usability of CAI systems. Nonetheless, I have found examples of violations of each of the guidelines in various production instruments used by a number of different organizations. The CAI systems we use are clearly not perfect, and there is much room for improvement. These guidelines may help us on the road to quality improvement in instrument design to facilitate the work of our interviewers.

Many of these guidelines are not new to CAI. A number of these principles have already been articulated with regard to CAI. Nicholls and House (1987), for example, note that one of the general objectives of CAI systems is that they should meet interviewer needs. They explicate further: " displays must be quickly comprehensible [Principle 6], interviewers should have access to all needed information [Principle 9], opportunities for interviewer error should be minimized [Principle 7], and interviewer movement through the questionnaire, either forward or backward, should be expedited [Principle 8]" (Nicholls and House, 1987: 96). Despite these and other efforts to articulate design guidelines for CAI systems, it appears that little progress has been made.

If the only contribution made by human-computer interaction research to CAI was in the development of a set of general design principles (such as those outlined above), we would not have gained much. There are two additional keys to the applicability of HCI research to CAI. The first is a theoretically-grounded understanding of the interaction between human and computer and how the interface impacts the user and his/her task. Human factors research traces much of its theoretical roots to cognitive psychology (see for example Carroll, 1991), and it is this body of literature that will be most helpful to CAI design. The second critical lesson to be learned from HCI is the application of research methods to measure and understand the usability aspects of CAI systems (and user interfaces in general). The utility of HCI research lies not only in <u>what</u> was found, but also <u>how</u> it was found. A variety of methods are used in HCI research that can be readily adapted for use in CAI. These include usability testing in laboratory settings, experimental studies, observation and so on (see Shneiderman, 1992). Both theory and measurement are important to the partnership between HCI and CAI.

With regard to theory it is important to note that not all the findings of HCI research are equally applicable to computer-assisted interviewing. The nature of the interviewing task may be

very different from that studied in HCI (use of programming languages, word processors, spreadsheets, etc.). The partnership with HCI does not mean uncritically applying all findings in that field to CAI. Rather, by focusing on similarities and differences between CAI and other tasks involving human-computer interaction, we can distinguish between what is useful and what requires further exploration.

As far as measurement is concerned, many of the techniques already used in survey research can be applied to study interaction aspects of the task. These include:

(a) Cognitive laboratory investigations of interviewer-computer interaction using observation, protocol analysis, think alouds, etc.
(b) Laboratory-based experiments testing alternative designs or focusing on particular issues and actions interviewers face in CAI.
(c) Scripted mock interviews which may include tests of particular types of actions.
(d) Observation of production interviewing using computers.
(e) Experiments embedded in production data collection.
(f) Measurement of interviewer production and process (e.g. keystroke files, time stamps, monitoring, behavior coding).

Many of these methods parallel those used to study human-computer interaction, and can be productively applied to CAI.

6. Conclusion

So where do we go from here? In this discussion I have tried not to be too prescriptive in terms of ways to design user interfaces for CAI. Rather, I am advocating more of a design philosophy that (a) explicitly takes the users into account; (b) involves measurement of progress toward usability goals (e.g reducing learning, minimizing errors, maximizing user satisfaction, etc.); and (c) attempts to extract empirically-derived principles and guidelines that have general applicability beyond the particular system or interface on which they were tested or developed. These tasks can be greatly facilitated by learning from the field of HCI research.

In terms of action steps, I believe the field of computer-assisted interviewing can make great strides by doing the following:

(a) Apply what is already known about human-computer interaction and usability to CAI.
(b) Adapt HCI research methods to understand and explore usability issues in CAI. Conduct both qualitative and quantitative research on the interface between interviewer and computer in CAI.
(c) Explicitly incorporate usability testing as an integral part of the instrument development process.

(d) Identify gaps in our knowledge in the human-computer interface in CAI, and undertake research to close these gaps.
(e) Think ahead to new technologies and what we require of them, rather than being constrained by the limitations of existing CAI systems.

In this discussion, I have tried to turn our attention to the future rather than the immediate past. I see these issues both as challenges and as great opportunities for survey research. Let us not just embrace the machine, let us understand it and thereby unleash its full potential.

In doing so, we can learn a great deal from HCI research. In the same way that the field of questionnaire design has reaped great benefits from the partnership with cognitive psychology, so too can CAI benefit from interaction and collaboration with human-computer interaction or human factors researchers. Indeed, the benefit may well be mutual.

REFERENCES

Baker, R.P. (1992), "New technology in survey research: Computer-assisted personal interviewing (CAPI)." Social Science Computer Review, 10 (2): 145-157.

Balestrino, R., Fortunato, E. and Montagna, S. (1992), "A survey questionnaire design methodology." Pp. 489-501 in Westlake, A., Banks, R., Payne, C. and Orchard, T., Survey and Statistical Computing. New York: North-Holland.

Bayerl, J.P., Millen, D.R. and Lewis, S.H. (1988), "Consistent layout of function keys and screen labels speeds user responses." Proceedings of the Human Factors Society 32nd Annual Meeting, pp. 344-346.

Brown, C.M. (1988), Human-Computer Interface Design Guidelines. Norwood, NJ: Ablex.

Carroll, J.M. (ed.) (1991), Designing Interaction: Psychology at the Human-Computer Interface. New York: Cambridge University Press.

Coll, R. and Wingertsman, J.C. (1992), "The effect of screen complexity on user preference and performance." International Journal of Human-Computer Interaction, 2 (3): 255-265.

Connett, W.E., Mockovak, W.P. and Uglow, D.A. (1994), "Future CAI needs from a user perspective." Paper presented at the Census Bureau's 1994 Annual Research Conference and CASIC Technologies Interchange.

Couper, M.P., Groves, R.M. and Kosary, C. (1991), "Methodological issues in CAPI." Proceedings of the American Statistical Association, Survey Research Methods Section.

Davis, S. and Bostrom, R. (1992), "An experimental investigation of the roles of the computer interface and individual characteristics in the learning of computer systems." International Journal of Human-Computer Interaction, 4 (2): 143-172.

Dibbs, R. and Hale, A. (1993), "Questionnaire design in a paperless society." Proceedings of the Bureau of the Census 1993 Annual Research Conference. Washington: Bureau of the Census.

Falzon, P. (ed.) (1990), Cognitive Ergonomics: Understanding, Learning and Designing Human-Computer Interaction. San Diego: Academic Press.

Galitz, W.O. (1993), User-Interface Screen Design. Boston: QED.

Gould, J.D. (1988), "Designing for usability: The next iteration is to reduce organizational barriers." Proceedings of the Human Factors Society 32nd Annual Meeting, pp. 1-9.

Gould, J.D. and Lewis, C. (1983), "Designing for usability -- key principles and what designers think." Proceedings of CHI '83: Human Factors in Computing Systems. New York: ACM, pp. 50-53.

Grabinger, R.S. (1989), "Screen layout design: Research into the overall appearance of the screen." Computers in Human Behavior, 5: 175-183.

Groves, R.M. and Nicholls, W.L. (1986), "The status of computer-assisted telephone interviewing: Part II -- Data quality issues." Journal of Official Statistics, 2 (2): 117-134.

Grudin, J. (1989), "The case against user interface consistency." Communications of the ACM, 32 (10): 1164-1173.

Hix, D. and Hartson, H.R. (1993), Developing User Interfaces: Ensuring Usability Through Product and Process. New York: Wiley.

Hunter, P. (1993), "Introducing CAI standards within OPCS." Paper presented at the Second International Blaise Users Conference.

Kreighton, K., Matchett, S. and Landman, C. (1994), "Building integrated systems of CASIC technologies at the U.SA. Bureau of the Census." Paper presented at the Census Bureau's 1994 Annual Research Conference and CASIC Technologies Interchange.

Marchionini, G. and Sibert, J. (1991), "An agenda for human-computer interaction: Science and engineering serving human needs; Report of an invitational workshop sponsored by the National

Science Foundation." SIGCHI Bulletin, 23 (4): 17-32.

Mayhew, D.J. (1992), Principles and Guidelines in Software User Interface Design. Englewood Cliffs, NJ: Prentice-Hall.

Mumford, L. (1934), Technics and Civilization. New York: Harcourt, Brace and Company.

Nicholls II, W.L. and House, C.C. (1987), "Designing questionnaires for computer-assisted interviewing: A focus on program correctness." Proceedings of the Bureau of the Census Annual Research Conference. Washington, DC: Bureau of the Census.

Nicholls II, W.L. and Kindel, K.K. (1993), "Case management and communications for computer assisted personal interviewing." Journal of Official Statistics, 9 (3): 623-640.

Nielsen, J. and Levy, J. (1994), "Measuring usability: Preference vs. performance." Communications of the ACM, 37 (4): 66-75.

Norman, D.A. (1983), "Design principles for human-computer interfaces." Proceedings of CHI '83: Human Factors in Computing Systems. New York: ACM, pp. 1-10.

Norman, D.A. (1988), The Design of Everyday Things. New York: Doubleday.

Norman, D.A. (1993), Things that Make Us Smart: Defending Human Attributes in the Age of the Machine. Reading, MA: Addison-Wesley.

Norman, K.L. (1991), The Psychology of Menu Selection: Designing Cognitive Control of the Human/Computer Interface. Norwood, NJ: Ablex.

Pierzchala, M. (1993), "Computer generation of mega-version instruments for data collection and interactive editing of survey data." Proceedings of the Bureau of the Census Annual Research Conference. Washington, DC: Bureau of the Census.

Powell, J.E. (1990), Designing User Interfaces. San Marcos, CA: Microtrends Books.

Ravden, S. and Johnson, G. (1989), Evaluating Usability of Human-Computer Interfaces. Chichester: Ellis Horwood.

Reason, J. (1990), Human Error. Cambridge: Cambridge University Press.

Shneiderman, B. (1992), Designing the User Interface: Strategies for Effective Human-Computer Interaction. (2nd. ed.) Reading, MA: Addison-Wesley.

Staggers, N. (1993), "Impact of screen density on clinical nurses' computer task performance and subjective screen satisfaction." International Journal of Man-Machine Studies, 39 (5): 775-792.

Tullis, T.S. (1983), "The formatting of alphanumeric displays: A review and analysis." Human Factors, 25 (6): 657-682.

Weeks, M.F. (1992), "Computer-assisted survey information collection: a review of CASIC methods and their implications for survey operations." Journal of Official Statistics, 8 (4): 445-465.

DISCUSSION OF TWO PAPERS ABOUT CASIC

Sandra Sperry
Westat, Inc.

This is such an interesting time in survey research. After many years in which there was very little change in the approaches used to collect data and prepare it for analysis, the past ten years have produced enormous change. First there was the growth of centralized telephone interviewing using computer-assisted data collection methods. Then in the late 80's, government agencies began to experiment with using laptop computers to collect data for in-person interviews. Now it seems to be completely accepted that an agency will, at the very least, consider computer-assisted survey information collection (CASIC) techniques for all its surveys.

The impact of these two papers about CASIC comes in part from the fact that the Census Bureau and the Bureau of Labor Statistics do things on such a grand scale. How often do survey researchers get to casually mention that their pretest of a CATI system included 10,000 cases? This is what Jeanette Mon tells us about the CATI test done in 1982 to prepare for the 1992 Census of Agriculture. And Cathy Dippo tells us that the Parallel Survey, which was carried out for 18 months to assess the overall effects of redesigning the Current Population Survey (CPS), was "relatively small," that is, only 12,000 households per month.

The scale of these two projects, along with their emergence at a time when so many organizations are considering how best to implement CASIC, makes these two papers especially valuable. The papers are, among other things, historical records that document aspects of the development and application of CASIC at a time when CASIC is moving from being unusual and experimental to being the data collection mode of choice.

Both papers tell us a lot about process, although each paper focuses on different aspects of the process of applying CASIC methods. Cathy Dippo's paper tells us about the methodological processes used to redesign the CPS questionnaire for computer-assisted data collection. Jeanette Mon's paper focuses more on the processes used to manage all aspects of developing and using CASIC methods. Both include much helpful information for the researcher embarking on a project that will use CASIC. I would like to mention a few of the points that I found most relevant.

I will start with Cathy Dippo's paper. She notes that preliminary work on designing a new CPS questionnaire included laboratory work on how interviewers and respondents understand labor force terminology. This work led to developing questionnaires that were also tested with laboratory techniques. Then special respondent debriefing techniques were used with 2,300 CPS telephone respondents. The results of this debriefing verified the laboratory findings. These results are reassuring about the value of using laboratory techniques for questionnaire design. Most researchers do not have the resources to carry out a respondent debriefing of this scope and, therefore, are not able to evaluate the results of their laboratory work with this precision.

I was also very interested in how the CPS staff used CASIC techniques to help them with questionnaire design. The use of behavior coding while monitoring CATI interviews is an excellent way to maximize the benefits of a CATI or CAPI pretest. Programming follow-up probes to debrief respondents about specific answers that they gave during the main interview seems like another inspired idea for getting the most out of a pretest.

The design features of a CASIC questionnaire that have been incorporated into the new CPS instruments, that is, complex skips, wording that is specially tailored to the respondent's situation, built-in consistency checks, and dependent interviewing in longitudinal data collections, are aspects of CASIC that I think have been incorporated into a number of CATI and CAPI surveys. It was interesting to read about some of the ways in which the CPS instrument has changed to make use of CASIC, but the concepts presented in this part of the paper were not as new to me as some of the approaches used for instrument development. To learn even more about the instrument development process, I think many researchers would be interested in reading some of the technical reports written by the CPS Overlap Analysis Team. The technical reports present more detail about work that was done on this project, a project that included much more developmental research than most projects can afford.

I would like to turn now to Jeanette Mon's paper. As I mentioned earlier, this paper addresses more of the issues encountered in managing a large CASIC design and data collection. I was particularly struck by the number of different Census divisions (five) that needed to cooperate and communicate with one another in order to carry out the 1992 Census of Agriculture's CATI follow-up. Because different divisions had different responsibilities with regard to the same data, it was very important for them to meet regularly and to document carefully everything that they did. It is often easy for researchers working on smaller studies or in organizations with a less structured approach to dividing up responsibilities to assume that they do not need the level of formal communication and documentation that was required for the Census of Agriculture follow-up. I would hypothesize, however, that all survey research projects would be better off if they included more preparation of the formal specifications that Ms. Mon mentions in her paper.

Ms. Mon also mentions that to develop the basic system specifications that are required by the CASIC programming staff, the subject matter experts needed to go through a learning period. She notes that in some instances, the requisite specs were so complicated that it took a considerable amount of time to learn how to prepare them. I think this is a problem that many of us have encountered in developing CASIC instrumentation. The process requires much collaborative work among staff involved in all different aspects of a CASIC project, and we must occasionally spend time just learning how to communicate with one another.

The description of how different states were scheduled to be called in each of two telephone centers presents another lesson in managing a large scale CATI operation. It was very important that the scheduling of states be kept flexible so that when one or the other of the phone centers began to run low on work at particular times (as happened on two different occasions), new states could be installed to provide the level of work needed to keep interviewers working efficiently.

The paper about the Census of Agriculture's CATI follow-up included one more section that I found particularly enticing. The CATI software created a file with a complete history of all actions made with regard to a case. Staff used these files "to produce tables and graphs for management analysis." Some of the tables and graphs are included as attachments to the paper. I found these tables and graphs very intriguing and would like to know more about how the management staff used them during the field operations phase of the project.

In conclusion, I will not suggest the usual call for more research. Rather, I will end with a call for more papers from the research on which these two papers are based. There is a great wealth of material here to be explored, and I think there is much for all of us to learn from it.

Session 8

LONGITUDINAL SURVEYS

PANEL DESIGN AND ESTIMATION STRATEGIES IN THE NATIONAL MEDICAL EXPENDITURE SURVEY

Steven B. Cohen

Agency for Health Care Policy and Research
2101 E. Jefferson Street
Rockville, Maryland 20852

Presented at the COPAFS Seminar on New Directions in Statistical Methodology
May 25-26, Bethesda Holiday Inn

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 acknowledge the contributions made by the following individuals who were co-authors of papers whose findings have been synthesized in this presentation: D.E.B, Potter, Barbara L. Carlson, and Ayah Johnson.

PANEL DESIGN AND ESTIMATION STRATEGIES IN THE NATIONAL MEDICAL EXPENDITURE SURVEY

Steven B. Cohen, Agency for Health Care Policy and Research

1. Introduction

The National Medical Expenditure Survey (NMES-2) was established to provide an assessment of the health care utilization, expenditures, sources of payment and health insurance coverage of the U.S. civilian noninstitutional population and the population using nursing and personal care homes and facilities for the mentally retarded for calendar year 1987. The core of the data collection effort for the non-institutionalized population was a series of interviews with a household sample that collected detailed information on health status, use of health care services, expenditures and sources of payments, insurance coverage, employment, income and assets, and demographic characteristics for calendar year 1987. The institutional population component was designed to obtain similar types of information for the institutionalized population residing in nursing and personal care homes (NH), and in facilities for the mentally retarded (MR) for calendar year 1987.

The NMES household survey was designed as a panel survey with an initial screening interview conducted in the fall of 1986 for a sample of approximately 35,600 addresses, to obtain information required for oversampling of specific policy relevant population subgroups. A subsample of about 15,000 households was selected for the detailed interviews. To reduce the deleterious impact of long recall periods on measurement error, data collection specifications required four separate interviews conducted with selected households at three to four month

intervals over a fifteen month period to obtain the required heath care data for calendar year 1987. The adoption of a panel design was also motivated by the analytical goal of measuring changes in health insurance coverage over the course of calendar year 1987.

The NMES Institutional Population Component was designed to yield unbiased national and regional estimates at the facility level and for the overall institutional user population. More specifically, the primary objective of the survey was to estimate the use of and expenses for health care services for all persons residing in institutions at any time during calendar year 1987. To obtain a nationally representative sample of the 1987 institutional user population, the survey included a sample of residents residing in selected facilities as of January 1, 1987, in addition to a representative sample of admissions to the selected facilities over the course of 1987. The union of these samples served to represent the 1987 institutional user population. The operational implications of a selection of both residents as of 1/1/87 and admissions over the course of 1987 required the adoption of a panel design for the survey. Interviewers made four distinct visits to each cooperating facility at approximately four month intervals to facilitate sample selection and data collection in the institutions.

This paper provides a summary of the sample design and estimation strategies adopted in the National Medical Expenditure Survey associated with the longitudinal features of the survey. For the Household Survey, particular attention is given to the consequences of adopting an address sample design in meeting specified survey design goals. The NMES Household survey also benefitted from a design strategy that included the selection of a sample of

households in the first round of data collection for the main study that were non-respondents to the screener interview. The impact of this design strategy on the overall survey response rate is also summarized. The paper includes a discussion of the estimation strategy adopted to adjusted for part-year nonresponse. With respect to the Institutional Population Component, both the sample design and the estimation strategy used to correct for the representation of individuals with multiple opportunities for sample selection over the course of 1987 are described in detail.

2. Address Sample Design in the NMES Household Survey

The Household Component of the National Medical Expenditure Survey (NMES) is a panel survey with 1987 as the reference period, collecting measures of health status, use of health care services, expenditures and sources of payment, insurance coverage, employment, income and assets, as well as demographic information for the U. S. civilian noninstitutional population. To meet analytical objectives, the survey included an oversample of the following policy relevant population subgroups: blacks, Hispanics, the poor and near poor, the elderly and persons with functional limitations. A separate screening interview was conducted in the fall of 1986 to facilitate the identification of these population subgroups. The screened households were selected for the main NMES household survey on the basis of the characteristics of the persons they included at the time of the screening interview. However, for the purposes of cost-efficiency and to maximize the response rate, the actual NMES Round One sample was characterized by an address sample design. An address sample design requires the interviewer to go back to the originally sampled address, whether the same household resides there or not. Consequently, the Round One sample consisted of all households living at the sampled addresses

at the time of the Round One interview, whether or not the screened households were still present.

The sample design considerations inherent in the selection of an address sample as an alternative to a household sample design are clearly not limited to the NMES, but generally applicable to all national household surveys that consider non-concurrent screening interviews to identify particular population subgroups targeted for oversampling. When the basic sampling units are defined to be the sample addresses themselves, rather than the households residing at the addresses at the time of the screening interview, a reduction in survey costs should be realized for the subsequent interview since movers would not have to be traced and interviewed at their new location. However, when the address sample design is used, each sample listing (or a representative subsample) should be returned to in the round of data collection following the screening interview, to interview all eligible respondents residing at the address (Cohen and Johnson, 1992). Addresses that were vacant at the time of the screening interview would have to be checked to determine their occupancy status. Dwelling units whose occupants refused to participate in the screener interview would need to be recontacted in addition to those units in which no eligible respondents were at home or available for being interviewed for the screening interview (Cox and Cohen, 1985; Cox et al., 1979).

The adoption of a household sample design would define households residing within the sampled addresses at the time of the screening interview as the basic sampling units. Individuals or families who moved out of the sample dwelling units would have to be traced and followed for interview. This design imposes greater control over the sample to insure that sample size targets are satisfied for the oversampled population subgroups. Furthermore, this design would

generally yield more precise survey estimates for the oversampled population subgroups than an address sample design, which is more vulnerable to the adverse effects of greater sampling weight variation. However, the advantages of the household sample design are less likely to be realized as the time period between the screening and follow-up interviews increases. Additional interviewers would have to be hired to conduct the Round One interview for the selected screener respondents that moved outside the PSUs that comprised the NMES screener sample. Furthermore, the field period would have to be extended to accommodate the time needed to locate the movers and conduct their interview.

The choice of design requires an evaluation of the potential loss in the precision of surveys estimates for the oversampled population subgroups, further contrasted with the potential savings in survey costs and higher response rate achieved by the address sample design. In this study, the consequences of adopting an address sample design for the National Medical Expenditure Survey are evaluated in terms of the realization of specified survey design goals.

The adopted NMES household survey sample design is a stratified area probability design with the following stages of selection: (1) selection of 165 primary sampling units (PSU's) which are counties, parts of counties or groups of contiguous counties; (2) selection of 2,317 segments within PSU's; (3) selection and screening of dwelling units within segments; and (4) selection of dwelling units based on demographic characteristics from the set of screened dwelling units (Cohen, DiGaetano and Waksberg,1991). The survey was sponsored by the Agency for Health Care Policy and Research and the NMES sample represents a union of the national sample frames of Westat, Inc., the prime contractor, and NORC, a subcontractor.

The NMES screener interview was conducted in the fall of 1986. The final NMES screener sample consisted of 35,634 addresses, of which 3,091 were identified as vacant, another 1,085 identified as not a dwelling unit and 250 addresses were determined to be ineligible for the survey. Of the 31,208 remaining dwelling units sampled, 28,458 responded to the screener interview, achieving a 91.2 percent response rate.

The Round One household sample of dwelling units was then selected from the screened households, using sampling rates that achieved the desired sample for specified population subgroups (DiGaetano,1987). The 15,130 dwelling units that constituted the targeted Round One sample, based on the demographic and health status profiles of their members at the time of the NMES screener interview, initially consisted of 16,615 responding reporting units. Reporting units were defined as individuals related by blood, marriage or adoption within a dwelling unit. These reporting units contained 39,885 individuals that represented the civilian non-institutionalized population as of Fall, 1986.

Figure 1 illustrates the transitions in sample composition that would occur between the screener and Round One interviews, as a consequence of the address sample design. The expected transitions include the movement of targeted households out of sampled addresses, changes in household composition, and the inclusion of replacement households that have moved into sampled addresses.

FIGURE 1 ABOUT HERE

It was also expected that some of the addresses contacted which were vacant in the screener field period would become occupied at the time of the Round One interview. Excluding

them from the sample would understate the number of recent moves in the sample. Consequently, a sample of 1,464 vacant addresses was selected to supplement the occupied addresses sampled from all screened households.

o Field Results

Field results indicated that the majority of the 15,130 sampled addresses consisted of the same reporting units that completed the screener interview. Transitions occurred as a consequence of the following situations:

1. movement by screener identified families out of sampled dwelling units,
2. creation of new reporting units, consisting of individuals related by blood, marriage or adoption, that have moved into sampled addresses since the time of the screener interview, and
3. changes in the Round One composition of reporting units that were initially identified at the time of the screener interview.

At the end of the Round One field period, it was determined that 722 (4.5 percent) of the 15,130 sampled addresses responding to the NMES screener interview were vacant. Another 253 sampled addresses were determined to be ineligible for the interview as a consequence of the death or institutionalization of targeted respondents, or due to a change in the original household composition to all military or student members. Furthermore, 847 reporting units that responded to the screener moved out of the sampled addresses, and were replaced by new reporting units at the time of the Round One interview. Overall, more than ten percent of the sampled addresses with screener respondents experienced at least one household move during the five to six month period that passed between interviews..

Within the 15,130 sampled addresses that constituted the targeted Round One sample, there were 15,590 reporting units identified as eligible or NMES interviews during the Round One field period out of 17,412 reporting units linked to the addresses during the screener or Round One interview. Of these, 14,060 responded (90.2 percent) to the Round One interview. The remaining 1,530 eligible RUs were classified as nonrespondents (9.8 percent) due to a refusal to complete the interview, unavailability during the field period, illness, or other nonresponse. Relative to the 1,464 sample addresses vacant at the time of the screener, 1,016 (69.4 percent) remained vacant at the time of the Round One interview. Of the 479 reporting units new to the sample, 408 (85.2 percent) responded to the first household interview. The joint screener-round one response rate for these targeted Round One addresses, including the screener vacant sample, was 82.1 percent. This was derived by multiplying the screener response rate (.912) with the combined Round One response rate for targeted sample and the vacant sample ((14,060 + 408)/(15,590 + 479) = .90).

At the person level, it was noted that 32,205 (80.7 percent) of the targeted 39,885 screener respondents also responded to the Round One interview. Of the 7,680 screener only-respondents, 3,150 individuals (41 percent) were in targeted reporting units that refused to complete the Round One interview. Another 224 individuals (2.9 percent) resided in reporting units that were determined to be ineligible at the time of the Round One interview. Consequently, 43.9 percent of the screener only- respondents would not have participated in the Round One interview, independent of the address sample design.

It was also determined that 1,673 of the screener only- respondents (21.8 percent) departed from the sample as a consequence of the movement of their reporting units (847 RUs)

away from the sampled address, which gained a new replacement household available for the Round One interview. Another 1,529 individuals (19.9 percent) left the sample due to the movement of their reporting units out of the sampled address, which was vacant at the time of the Round One interview. The remaining 1,104 screener-only respondents that departed from their sampled addresses (14.4 percent), were associated with reporting units for which at least one respondent completed both the screener and Round One interview at the sampled address. Consequently, 56.1 percent of the screener only- respondents (4,306) were not sampled in Round One as a function of the address sample design. The final set of respondents to the first round of the NMES household survey consisted of 36,259 individuals. Of the 4,054 individuals that were Round One only-respondents, 841 (20.7 percent) were associated with initially sampled reporting units that did not respond to the NMES screener interview, but were refielded to improved the overall NMES response rate. Independent of an address sample design, a decision to refield these cases could have been incorporated in the NMES sample design. The remaining 3,213 Round One only-respondents (79.3 percent) were added to the NMES sample as a function of the address sample design.

o A COMPARISON OF THE DEMOGRAPHIC CHARACTERISTICS OF NMES MOVERS

Of the 7,680 screener only-respondents, 43.9 percent would not have completed a Round One interview, independent of the address sample design for NMES. Under an alternative sample design that tracked all screener respondents that moved, efforts would have been made to locate or determine the status of 4,306 individuals linked to more than 1,500 reporting units.

The demographic characteristics of these targeted sample movers are presented in Table 1, and contrasted with the characteristics of the 3,213 replacement individuals that are new to the sample as a function of the address sample design (excluding the 841 new Round One respondents associated with RUs that were screener nonrespondents).

TABLE 1 ABOUT HERE

According to Current Population Reports, (U.S. Bureau of the Census, 1987) 18 percent of the population experienced a move during the course of the year. The NMES national estimate of population transition (10 percent) during the five to six month period that transpired between the screener and Round One interviews, compares well with census figures.

Overall, the 3,213 individuals that were new to the sample in Round One represented a 25 percent shortfall in targeted sample size, relative to the movers. The population subgroup that consistently experienced the greatest proportionate differential from sample targets were the elderly. In addition, a greater shortfall in targeted sample size characterized the black and Hispanic population subgroups, relative to the domain that represented whites and other races. It was recognized, however, that even if the targeted sample of movers were to be traced, the expected Round One yield would need to reflect nonresponse and loss in sample due to death, institutionalization, or inability to locate.

Under the assumption of a ten percent loss in sample due to nonresponse (which was the experience for non-movers), and a conservative assumption of an additional five percent loss due to death, institutionalization, or inability to locate, the expected Round One yield for the targeted NMES movers was 85 percent. Relative to the expected number of movers completing the Round One interview, the effective overall shortfall in sample size was only 12 percent. When

focused on specific population subgroups, a more pronounced shortfall is sample size was noted, particularly for the elderly. The address sample design was more effective for subpopulation subgroups defined by race or ethnicity, where black or Hispanic households were likely to be replaced by other black or Hispanic households.

The additional measures of poverty status and functional impairment considered in the sampling scheme were not examined, as a consequence of transitions in classification over time that characterize these measures. Screening for these measures is further complicated by the considerable degree of movement into and out of poverty in any two years (Moeller and Mathiowetz, 1990) and the potential movement of elderly individuals from a state of good health to that of disability over time.

Although addresses that experienced a sample movement out were characterized by a shortfall in expected sample size by virtue of the address sample design, a more important consideration was the effect on sample yields for the overall Round One sample. A comparison of the targeted overall Round One sample (36,511) and the resultant Round One sample (35,418) revealed a three percent shortfall. Relative to the targeted Round One sample, a greater shortfall in sample size characterized the black and Hispanic population subgroups (four percent), when contrasted with the white and other races population subgroup (2 percent). Clearly, the overall effect of the address sample design on the Round One sample yields was minimal.

o EFFECT OF ADDRESS SAMPLE DESIGN ON PRECISION

The precision of the NMES survey estimates for the population subgroups of analytical

interest was expected to decrease only marginally, as a function of the small loss in sample size attributable to the address sample design. Even for the population subgroup characterized by the largest relative sample loss (4.3 percent for Hispanics), the expected increment in the standard errors of their associated survey estimates was only two percent. The major cause for concern, however, was the greater variation in sampling weights for the population subgroups of analytical interest (Kish. 1965). Table 2 presents a summary of the variation in the sampling weights that characterized the targeted and actual Round One samples. More specifically, the distributions of the unadjusted sampling weights, and their coefficient of variation are compared across the alternative sample designs for Round One.

TABLE 2 ABOUT HERE

The actual Round One sample was characterized by a higher level of variation in sampling weights across all population subgroups targeted for oversampling. This was most obvious for the black and Hispanic population subgroups. The greater variation in their sampling weights was primarily a function of the movement of new individuals, with characteristics targeted for oversampling, into sampling addresses selected in NMES at a lower sampling rate based on their composition at the time of the screener interview.

The impact of the address sample design on the precision of survey estimates was directly examined by a comparison of the standard errors for a representative set of demographic and health status measures that were available from the screener interview. It was necessary to restrict this comparison to these measures, since the screener only-respondents did not have any

of the information that was collected only in the Round One interview.

The person level demographic measures under investigation included region, size of city, marital status and veteran status. In addition, the health status measures under investigation included all questions in the NMES screener data base that identified functionally impaired individuals for the purposes of oversampling. For each of the survey measures under consideration, the ratio of the standard errors derived from the address sample design and from the potential Round One sample was computed. Study findings indicate only a slight increase in the standard errors of the survey estimates that characterize the address sample design (ratios generally ranged from 1.01 to 1.04). Even for the population subgroup characterized by the greatest loss in precision (white/other race, 65+), the average increment in the standard errors of their associated survey estimates was only six percent.

The cost savings achieved by the address sample design were estimated to be $175,000, and were attributable to not having to locate the targeted individuals that had a change of address since responding to the NMES screener interview. The availability of these additional funds allowed for the implementation of more intensive survey nonresponse conversion techniques to enhance the overall NMES household survey response rate. As a consequence of the address sample design, the response rate that characterized the NMES Round One interview was not affected by a component of nonresponse due to inability to locate.

Under a NMES design that attempted to include the movers in the sample, the field period for the Round One interview would have had to be extended. This would be required in order to accommodate the time needed to locate the movers and conduct the interview. This

extension in the field period would have a potentially adverse effect on the quality of data obtained from the movers, due to the extended length of the recall period that characterized their interview. Use of an address sample design for NMES eliminated the component of response error that would be attributable to this particular extension in recall period. Alternatively, the problems inherent with tracing movers in a NMES survey that considered a household design could be noticeably reduced by a field redesign that facilitated a significant reduction in the average lag period between the screener and Round One interviews.

The NMES address sample design was cost effective and achieved an acceptable response rate relative to a design that would have tracked movers. In addition, the quality of the Round One data was expected to be improved, as a consequence of the shorter length of recall period (where respondents were required to provide health care information at the time of the Round One interview relative to 1/1/87) that characterized the respondents of the address sample design. The major potential limitation of the adopted design was with respect to the precision of survey estimates. Even on this sensitive dimension, study findings indicated only a slight loss in the precision of survey estimates derived from the address sample design. Consequently, the overall benefits of the address sample design were realized in the National Medical Expenditure Survey, with only minimal effects on the precision of survey estimates.

3. Refielding A Sample of Screener Refusals and Other Nonresponding Dwelling Units in the First Round of the NMES Household Survey.

A supplemental sample of refusals to the screener and other nonresponding dwelling units was

also included in the round 1 sample, to improve the overall NMES response rate over the four rounds of data collection. In addition to 1,600 dwelling units classified as screener interview refusals, another 916 dwelling units were categorized as "other nonresponse". The other nonresponse classification included dwelling units where no one was home after four calls, where the potential household respondents were unavailable during the screener field period, and where respondents were considered too ill to complete the interview.

Refusals and other nonresponding dwelling units were classified into the same four strata used for sampling vacant addresses. A sample of 645 addresses with dwelling units categorized as refusals to the screener was then selected from the 1,600 addresses with such eligible units and a sample of 376 addresses with dwelling units categorized as other nonresponding units were selected from 916 addresses with eligible units with this classification. Consequently, the only set of screener nonrespondents that were not considered for selection in the first round of the NMES household survey consisted of dwelling units with a language problem, dwelling units where the interviewer was unable to enter the structure, and those classified in the residual nonresponse category. These 234 dwelling units were not targeted for selection in round 1 because of the low probability of converting them to participate in the four rounds of the household survey. By recontacting the nonrespondents to the NMES screener interview, 99.26 percent of the sample of all households were given a chance to participate in the NMES round 1 interview. It should be noted, however, that any new dwelling units built during the five month lag time between the screener and the round 1 interview were not included in the round 1 sample frame. The final round 1 sample is summarized in Table 3.

TABLE 3 ABOUT HERE

Round One Field Results

As noted, the joint screener-round one response rate for the targeted round 1 address sample based on responses to the screener interview, including the screener vacant sample, was 82.1 percent. This round 1 response rate was improved by recontacting nonrespondents to the NMES screener interview. In the set of refielded addresses which were characterized by screener nonresponse, 372 (40.1 percent) of the 928 eligible reporting units responded in round 1. This sampling approach contributed an additional 3.3 percent to the joint response rate. Overall, the joint screener round one response rate for the NMES was 85.4 percent.

4. Part-Year Nonresponse in the NMES Household Survey

Panel surveys are subject to wave nonresponse, which occurs when responses are obtained for some but not all waves of the survey. Prior to selecting an adjustment strategy, it is essential that the patterns of nonresponse and potential correlates are examined. When levels of sample attrition are high, it has been suggested that wave nonresponse is related to level of respondent burden (Czajka, 1986). Generally, the process of sample attrition is non-random (Short and McArthur, 1986). Partial respondents are often distinguished from their complete respondent counterparts on a number of dimensions. These distinctions should be considered in the determination of a nonresponse adjustment strategy.

Two general strategies to correct for sample attrition have gained acceptability in the statistical community: sampling weight nonresponse adjustments and imputation (Kalton, 1986;

Czajka, 1986; Herringa and Lepkowski, 1986). Weight adjustments for sample attrition are rather straightforward to implement and avoid the increase in sampling variance expected in survey estimates as a function of imputation. The primary limiting features of the technique are the loss of large amounts of useful data provided by partial respondents, and the deleterious impact of large nonresponse adjustments to sampling weights on the precision of survey estimates. Imputation for panel attrition allows for the inclusion of partial respondents in the derivation of survey estimates, and the use of the data they have provided. Within the imputation framework, there are several general methodologies that are considered to compensate for panel nonresponse: longitudinal hot deck imputation (Herringa and Lepkowski, 1986). Due to the large number of time dependent analytical measures that are directly affected by sample attrition and the sophisticated software requirements to implement the technique, imputation compensation strategies are the more costly of the two types of strategies.

o Characteristics of the NMES Part-Year Respondents

Of 36,753 key participants in the NMES household survey, 2,294, or 6.2 percent, responded for some, but not all, of the time period in 1987 for which they were eligible. Key sample respondents to the household survey consisted of all civilian non-institutionalized individuals who responded to the Round One interview, in addition to individuals who joined responding Round One reporting units and did not have an opportunity for selection during the period of time that spanned the Round One field period (new babies, military returning to civilian status, individuals in institutions or outside the country returning to their primary

residence). To ascertain the potential level of nonresponse bias that was attributable to partial response in NMES, it was necessary to determine whether the part-year respondents differed systematically from their full year counterparts. To facilitate these comparisons, demographic profiles of these two distinct respondent groups were compared for the following measures: gender, age, race/ethnicity, marital status, years of school completed, census region, size of city, and indicators of functional status.

Since the nonresponse adjustment strategies employed to correct for part-year nonresponse would be dependent on a respondent's data profile for his period of participation in the survey, it was necessary to impose a thresshold on what constituted a minimally acceptable time representation of partial data for making annual national health care estimates. In NMES-2, the minimum part-year response requirement of data for more than one-third of a respondent's period of eligibility followed the approach taken in the 1980 National Medical Care Utilization and Expenditure Survey (sponsored by the National Center for Health Statistics). In NMES-2, 48.6 percent of the part-year respondents (1,114 individuals) who constituted 6.2 percent of the person level sample, did not satisfy this criterion. Consequently, they were treated as total nonrespondents and a standard weighting class adjustment for non-response was applied to the sampling weights of the remaining respondents to correct for their exclusion.

The comparisons of the race/ethnicity distributions for the part-year respondents with data for at least one-third of their period of eligibility in 1987 (henceforth referred to as part-year respondents) and full year respondents revealed a significantly greater representation of whites

and other races (non-black, non-Hispanic) among the individuals that provided complete data when compared to the partial respondents (Cohen, Johnson and Carlson, 1989). This pattern was also observed in the 1977 National Medical Care Expenditure Survey (NMCES), which was the predecessor of NMES (Cohen, 1982). Alternatively, the partial respondents had a higher representation of Hispanics than their complete data counterparts. There was also a significantly higher representation of the partial respondents living in the 19 largest Standard Metropolitan Statistical Areas (SMSAs) in the nation, indicating a greater likelihood of sample attrition in NMES in the large urban metropolitan areas. Furthermore, partial respondents were more likely to reside in the Northeast region of the United States than individuals who provided full year response profiles.

No significant gender differentials were noted across respondent groups. With respect to marital status, the full year respondents were more likely to be married than their partial respondent counterparts. Alternatively, the partial respondents had a greater representation of never married individuals, which mirrored the 1977 NMCES experience. Furthermore, a comparison of the age distributions that characterized the respective respondent groups revealed that partial respondents were more likely to be aged 20-29. Since this age group represents a highly mobile population subgroup, this suggests that the sample attrition that they displayed in the NMES was partially a function of migration. Furthermore, complete respondents had a higher representation of elderly individuals between the ages 70-74 than their partial respondent counterparts. With respect to years of education completed, the partial respondents had a higher representation of individuals with some high school training, as indicated by at least 9-11 years

of education.

To minimize the nonresponse bias in survey estimates attributable to partial response, an appropriate estimation strategy is needed to adjust the data for the remaining sample members who did not respond for their entire period of eligibility. In view of the programming time and cost necessary to implement an imputation strategy to correct for partial nonresponse, and the relatively small representation of partial nonrespondents in NMES, the advantages of this technique for NMES application were not obvious. The technique would require the linkage of partial respondents to complete respondents with matching demographic and health status profiles, the extraction of data from the complete respondent which corresponded to the nonresponding time period of the partial respondent, and its imputation to the partial respondent for each time dependent variable in NMES-2, representing a complex and expensive process.

Traditionally, when the level of partial response is low, it is often preferable to treat partial respondents as complete nonrespondents. Using this approach, only those sample participants providing complete data would be used in the analysis. This was the approach taken for the remaining part-year respondents in the NMES-2 household survey. Weighting classes were formed by cross-classifications of the following measures: race/ethnicity, age and gender. The person level sampling weights for the full year respondents were further post-stratified to poverty status estimates derived from the Current Population Survey.

5. An Estimation Strategy to Represent the Institutional User Population in the NMES-2

Institutional Population Component

The Institutional Population Component (IPC) of the National Medical Expenditure Survey (NMES) was established to provide an assessment of the health care utilization, costs, sources of payment and health insurance coverage of the U.S. institutionalized population residing in nursing and personal care homes (NH), and in facilities for the mentally retarded (MR). The primary objective of the survey was to estimate the use of and expenses for health care services for all persons residing in institutions at any time during calendar year 1987. To obtain a nationally representative sample of the 1987 institutional user population, the survey included a sample of residents residing in selected facilities as of January 1, 1987, in addition to a representative sample of admissions to the selected facilities over the course of 1987. The union of these samples served to represent the 1987 institutional user population.

o Sample Design

The adopted NMES institutional population survey is a stratified, two stage probability design with two phases of facility selection. Current residents (residents on January 1, 1987) and admissions (persons admitted between January 1, and December 31, 1987) were sampled within participating facilities at the second stage.

The IPC facility sample consisted of 851 eligible nursing and personal care homes and 730 eligible facilities for the mentally retarded. Facilities were considered to be respondents to the survey when they completed a Facility Questionnaire. Consequently, the IPC facility level response rate was 95.2 percent for nursing and personal care homes, and 94.7 percent for

facilities for the mentally retarded.

The design of the survey required that the institutional use and expenditure data for current residents were to be collected for their entire period(s) of institutionalization in 1987. In contrast, IPC data collection for the admissions sample began with their first admission to a sampled IPC eligible facility, independent of prior institutional stays over the course of 1987. Consequently, their 1987 institutional data collection period was constrained. For estimation purposes, individuals who responded for at least a third of their eligibility period of institutional data collection were considered respondents.

In the nursing and personal care IPC sample, 805 participating facilities (94.6 percent) allowed for the selection of a sample of their residents as of January 1, 1987. Overall, 3,392 eligible residents were selected, representing a national nursing and personal care home population of 1.5 million residents. Similarly, in the IPC sample of facilities for the mentally retarded, 685 participating facilities (93.8 percent) allowed for "current" resident sampling. Overall, 3,738 eligible residents were selected, representing a national population of 212,000 residents in facilities for the mentally retarded.

Overall, the response rate in the IPC for current residents providing data for at least one-third of their period of institutionalization in 1987 was 89.5 percent for residents in nursing and personal care homes (.946 facility level response rate x .946 resident level response rate), and 88.4 percent for residents in facilities for the mentally retarded (.938 facility level response rate x .942 resident level response rate). This data was to be obtained in the IPC through the administration of the Institutional Use and Expenditure Questionnaire (IUEQ), to be completed

by facility staff (Edwards and Edwards, 1989).

The admissions sample consisted of 2,608 eligible sampled admissions to nursing and personal care homes, and 889 eligible sampled admissions to facilities for the mentally retarded. Sampled admissions were defined to be individuals who were admitted to the sampled IPC facility during 1987 and had no prior admissions to that facility during the survey year.

In the nursing and personal care home sample, 758 participating facilities (89.1 percent) allowed for the sample selection of admissions at all rounds of data collection. Similarly, 657 facilities for the mentally retarded (90 percent) allowed for the sample selection of new admissions at all rounds of data collection. .

Overall, the response rate for new admissions providing data for at least one-third of their period of institutionalization in 1987 was 81.2 percent for those sampled in nursing and personal care homes (.891 facility level response rate x .911 admission response), and 81.3 percent for admissions sampled in facilities for the mentally retarded (.900 facility level response rate x .903 admissions response rate).

Data collected from facility respondents included facility level characteristics, physical and mental health status and functional limitations of sampled persons, and their socio-demographic characteristics and residential history in and immediately before admission to sampled facilities. Information collected on health care services use and expenses included facility services provided, charges and sources of payment, hospitalizations during the institutionalized period and associated conditions, number of physician contacts, and contacts with other medical care providers and therapists.

This data collection effort was referred to as the Survey in Institutions (SII). During each visit,

interviewers obtained or constructed lists of residents from each cooperating facility and proceeded to select the sample (Edwards and Edwards, 1989). The current resident sample was selected from a list of all residents in sample facilities as of January 1, 1987. Similarly, the admission samples were selected on three separate occasions in cooperating facilities from separate lists of all admissions that occurred during the following time periods in 1987: January 1 to April 30, May 1 to August 31, and September 1 to December 31. Sampled persons were followed throughout 1987. For those who left the facilities in which they were selected, facility use and expenditure data were collected up to the time of discharge. If a sample person entered another IPC-eligible facility, the institutional data collection procedures were continued in the new facility.

Since study objectives required data that facility staff could not be expected to provide, the IPC also included a Survey of Next of Kin. This survey consisted of a set of questionnaires administered to community respondents who knew about sampled persons and their lives outside of institutions. Data were obtained on use and expenditures linked to specific residence periods, living arrangements outside of sampled institutions, perceptions of health status and functional limitations, and arrangements for informal care.

o The NMES Institutional User Population

The IPC sample design consisted of two distinct selections of 1987 institutional users: the first selection was designed to provide a representative national sample of residents in IPC

eligible facilities as of January, 1, 1987 (current residents); and the other selection was designed to provide a nationally representative sample of 1987 admissions to IPC eligible facilities. The strict requirement of a single day of sample eligibility for the current resident sample resulted in a single opportunity of selection for each sampled current resident as of 1/1/87. Imposition of a similar restriction for the selection of admissions, requiring the selection of individuals experiencing their first institutional stay in 1987, would have simplified the sample design by allowing each sampled institutional user a single opportunity of selection. Since this information regarding an individual's prior periods of institutionalization was not available at the time of sample selection, and often unavailable from facility records, such a restriction could not be imposed. Resident history information for sampled admissions was often obtained through the IPC Survey of Next of Kin, whereby community respondents who knew about sampled persons would be the primary source for information regarding prior institutional stays.

As a consequence of the sample selection scheme that was employed, an individual who experienced more than one institutional stay over the course of 1987 had multiple chances of selection into the IPC sample. Furthermore, a subset of sampled admissions was determined to have also resided in an IPC eligible facility on 1/1/87, indicating an overlap with the independent sample of January 1 residents. In order to identify the sample of institutional users that had multiple opportunities of selection in the IPC sample, it was necessary to further classify the IPC sample of institutional users according to their institutional experience over the course of 1987.

o Classification of Current Resident Sample

With respect to the population of institutional users that resided in an IPC eligible facility on January 1, 1987 (referred to as current residents), four mutually exclusive and exhaustive classifications are specified in order to characterize their institutional experience over the course of 1987 (Figure 1). More specifically, the first group of current residents consists of institutional users who remained in the same facility over the course of 1987. Residents in this class are referred to as **static full year residents (Group 1)**. The next class of current residents consists of institutional users who remained in the same facility for only part of calendar year 1987, with no subsequent admissions to IPC eligible facilities (i.e., nursing and personal care homes and facilities for the mentally retarded that met the definition for eligibility in the NMES IPC) over the course of 1987. These institutional users are referred to as **single stay part-year residents (Group 2)**. Current residents with this classification could have returned to the community as a member of the civilian non-institutionalized population, been transferred to an out-of scope facility or institution (e.g., acute care hospital, psychiatric institution), or died while in the institutional setting.

FIGURE 2 ABOUT HERE

The remaining current residents experienced at least one subsequent admission to an IPC eligible facility over the course of 1987. They are distinguished in the following manner. The first group consists of current residents who were formally discharged from the facility they resided in as of January 1, 1987 and subsequently were readmitted to the same facility over the course of 1987. These institutional users are referred to as **current residents with re-admissions to same facility (Group 3)**, and consist of residents with one or more re-admissions restricted to the same facility over the course of 1987. Alternatively, the remaining group consists of

current residents with admissions to different facilities (Group 4). This classification also includes institutional users who were re-admitted to the same facility they resided in on 1/1/87 **and** who also experienced at least one admission to a different IPC eligible facility over the course of 1987.

Classification of Individuals Sampled as Admissions

In a complementary manner, four mutually exclusive and exhaustive classifications are specified in order to characterize individuals who experienced at least one admission to an IPC eligible facility over the course of 1987: these individuals are referred to as sampled admissions (Figure 1). The first class of affected institutional users consists of individuals who were not residents in IPC eligible facilities as of 1/1/87 and whose first institutional admission in 1987 was in a sampled IPC facility (Primary Sample Facility (PSF)). Institutional users in this class are referred to as **sampled admissions with initial 1987 admission to an IPC Primary Sample Facility (Group 5)**. This group includes individuals with one or more unique admissions to eligible institutions over the course of 1987 (Groups 5a or 5b, and Groups 5c or 5d, respectively) .

The next classification identifies individuals with a 1987 institutional admission to an IPC Primary Sample Facility, who were also residents in the same facility as of 1/1/87. Institutional users in this group are referred to as **residents sampled as admissions with 1987 admission(s) to the same IPC Primary Sample Facility (Group 6)**. This group of institutional users was already represented in the NMES Institutional Population Component Survey by current residents classified as **residents with re-admissions to same facility (Group 3)**, and by a subset of the

residents with admissions to different facilities (Group 4) who were also re-admitted to the Primary Sample Facility. As a consequence of the ease in identifying these current residents who were re-admitted to the same facility over the course of 1987, they were not considered eligible for IPC data collection. A related group of institutional users consists of individuals with an admission to an IPC Primary Sample Facility who were also residents in an non-sampled IPC eligible facility as of 1/1/87. Such individuals are referred to as **residents in non-sampled facilities with 1987 admission(s) to an IPC Primary Sample Facility (Group 7)**. This class of institutional users was also dually represented in the NMES IPC sample by a subset of the **current residents with admissions to different facilities (Group 4)**. These sampled admissions were not excluded from IPC data collection as a consequence of being unable to determine, at the time of sampling, whether they were institutionalized in some other IPC eligible facility on January 1, 1987.

The remaining group of institutional users with 1987 admissions consists of individuals who did not reside in IPC eligible settings as of 1/1/87, and who were admitted to non-sampled IPC eligible facilities in 1987 prior to an admission to a Primary Sample Facility. Institutional users in this class are referred to as **admissions in IPC Primary Sample Facilities with initial 1987 admission to a non-sampled IPC eligible facility (Group 8)**. This class of institutional users was also dually represented in the NMES IPC sample by a subset of the **new admissions with initial 1987 admission to an IPC Primary Sample Facility (Group 5d)**.

IPC Sample Distribution of Institutional Users

A summary of the IPC sample distribution of institutional users, further distinguished by

facility setting, is presented in Table 4. These tabulations include individuals with response profiles for utilization and expenditure data for at least a third of their period(s) of institutionalization in 1987, beginning with their **sampled stay.** For current residents, this translated to their entire period(s) of institutionalization in 1987. Alternatively, individuals with 1987 admissions were classified as respondents when response profiles were obtained for at least a third of their institutional experience in 1987, beginning with their **sampled admission.** Inclusion of these partial respondents in the derivation of national health care utilization and expenditure estimates for the institutional user population requires implementation of an imputation procedure to adjust for missing time dependent data (Cohen and Potter, 1990). Since IPC data collection for the admission sample began with their first admission to a sampled IPC facility, their 1987 institutional data collection period was constrained. When resident history information was not obtained either through the IPC Survey of Next of Kin or the IPC Survey in Institutions for periods in 1987 prior to their **sampled admission,** resident history profiles were imputed for the missing time gaps in 1987 (Potter and Cunningham, 1990). Inclusion of these sampled admissions with prior periods of institutionalization in the derivation of national health care utilization and expenditure estimates for the institutional user population also requires implementation of **additional** imputation procedures to correct for missing time dependent data (Cohen and Potter, 1990).

The sample of institutional users in nursing and personal care homes consisted of 5,585 respondents, with 3,209 (57.5 percent) sampled as current residents and 2,376 (42.5 percent) sampled as 1987 admissions (Table 4). Relative to the current resident sample, 2,586 (80.6 percent) were classified as static full year residents (Group 1), another 150 (4.7 percent) were

single stay part-year residents (Group 2), with the remaining 473 (14.7 percent) experiencing subsequent admissions to eligible facilities in 1987 (Groups 3 and 4). After excluding the 448 sampled admissions that were considered ineligible for IPC data collection (Group 6), the admission sample was dominated by 2,002 (84.3 percent) institutional users who were not institutionalized on 1/1/87 (Groups 5 and 8).

Alternatively, the sample of institutional users in facilities for the mentally retarded consisted of 4,323 respondents, with 3,520 (81.4 percent) sampled as current residents and 803 (18.6 percent) sampled as 1987 admissions (Table 4). As a consequence of the low representation of **sampled admissions** in these types of facilities in any given year, IPC sample size specifications for the admission sample in facilities for the mentally retarded did not assume separate national estimates would be made for the **sampled admissions**. Relative to the current resident sample, 3,089 (87.8 percent) were classified as static full year residents (Group 1), and another 73 (2.1 percent) were single stay part-year residents (Group 2). The remaining 358 experiencing subsequent admissions to eligible facilities in 1987 (Groups 3 and 4) with the majority (316) experiencing admissions to non-sampled facilities (Group 4). After excluding the 76 sampled admissions that were considered ineligible for IPC data collection (Group 6), the admission sample was primarily represented by 432 (53.8 percent) institutional users who were not institutionalized on 1/1/87 (Groups 5 and 8).

TABLE 4 ABOUT HERE

o Estimation Strategy for the Institutional User Population

The stratified, multi-stage probability sample design adopted for the IPC institutional user population allows for the derivation of approximately unbiased estimates of health care parameters at the national level. This is conditioned upon the application of sampling weights to the sample data that properly reflect the sample selection scheme. The sampling weight for a sample member is defined as the reciprocal of a sample unit's probability of selection (Cox and Cohen, 1985). The estimation strategy for the IPC includes additional adjustments for all levels of nonresponse experienced in the survey. Nonresponse adjustments to the sampling weights have been implemented at the facility level and the institutional user level. To further improve the precision of survey estimates that characterize the IPC sample, post-stratification adjustments on facility level and resident level characteristics have also been implemented, using information from the 1986 Inventory of Long Term Care Places (Flyer, 1992).

One estimation strategy under consideration attempted to maximize the precision in survey estimates that characterize the institutional user population by the inclusion of all responding sampled institutional users in the estimation process. However, the implementation of this estimation strategy is not without penalty. In order to derive national estimates of the health care utilization and expenditure experience for the institutional user population, an imputation strategy must be considered to correct for missing time dependent health care data associated with institutional stays in 1987 prior to the sampled admission. Greater programming resources are required to implement the imputation process that corrects for missing time dependent health care utilization and expenditure data. The approach requires a determination of the exact time period for which institutional data is missing, a linkage between the institutional user with missing time dependent data to the best matching donor with a complete

data profile (using minimum distance function techniques), and imputing the appropriate time dependent data from donor to recipient (Cohen, 1992). The inclusion of sampled units that have multiple opportunities of selection in the derivation of national estimates requires additional adjustments to the survey sampling weights that reflect corrections to sample unit selection probabilities. Furthermore, the inclusion of a multiplicity adjustment to the estimation weights of institutional users selected from the IPC admission sample adds greater variability to their sampling weight distribution, partially limiting the expected gain in precision associated with an increase in sample size.

Given the complexities associated with the implementation of this strategy, and the need to provide timely national health care expenditure estimates of the institutional user population, an alternative approach was adopted for the derivation of use and expenditure estimates and other time dependent measures. More specifically, the alternative estimation strategy restricted the admission sample to institutional users whose first institutional stay in 1987 was in a sampled facility (i.e., Group 5). Adoption of this approach obviates the need for an imputation strategy to correct for missing time dependent data associated with institutional stays in 1987 prior to an institutional user's sampled admission. Furthermore, the restriction of the IPC admission sample to a sample of first institutional stays in 1987 obviates the need for a multiplicity adjustment to estimation weights and an adjustment for dual frame representation of residents in facilities as of 1/1/87. This is a consequence of limiting the sample of institutional users to a single opportunity of selection. Institutional users determined to have experienced institutional stays prior to their sampled admission (Groups 7 and 8) would be defined as ineligible for the

purposes of estimation.

Implementation of this approach results in a sample diminution of 513 out of 5585 respondents for the sample in nursing and personal care homes (9.2 percent reduction) and a comparable loss of 403 out 4,323 respondents in the sample of facilities for the mentally retarded (9.3 percent reduction). The impact of this sample restriction on the precision of survey estimates translates to a 4.9 percent increase in the standard errors that characterize the survey estimates of all institutional users in nursing and personal care homes and a corresponding 5 percent increment for the sample in facilities for the mentally retarded. Greater increments in standard errors are to be noted for the subset of institutional users that experience admissions during 1987. Although the magnitude of this loss in precision is non-negligible, it falls within acceptable levels when contrasted with the time and resource demands inherent in the implementation of an imputation strategy to correct for missing time dependent data associated with institutional stays in 1987 prior to an institutional user's sampled admission. Furthermore, consideration of the restricted first admission sample for the purposes of estimation eliminates exposure to a component of nonresponse bias due to missing time dependent data for prior institutional stays. This component of bias is often only partially reduced through application of imputation techniques. Implementation of imputation strategies to correct for missing time dependent health care utilization and expenditure data associated with prior institutional stays will inform future methodological investigations regarding the impact on survey estimates and their precision due to the inclusion of individuals whose sampled admission was not their first institutional stay in 1987.

6. Summary

The next cycle of the National Medical Expenditure Survey (NMES-3) will be in the field from 1996 through 1997, in order to obtain health care utilization, expenditure and insurance coverage information that characterizes the health care experience of the civilian non-institutionalized population and the population in nursing home for calendar year 1996. A number of design strategies that have been adopted in NMES-2 will also be considered for the NMES-3. One design feature of the NMES-2 Household Survey that will not be adopted for NMES-2 is the address sample design. This decision was not based on any design limitations that were identified in NMES-2 with the adoption of the address sample design, but driven by additional analytical demands placed on the survey. One of the primary motivations for the choice of 1996 as the time period for data collection was the need to have baseline data to assess the impact of health care reform on the nation's health care experience. A number of fast-track states are already in the process of implement health care reform initiatives. Consequently, the households identified for sample selection through the administration of a screening interview will also be subject to an additional interview in the fall of 1995. to gauge their satisfaction with the health care delivery system, their perceptions regarding access, and their current level of health insurance coverage. Since one of the analytical objectives of the NMES-3 household survey will be to assess the longitudinal changes in insurance coverage and access to health health care system over time, including data from the fall of 1995, it will be necessary to follow the same individuals that complete the screening and baseline interviews in the fall of 1995.

The NMES-2 was successful with its strategy to re-field a supplemental sample of

refusals to the screener and other nonresponding dwelling units in the first core round of data collection for the Household Survey. This strategy resulted in a three percent improvement to the overall NMES-2 response rate over the four rounds of data collection. As response rates are monitored for both the NMES-3 household and nursing home surveys, this design strategy will again be given serious consideration.

As a consequence of the panel designs of the NMES-3 household and nursing home surveys, wave nonresponse will remain a concern. As in NMES-2, methodological investigations will be conducted to determine whether the part-year respondents differ systematically from their full year counterparts. Both the level of nonresponse encountered in NMES-3 and the results of the evaluation of the patterns of nonresponse will guide the choice of the nonresponse adjustment strategy that is to be implemented.

Based on the estimation strategy considered in NMES-2 to represent the institutional user population, the NMES-3 nursing home survey design will employ a data collection scheme that limits the likelihood of multiple opportunities of selection into the survey. The planned design will restrict the admission sample for estimation purposes to a sample of individuals experiencing their first institutional stay in a sampled facility for the targeted survey year. Results from the NMES Institutional Population Component Feasibility Study revealed that facility respondents are able to provide accurate information regarding an individual's prior period of institutionalization in a given year. It is recognized that the restriction of the sample of admissions to the individuals experiencing their first institutional stay in a sampled facility for

the targeted survey year cannot be completely implemented at the time of sample selection. However, the ability to use facility information on prior admissions for the selected sample will result in significant cost savings to the survey, based on a reduction in unnecessary data collection activities associated with cases that are not eligible for estimation purposes.

7. References

Cohen, S. B. (1982). "An Analysis of Alternative Imputation Strategies for Individuals with Partial Data in the National Medical Care Expenditure Survey", Review of Public Data Use, 10:153-165.

Cohen, S. B., P. Flyer and D.E.B. Potter (1987). "Sample Design of the Institutional Population Component of the National Medical Expenditure Survey". Presented at the Annual Meeting of the American Public Health Association, New Orleans, LA.

Cohen, S. B., R. DiGaetano and J. Waksberg (1991). Sample Design of the National Medical Expenditure Survey Household Survey, Methods 3. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research.

Cohen, S. B., D.E.B. Potter and P. Flyer (1993). Sample Design of the National Medical Expenditure Survey Institutional Population Component, Methods 6. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research.

Cohen, S.B., A.E. Johnson and B.L. Carlson (1989). "An Analysis of Part-Year Nonresponse in the Household Component of the National Medical Expenditure Survey". Journal of Economic and Social Measurement, 15:281-299.

Cohen, S.B., and D.E.B. Potter (1993). "An Estimation Strategy to Represent the Institutional User Population in the 1987 National Medical Expenditure Survey". Journal of Economic and Social Measurement, 19:159-177.

Cohen, S.B. and A.E. Johnson (1992). "Address Sample Effects the National Medical Expenditure Survey". The American Statistician 46: 100-105.

Cohen, S.B.and D.E.B. Potter (1990). "An Analysis of Part-Year Nonresponse in the Institutional Population Component of the National Medical Expenditure Survey". American Statistical Association, Proceedings of the Section on Survey Research Methods.

Cox, B.G., and S.B. Cohen (1985). Methodological Issues for Health Care Surveys. New York, Basel: Marcel Dekker, Inc.

Cox, B., R. Folsom, B. Moser and T. Virag (1979). Housing Unit Versus Household as the Sampling Unit, Report No.RTI/1725/01-01S, Research Triangle Park, North Carolina.

Czajka, J.L. (1986). "Discussion of Papers on Nonresponse in the Survey of Income and Program Participation". Proceedings of the Section on Survey Research Methods.

DiGaetano, R.G. (1987). Sampling Report for the Selection of Dwelling Units for the Household Survey of the NMES. NMES Report No. 1.00. Westat, Inc., Rockville, Maryland.

DiGaetano, R.G., and J. Waksberg (1986). Sample Design of the Household Survey (HHS) and the Survey of American Indians and Alaska Natives (SAIAN). NMES Report No. 1.46. Westat, Inc., Rockville, Maryland.

Edwards, W. S. and B. Edwards (1989). National Medical Expenditure Survey: Questionnaires and Data Collection Methods for the Institutional Population Component. Methods 1. Department of Health and Human Services, Public Health Service, National Center for Health Services and Health Care Technology Assessment, DHHS Pub. No. (PHS) 89-3440.

Flyer, P. (1992). Sample Design for the Institutional Population Component (IPC) of the National Medical Expenditure Survey. NMES Internal Report. Westat, Inc., Rockville, Maryland.

Herringa, S.G. and J.M. Lepkowski (1986). "Longitudinal Imputation for SIPP". ASA.

Proceedings of the Section on Survey Research Methods.

Kalton, G. "Handling Wave Nonresponse in Panel Surveys". Journal of Official Statistics, 2:303-314.

Kish, L. (1965). Survey Sampling. New York: New York, John Wiley and Sons.

Moeller, J.F. and N. A. Mathiowetz (1990). Problems of Screening for Poverty Status. American Statistical Association: Proceedings of the Section on Survey Research Methods.

Potter, D.E.B. and P.J. Cunningham (1990). "Constructing Year-Long Profiles of Residence Histories for a National Sample of Nursing Home Users." American Statistical Association, Proceedings of the Section on Survey Research Methods.

Sirken, M. G. (1970). "Household Surveys with Multiplicity." Journal of the American Statistical Association, 65,257-266.

Short, K. and E. McArthur (1986). "Life Event and Sample Attrition in the Survey of Income and Program Participation". ASA, Proceedings of the Section on Survey Research Methods.

U.S. Bureau of the Census (1987). Current Population Reports, Series P-25, No. 957.

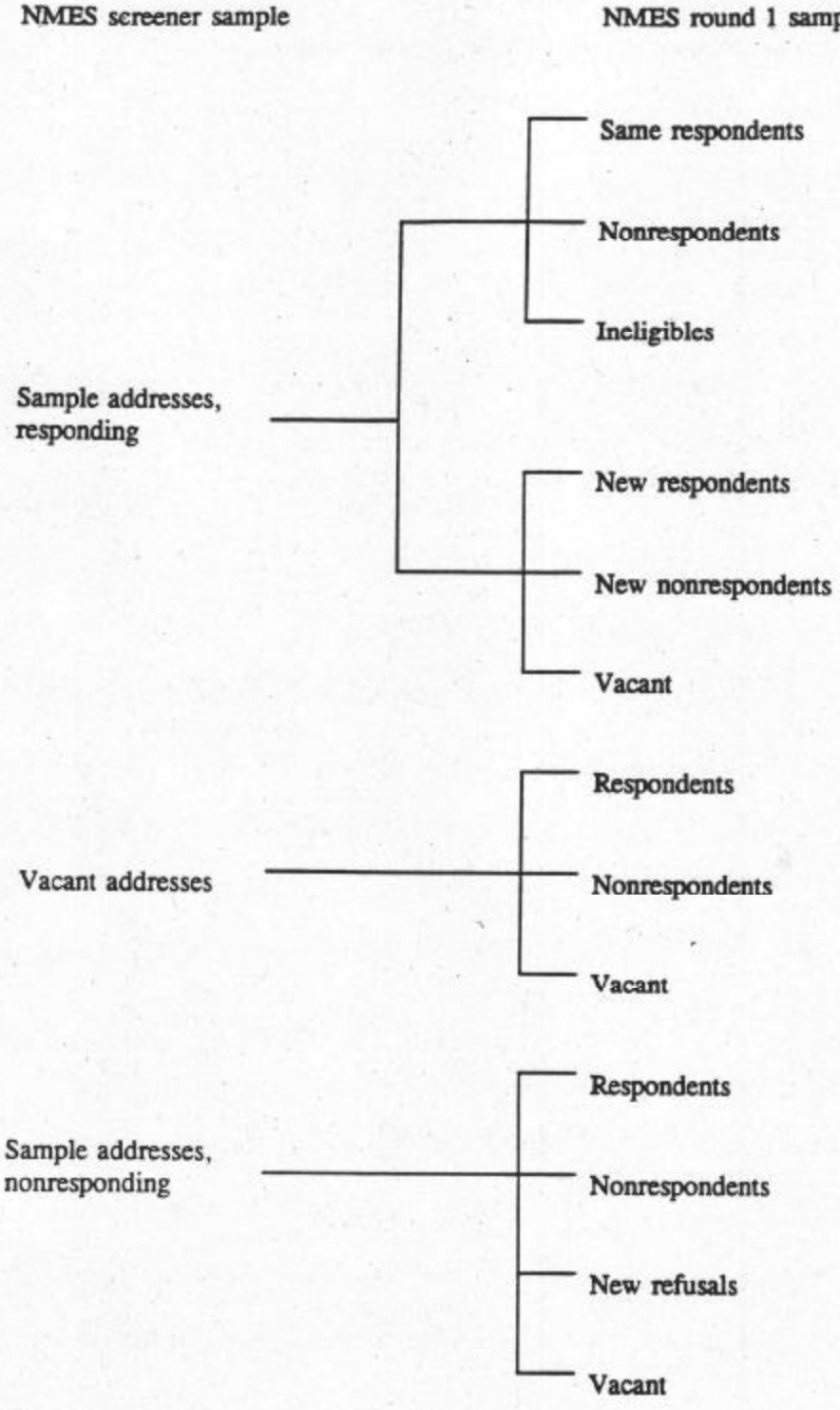


Figure 1. Results of NMES Address Sample Design: Transitions from Screener to Round 1.

Table 1. Comparison of Demographic Characteristics of NMES Movers

Race/ethnicity	Age	Screener-only respondents[a]	Expected[b] Round 1 response	Round 1-only respondents[c]	Round 1-only respondents relative to: Targeted sample (%)	Round 1-only respondents relative to: Expected sample (%)
Hispanic	<65	615	523	447	72.7	85.5
	65+	13	11	6	46.2	54.6
Subtotal		628	534	453	72.1	84.8
Black/non-Hispanic	<65	978	831	686	70.1	82.6
	65+	33	28	20	60.6	71.4
Subtotal	<65	1,011	859	706	69.8	82.2
White/other	65+	2,400	2,040	1,926	80.3	94.4
		267	227	128	47.9	56.4
Subtotal		2,667	2,267	2,054	77.0	90.6
Total		4,306	3,660	3,213	74.6	87.8
National estimates (in thousands)		(24,060)		(21,500)		

[a] Screener-only respondents are those who moved away from sampled addresses prior to the administration of the NMES Round 1 interview.

[b] Under assumption of 85% response rate.

[c] Round 1-only respondents are individuals new to the NMES sample as a function of the address sample design.

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

Table 2. Variation in Sampling Weights and Mean Standard Error Ratios

Race/ethnicity	Age	Distribution of weights						
		Potential Round 1 sample			Actual Round 1 sample			
		Coefficient of variation	Percentiles		Coefficient of variation	Percentiles		
			1st	99th		1st	95th	99th
Hispanic	<65	.08	2562	3420	.19	2562	3420	7935
	65+	.08	2562	3925	.19	2562	3420	7935
Black/non-Hispanic	<65	.07	2562	2659	.21	2562	2959	5572
	65+	.05	2562	2949	.12	2562	2949	7935
White/other	<65	.31	2659	8235	.25	2863	8235	8235
	65+	.21	2863	8235	.36	2659	5193	8235
Total		.46	2562	8235	.46	2562	8215	8235

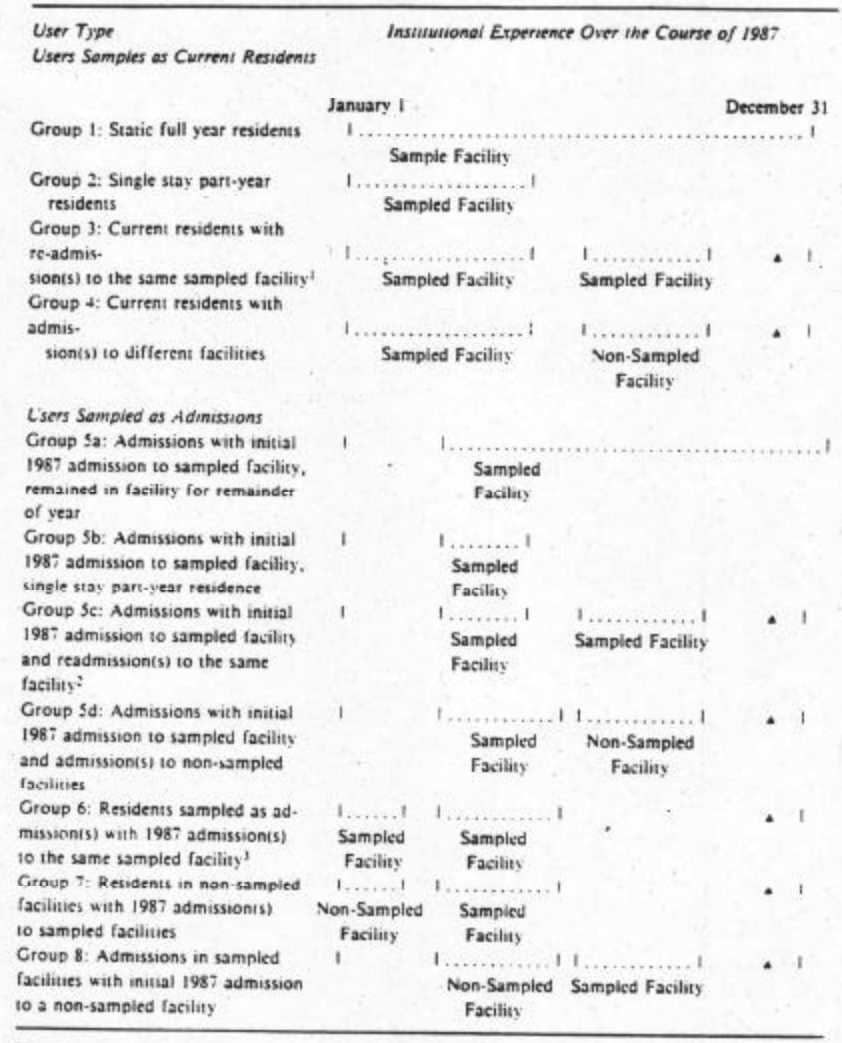
* Mean of standard error ratios for address sample design relative to potential Round 1 sample from household sample design.
Source: Agency for Health Care Policy and Research. National Medical Expenditure Survey, United States, 1987.

Table 3. NMES Household Survey field results at the RU level, round 1

Screener disposition	Number of addresses targeted for round 1	Round 1 responding	Round 1 refusal	Round 1 other non-response	Movers	Other ineligible	Vacant, not a dwelling unit	Total	Overall response rate
Responding	15,130	14,060	1,234	296	847[a]	253[a]	722[a]	17,412	
Response rate	0.912	0.902	0.079	0.019					0.821[b]
Vacant	1,464	408	47	24	0[a]	15[a]	1,016[a]	1,510	
Response rate	--	0.852	0.098	0.050					
Nonresponding	1,021	372	470	86	2[a]	11[a]	111[a]	1,052	
Response rate	0.081	0.401	0.506	0.093					0.033
Nonresponding not fielded	234								
Response rate	0.007	–	–	–	–	–	–	–	0.000
Total		14,840	1,751	406	849	279	1,849	19,974	0.854

[a]Ineligible for round 1.
[b]Combined round 1 response rate for responding and vacant addresses based on screener interview.
Source: Agency for Health Care Policy and Research. National Medical Expenditure Survey.

Figure 2.
Graphic Representation of Institutional User Population in the 1987 National Medical Expenditure Survey



[1]Sampled persons meeting the criterion for User Type 3 and User Type 4 were classified as User Type 4.
[2]Sampled persons meeting the criterion for User Type 5c and User Type 5d were classified as User Type 5d.
[3]Sampled persons meeting criterion for User Type 6 were classified as ineligible.
▲ Potential additional admissions.

Table 4
Distribution of the Institutional User Population, by User and Facility Type, Unweighted (National Medical Expenditure Survey – Institutional Population Component: United States, 1987)

User type	Nursing and personal care homes		Facilities for the mentally retarded	
	Number	Percent	Number	Percent
Users Sampled as Current Residents				
Group 1: Static full year residents	2,586	80.6	3,089	87.8
Group 2: Single stay part-year residents	150	4.7	73	2.1
Group 3: Current residents with re-admission(s) to the same sampled facility	269	8.4	42	1.2
Group 4: Current residents with admission(s) to different facilities	204	6.3	316	8.9
Total current residents	3,209	100.0	3,520	100.0
Users Sampled as Admissions				
Group 5a: Admissions with initial 1987 admission to sampled facility, remained in facility for remainder of year	1,106	46.6	283	35.2
Group 5b: Admissions with initial 1987 admission to sampled facility, single stay part-year residence	432	18.2	68	8.5
Group 5c: Admissions with initial 1987 admission to sampled facility and readmission(s) to the same facility	159	6.7	21	2.6
Group 5d: Admissions with initial 1987 admission to sampled facility and admission(s) to non-sampled facilities	166	6.9	28	3.5
Group 6: Residents sampled as admission(s) with 1987 admission(s) to the same sampled facility	448[a]	0.0[a]	76[a]	0.0[a]
Group 7: Residents in non-sampled facilities with 1987 admission(s) to sampled facilities	374	15.7	371	46.2
Group 8: Admissions in sampled facilities with initial 1987 admission to a non-sampled facility	139	5.9	32	4.0
Total sampled as admissions	2,376[a]	100.0[a]	803[a]	100.0[a]
Overall Total	5,585[a]		4,323[a]	

[a]Users classified into Group 6 were ineligible for data collection and were thus excluded from the totals and percent distributions.
Source: Agency for Health Care Policy and Research.

Council of Professional Associations on Federal Statistics

SEMINAR ON NEW DIRECTIONS IN STATISTICAL METHODOLOGY
Session on Longitudinal Surveys

Methodological Issues Encountered in Following a Cohort of Eighth Graders

Steven J. Ingels, NORC; Jeffrey A. Owings, NCES

Bethesda, Maryland
May 26, 1994

This paper is based on the National Education Longitudinal Study of 1988 (NELS:88). NELS:88 is a longitudinal micro-database sponsored by the National Center for Education Statistics, U.S. Department of Education. The ideas and opinions expressed in this paper are those of the authors and do not necessarily represent positions or conclusions endorsed by NCES.

Introduction. Longitudinal studies typically employ a probability sample of a unit (for example, individuals, institutions [e.g., schools], groups [e.g., families] or inanimate objects [e.g., dwelling units]) that is drawn at one point in time, then repeatedly observed, so that change in units can be measured over time. Longitudinal designs provide a powerful vehicle for reliably[1] measuring individual-level change and development as well as for describing the dynamics of change and the processes that are associated with it. At the same time, longitudinal studies have both inherent and potential limitations (see, for example, Pearson 1989 for a discussion of advantages and disadvantages of longitudinal surveys).

Some of the most important of these limitations can in large measures be overcome if the study is properly executed or if potential limitations are explicitly addressed in the design. This is the case, in particular, for three specific threats to longitudinal sample representativeness.[2]

(1) *Undercoverage*. Potentially biasing undercoverage[3] may arise from any of several sources. It may arise from deliberate or inadvertent exclusion of part of the baseline "target" sample, or may arise if baseline nonrespondents are not pursued in subsequent waves. Undercoverage problems may also arise if the eligibility of ineligible baseline students whose eligibility status is subject to change is not reassessed in succeeding rounds.

(2) *Need for Freshening*. The unit or cohort being studied may become less statistically representative of the target population (or less policy-relevant) over time. For example, a sample of individuals in a given geographical area may become less representative of that area as sample members disperse and other individuals move into the area. A sample of eighth graders two years later is not fully representative of the nation's tenth graders at the second point in time.

(3) *Attrition and Nonresponse*. Sample attrition poses substantial risks for a longitudinal study's representativeness. This danger can be overcome if high response rates are maintained across all rounds, and may be partially compensated for in weighting.

In this paper, the National Education Longitudinal Study of 1988 (NELS:88) serves as an example of how these three specific problems of representativeness in a longitudinal study may be approached and overcome. We more briefly comment on some additional potential sources of survey error.

1. Description of NELS:88

As a point of entry into our topic, we briefly describe NELS:88 by summarizing its goals, surveys administered, response rates, and analysis potential.

1.1 *Goals of NELS:88*

Beginning in 1988 with a cohort of 26,432 eighth graders attending 1,052 public and private schools across the nation, NELS:88 was designed to provide longitudinal data about critical transitions experienced by students as they leave eighth grade school settings, progress through high school (or drop out), enter and leave postsecondary institutions, and enter the work force. The 1988 eighth grade cohort has been followed at two-year intervals (specifically, first follow-up – 1990; second follow-up – 1992) with a third follow-up currently (spring 1994) underway.

Major features of NELS:88 include:

- the integration of student, dropout, parent, teacher, school administrator and school records (transcript) surveys;
- the initial concentration on an eighth grade student cohort with follow-ups at two year intervals;

- the inclusion of supplementary components to support analyses of geographically or demographically distinct subgroups (for example, selected state supplements; oversamples of Asians and Hispanics, and of students in private schools); and

- the design linkages to previous longitudinal studies (High School and Beyond [HS&B], the National Longitudinal Study of the High School Class of 1972 [NLS-1972]) and other current studies (for example, the National Assessment of Educational Progress [NAEP] testing program and high school transcript data collections).

The longitudinal design of NELS:88 permits the examination of change in young people's lives and the role of schools, teachers, community, and family in promoting growth and positive life outcomes. In particular, data from NELS:88 can be used to investigate issues in the context of the family, community, school, and classroom including:

- Students' academic growth over time;

- The transition from eighth grade to high school and the transition from high school to the labor market or postsecondary education;

- The process of dropping out of school, as it occurs from the end of eighth grade on;

- The role of schools in helping the disadvantaged;

- The school experiences and academic performance of language minority students;

- The influence of ability grouping, program type, and coursetaking patterns on future educational persistence and achievement; and

- The features of effective schools.

1.2 Surveys administered

NELS:88 components, by wave, are summarized in Figure 1.

Figure 1: Base Year Through Fourth Follow-Up – NELS:88 Components

BASE YEAR	FIRST FOLLOW-UP	SECOND FOLLOW-UP	THIRD FOLLOW-UP	FOURTH FOLLOW-UP
spring term 1988 GRADE 8 **Students:** Questionnaire, Tests* **Parents:** Questionnaire **Principals:** Questionnaire **Two Teachers per student:** (taken from English, social studies, mathematics, science)	spring term 1990 MODAL GRADE = SOPHOMORE **Dropouts, Students:** Questionnaire, Tests **Principals:** Questionnaire **Two Teachers per student:** (taken from English, social studie mathematics, or science)	spring term 1992 MODAL GRADE = SENIOR **Dropouts, Students:** Questionnaire, Tests, H.S. Transcripts **Parents:** Questionnaire **Principals:** Questionnaire **One Teacher per student:** (taken from mathematics or science)	spring 1994 MODE = H.S. + 2 YEARS **All Individuals:** Questionnaire	spring 1997 HS + 5 YEARS **All Individuals:** Questionnaire

* Reading, social studies, math and science tests are administered in the three in-school rounds.

1.3 *Analysis potential - longitudinal vs. cross-sectional applications.*

Analytic Levels. The NELS:88 design enables researchers to conduct analyses on three principal levels: (1) within-wave (or cross-sectional) analysis at a single time point, (2) cross-cohort analysis (by comparing cross-sectional NELS:88 findings to those of comparable populations studied earlier in HS&B and NLS-72) and (3) cross-wave (or longitudinal[4]) analysis.

The first analytic level within NELS:88 is cross-sectional. By beginning with a cross-section of 1988 eighth graders, following a substantial subsample of these students at two-year intervals, and freshening the 1990 and 1992 samples to obtain representative national cross-sections of tenth and twelfth graders, the study also provides a statistical profile of America's eighth graders, high school sophomores, and high school seniors.

A second analytic level extends representative cross-sections to intercohort comparisons. NELS:88 provides researchers with data for drawing comparisons with previous NCES longitudinal studies. After the release of NELS:88 first follow-up data, researchers were able to conduct trend analyses with the 1980 sophomore cohort of HS&B. With completion of the NELS:88 second follow-up, comparisons may be made among NELS:88, HS&B, and NLS-72 senior cohorts. To facilitate cross-cohort comparisons, some of the questionnaire items used in the NLS-72 and HS&B high school surveys were repeated in NELS:88, and data processing and file conventions were kept consistent, to the maximum possible extent, with HS&B and NLS-72.[5]

The third analytic level is longitudinal, and utilizes repeated measurements on the same individuals over time. However, because NELS:88 comprises three nationally representative grade- and year-defined cross-sections, it supports multiple panels:

1988	eighth graders two, four, six, and nine years later
1990	sophomores two, four, and seven years later
1992	seniors two and five years later[6]

Change Analysis. Cross-sectional analysis provides a snapshot at a single point in time. Repeated cross-sectional analysis, and longitudinal analysis, permit the measurement of change over time. Change (and stability) over time can be measured at the group or individual level:

(1) *At the group level*, change can be measured across the successive cross-sections--eighth graders in 1988, sophomores in 1990, and seniors in 1992. In the same way, multicohort assessments such as NAEP can estimate overall and subgroup gains in specific subject matter proficiency across selected points in the school career (e.g., between fourth, eighth, and twelfth grade). In addition, NELS:88 and comparable studies (e.g., NLS-72 and HS&B) can be analyzed as repeated cross-sections (e.g. of seniors in 1972, 1980/82, and 1992) to measure trends. A cross-sectional time-series such as NAEP also measures trends (e.g. in math achievement for 17 year olds from 1973 to 1990 for the nation and subgroups).

A principal weakness of change measurement at the group level--whether one is looking at rolling (e.g., eighth graders in 1988, sophomores in 1990, seniors in 1992) or repeated (e.g., eighth graders in 1988, 1990, and 1992) static cross-sections[7] is that it sometimes masks individual change; high levels of individual change are not incompatible with stability at the aggregate level. Thus, for example, looking at the proportion of 1988 eighth graders in 1988 who were out of school in 1990 (6.8%) and comparing this to the proportion out of school in 1992 (11.6%) masks the cumulative number of individuals who were

1990 or 1992 spring term dropouts, since some 1990 dropouts had returned to school by 1992.[8] A *locus classicus* of this phenomenon is found in studies of poverty and welfare recipience. While the proportion of adolescent mothers receiving AFDC over time is relatively constant, the AFDC population is not. Mobility onto, and off, the AFDC's rolls is demonstrated by longitudinal data provided by the NLSY, but would not be apparent from repeated cross-sectional results. Likewise, PSID data show that while poverty rates may be roughly stable over time, poverty spells for individuals and households tend to be relatively brief.[9]

(2) Change can also be analyzed *at the individual level* over time. The latter possibility–true longitudinal measurement–represents, for most purposes, the unique strength of the NELS:88 design. Following individual educational histories generally provides the best basis for drawing causal inferences about educational processes and their effects. Two broad kinds of analysis scenarios are possible. Longitudinal analysis can involve repeated measures of the same outcome–for example, test data can be used to measure growth in academic achievement over time. Or longitudinal analysis can show how conditions at an earlier time point are predictive of outcomes at a later time point. For example, one might examine how eighth graders with single or clustered "risk factors" (for example, such status risk factors as coming from a low-income home, having parents who did not finish high school, and so on; or such behavioral risk factors as cutting classes, lack of participation in extracurricular activities, and so on) fared two years later (for example, what proportion had dropped out, repeated a grade, and so on).

While longitudinal studies are prospective, in that they offer the opportunity to record new events, longitudinal analysis may be either retrospective or prospective. In NELS:88, priority in the baseline was given to questions predictive of future behavior. However, while questions that asked for *reasons* for past behavior were deliberately avoided, some retrospective questions were posed, when their focus was on simple descriptions of salient past events. For example, parents were asked whether their eighth grader had attended a Head Start program[10] or kindergarten or preschool, whether other of their children (respondent's elder siblings) had dropped out of school, and so on.

2. Sample Representativeness

This section discusses three key issues. *First*, eligibility and exclusion rules, particularly as applied in the NELS:88 base year, and the measures taken in later rounds of the study to deal with the potential for undercoverage biases that might result from these exclusions. *Second*, the need for sample freshening to ensure representative sophomore and senior cohorts in 1990 and 1992, and the procedures undertaken to bring that sample freshening about. And *third*, attempts to minimize sample attrition and nonresponse error.

2.1 Eligibility: Excluded Students and Undercoverage Bias.

In the base year of NELS:88, students were sampled through a two-stage process. First, stratified random sampling and school contacting resulted in the identification of the school sample; second, students were randomly selected (with oversampling of Hispanics and Asians) from within cooperating schools.

The target population for the base year comprised all public and private schools containing eighth grades in the fifty states and the District of Columbia. Excluded from the NELS:88 school sample are Bureau of Indian Affairs (BIA) schools, special education schools for the handicapped, area vocational schools that do not enroll students directly, and schools for dependents of U.S. personnel overseas.[11] The student population excludes students with severe mental handicaps, students whose command of

the English language was not sufficient for understanding the survey materials (especially the cognitive tests), and students with physical or emotional problems that would make it unduly difficult for them to participate in the survey. This chapter discusses (1) the consequences of student exclusion for the research design and results, and (2) the special measures that have been undertaken in NELS:88 to compensate or correct for the effects of exclusion. Before either of these two topics is pursued in detail, however, it will be desirable to say more about student exclusion in the NELS:88 base year–the 1987-88 school year during which the eighth grade cohort was selected and surveyed.

To better understand how excluding students with mental handicaps, language barriers, and severe physical and emotional problems affects population inferences, data were obtained on the numbers of students excluded as a result of these restrictions.

Seven ineligibility codes defining categories of excluded students were employed at the time of student sample selection:

A - attended sampled school only on a part-time basis, primary enrollment at another school.

B - physical disability precluded student from filling out questionnaires and taking tests.

C - mental disability precluded student from filling out questionnaires and taking tests.

D - dropout: absent or truant for 20 consecutive days, and was not expected to return to school.

E - did not have English as the mother tongue AND had insufficient command of English to complete the NELS:88 questionnaires and tests.

F - transferred out of the school since roster was compiled.

G - was deceased.

Before sampling, school coordinators–members of the school staff, typically an assistant principal or guidance counselor who acted as liaison between the school and the study–were asked to examine the school sampling roster and annotate each excluded student's entry by assigning one of the exclusion codes. Because eligibility decisions were to be made on an individual basis, special education and Limited English Proficiency (LEP) students were not to be excluded categorically. Rather, each student's case was to be reviewed to determine the extent of limitation in relation to the prospect for meaningful survey participation. Each individual student, including LEPs and physically or mentally handicapped students, was to be designated eligible for the survey if school staff deemed the student capable of completing the NELS:88 instruments, and excluded if school staff judged the student to be incapable of doing so. School coordinators were told that when there was doubt, they should consider the student capable of participation in the survey. Exclusion of students after sampling ("post-roster ineligibles") occurred either during the sample update just prior to survey day, or on survey day itself. Such exclusion after sampling normally occurred because of a change in student status (for example, transfer, death). However, in very rare instances such exclusions reflected belated recognition of a student's pre-existing ineligibility–that is, if an annotation error was made and an ineligible student selected for the sample in consequence of such an error, ineligibility became apparent later in the survey, whereupon the student was excluded.

Excluded students were divided into those who were full-time students at the school (categories B, C, and E) and those who were not (categories A, D, F, & G). Our main concern here is with students who were full-time students at the school but who were excluded from the sample. Excluding these students will affect estimates made from the sample.

Students in categories A (n=329), D (n=733), F (n=3,325), and G (n=6) were either not at the school or were present only part time (with primary registration at another school, hence a chance of selection into NELS:88 at another school). Thus excluding students in these categories has no implications for making estimates to the population of eighth grade students.

It should be noted that students in category F, those who had transferred out of the sampled school, had some chance of being selected into the sample if they transferred into another NELS:88 sampled school just as transfers into NELS:88 schools from non-NELS:88 schools had a chance of selection at the time of the sample update. The sampling of transfer-in students associated with the sample update allowed NORC to represent transfer students in the NELS:88 sample.

The total eighth grade enrollment for the NELS:88 sample of schools was 202,996. Of these students, 10,853 were excluded owing to limitations in their language proficiency or to mental or physical disabilities. Thus 5.37 percent of the potential student sample (the students enrolled in the eighth grade in the 1,052 NELS:88 schools from which usable student data were obtained) were excluded. Less than one half of one percent of the potential sample was excluded for reasons of physical or emotional disability (.41 percent), but 3.04 percent was excluded for reasons of mental disability, and 1.90 percent because of limitations in English proficiency.

Put another way, of the 10,853 excluded students, about 57 percent were excluded for mental disability, about 35 percent owing to language problems, and less than 8 percent because of physical or emotional disabilities. Because current characteristics and probable future educational outcomes for these groups may depart from the national norm, the exclusion factor should be taken into consideration in generalizing from the NELS:88 sample to eighth graders in the nation as a whole. This implication for estimation carries to future waves. For example, if the overall propensity to drop out between the eighth and tenth grades is twice as high for excluded students as for non-excluded students, the dropout figures derivable from the NELS:88 first follow-up (1990) study would underestimate early dropouts by about ten percent. (In point of fact, the 1988-90 status dropout rate derivable from the eligible NELS:88 sample representing about 94.6 percent of the cohort is between 6.0 and 6.1 percent, and from the expanded--eligible + ineligible--1988 sample representing [virtually] 100 percent of the cohort, 6.8 percent.)

Undercoverage of course affects the power of a study both to produce national estimates, and, yet more dramatically, to produce estimates for the particular group that is not fully covered.[12] Undercoverage, moreover, poses some special difficulties for the representativeness of a multi-cohort longitudinal study such as NELS:88.

In a school-based longitudinal survey such as NELS:88, baseline excluded students affect the representativeness of freshened grade cohorts in future waves. To achieve a thoroughly representative tenth grade (1990) and twelfth grade (1992) sample comparable to the High School and Beyond 1980 sophomore cohort (or, for 1992, the HS&B 1980 senior cohort and the base year of NLS-72), the NELS:88 follow-up samples must approximate those which would have come into being had a new baseline sample independently been drawn at either of the later

time points. In 1990 (and 1992) one must therefore freshen, to give "out of sequence" students (for example, in 1990, those tenth graders who were not in eighth grade in the spring of 1988) a chance of selection into the study. One must also accommodate excluded students whose eligibility status has changed, for they too (with the exception of those who fell out of sequence in the progression through grades) would potentially have been selected had a sample been independently drawn two years later, and must have a chance of selection if the representativeness and cross-cohort comparability of the follow-up sample is to be maintained. Thus, for example, if a base year student excluded because of a language barrier achieves the level of proficiency in English that is required for completing the NELS:88 instruments in 1990 or 1992, that student should have some chance of re-entering the sample.

A substantial subsample of the base year ineligibles was, accordingly, followed in 1990 and 1992, to reassess eligibility status and gather information about excluded students' demographic characteristics, educational paths, and life outcomes. Data on persistence in school to be obtained from this subsample has been used to derive an adjustment factor for national estimates of the eighth grade cohort's dropout rates between spring of 1988 and spring of 1990, and from 1988 and 1990 to 1992.

The base year ineligibles study largely compensates for population undercoverage. Small populations who remain outside the baseline sampling frame include students who are educated at home or in private tutorial settings, those who are in excluded categories of schools[13] and those who have dropped out of school before reaching the eighth grade.

Table 1 shows that by 1992, a substantial portion of the sample of base year ineligible students had been reclassified as eligible. Excluded students who were later classified as eligible were included in NELS:88 follow-up surveys.

Reclassifications reported in Table 1 reflect multiple phenomena. In some cases--and presumably this is particularly the case for the language exclusions--reclassification reflects change in the eligibility status of the sample member over time. In other cases, change represents the unreliability of exclusion judgments, particularly for exclusion reasons that are more open to interpretation (e.g., mental as opposed to physical handicaps) or that apply to individuals at the margin of the classification--different individuals were asked to assess eligibility at different points in time. Finally, some of the change registered in Table 1 reflects the fact that in the follow-ups we provided more detailed interpretation for the guidelines, so that the validity of exclusion judgments would be enhanced. All in all, however, if any individuals in the target population are to be subject to exclusion from the baseline of a longitudinal study, it is of some importance to reassess their eligibility over time, particularly, in a school-based survey, if the panel is to represent additional grade cohorts.

Table 1: 1992 Status Ns of 1988 Excluded Students

1988 reason for exclusion:	ELIG.	INELIG.	OUT OF SCOPE	N.A.	SAMPLING ERROR
language	125	22	25	30	23
physical	13	9	0	1	1
mental	166	140	5	25	16
unknown	30	15	2	10	16
TOTAL	334	186	32	66	56

* N.A. = status not ascertained.

2.2 Representativeness and New Grade Cohorts: Sample Freshening.

Pearson (1989) notes that a potential limitation of longitudinal samples is that they may provide estimates of the population from which they were originally drawn, but not of the current population. It is of interest to follow a sample of 1988 eighth graders. Nevertheless, an eighth grade panel two years later will not by itself provide a representative sample of the nation's high school sophomores, nor four years later a representative sample of seniors. Representative sophomore and senior samples are analytically desirable at all three levels of NELS:88 analysis *First*, it is desirable to be able to make cross-sectional generalizations about the nation's sophomores in 1990 and seniors in 1992. *Second*, it is desirable to be able to make intercohort comparisons between HS&B 1980 sophomores and 1990 NELS:88 sophomores; between NLS-72 (1972) and HS&B (1980) seniors and NELS:88 (1992) seniors; and between the transcript records of HS&B (1982), NAEP (1987 and 1990), and NELS:88 (1992) seniors. *Third*, it is desirable to be able to conduct longitudinal analyses of 1990 sophomores two, four, and more years later, and of 1992 seniors two and more years later.

Hence a major sampling objective of NELS:88 was to create a valid probability sample of students enrolled in tenth grade in the spring term of the 1989-1990 school year and of students enrolled in the twelfth grade in the spring term of the 1991-92 school year. This goal was achieved by a process we have termed "freshening." The 1990 freshening procedure was carried out in four steps:

1. For each school that contained at least one base year 10th grade student selected for interview in 1990, a complete alphabetical roster of all 10th grade students was obtained.

2. For each base year sample member, we examined the next student on the list; if the base year student was the last one listed on the roster, we examined the first student on the roster (that is, the roster was "circularized").

3. If the student who was examined was enrolled in the 8th grade in the U.S. in 1988, then the freshening process terminated. If the designated student was not enrolled in the 8th grade in the U.S. in 1988, then that student was selected into the freshened sample.

4. Whenever a student was added to the freshened sample in step 3, the next student on the roster was examined and step 3 was repeated. The sequence of steps 3 and 4 was repeated (adding more students to the freshened sample) until a student who was in the 8th grade in the U.S. in 1988 was reached on the roster.

At a given first follow-up school, the freshening process could yield zero, one, or more than one new sample member. Altogether, 1,229 new students were added to the tenth grade sample—on average, just less than one student per school.[14] This procedure was repeated in 1992, to generate a probability sample of the nation's high school seniors.

This freshening procedure is an essentially unbiased method[15] for producing a probability sample of students who were enrolled in the tenth grade in 1990 (or twelfth grade in 1992) but were not enrolled in the eighth grade in the U.S. in 1988. There is a very small bias introduced by the omission of eligible tenth (or twelfth) graders attending schools that included *no* students who were eighth graders in 1988. There is an additional small bias introduced by not freshening on the members of the sample of base year ineligibles. All other 1990 sophomores (or 1992 seniors) who qualify for the freshening sample have some chance of selection. This is because every student who was in the tenth grade in 1990 (or twelfth grade in 1992) but not in the eighth grade in 1988 is linked to exactly one student who was a 1988 eighth grader—this is the 1988 eighth grader who would immediately precede the candidate for the freshening sample on a circularized, alphabetical roster of tenth graders at the school. Because each 1988 eighth grader had a calculable, non-zero probability of selection into the base year and first follow-up samples, we can calculate the selection probabilities for all students eligible for the freshening sample. Thus, the freshening procedure produces a student sample that meets the criterion for a probability sample.

The NELS:88 school sample in 1990 and 1992—the schools to which 1988 eighth graders matriculated—was of course not a nationally representative sample of schools. However, for a select subset of schools, in order to provide a basis for studying school effects, feeder pattern information was collected so that tenth grade school selection probabilities could be approximated, and student samples augmented to make them robust and representative of the school's tenth grade class.[16]

2.3 Nonresponse Error as a Potential Source of Bias: Measures to Maximize Response Rates.

Cumulative nonresponse poses a special threat to longitudinal studies. Some individuals are missed in the baseline measurement, and may enter the study late. Other individuals may be lost, through mobility and the inability to locate them at a later date, or may cease to participate in the study. Still others may participate in the baseline, become temporarily out of scope by leaving the country or become nonrespondents by refusing to participate in the initial follow-up, then re-enter the study in a later follow-up. A longitudinal study must maximize the number of individuals who have data at all data points. Although weighting may help to adjust for nonresponse, the representativeness of the panel depends, in the final analysis, on maintaining high participation rates.

NELS:88 Response Rates. High response rates have been achieved by the study. In the NELS:88 base year (1988) 93.1 percent of selected eighth graders participated. In the NELS:88 first follow-up (1990), 93.9 percent of student and dropout sample members (19,264 of

20,524) took part. In the second follow-up, 90.7 percent of student and dropout sample members took part.

However, from the point of view of longitudinal analysis, a more critical statistic is the proportion of the sample with data at all time points (or, the proportion of baseline participants with data for all follow-ups). Of the 18,261 base year participants retained in the first follow-up, 17,424–or 95.4 percent–were successfully resurveyed. From this base of eighth grade cohort members with both (1988 and 1990) data points, 95.1 percent were resurveyed in the second follow-up.

Table 2A shows overall and subgroup results for the base year-first follow-up respondents for whom a reinterview was attempted in 1992. While, as noted above, around 95 percent were successfully resurveyed (that is, completed a student or dropout questionnaire) in 1992 and thus have data for all three waves, far fewer (72 percent) completed the cognitive test in all three rounds. Table 2B depicts the across-round questionnaire completion status of base year-first follow-up participants who were second follow-up *students*, and the likelihood that school contextual data was available for them for all three rounds. These tables show that completion rates were very similar across different school control types, urbanicity, region, and high and low minority enrollment, and that similar response rates were obtained for members of different racial and ethnic groups.

However, even with these high rates of success in baseline and follow-up data collection, the proportion of the 1988 eighth grade cohort in 1992 with all three data points drops to 84 percent (16,489 of 19,645) when all students missing one or more data point owing to base year, first or second follow-up nonresponse or any other source of sample attrition–being deceased, sample members who suffered grave impairments in the course of the study that did not permit them to be surveyed, individuals out of scope for either follow-up round by virtue of being outside the country–are factored in.

Overall, then, NELS:88 has achieved reasonably high student panel response rates. In addition, final weights have been adjusted for nonresponse, using nonresponse adjustment cells based upon combinations of classification values reflecting race, gender, and data collection status (e.g., dropout; in school in expected grade; in school in another grade; and so on).[17]

Means of Achieving High Response Rates. The means by which these high response rates were achieved may be concisely summarized. Most individuals changed schools, and many changed home addresses, between the base year and the follow-up surveys. About 99 percent of students were successfully traced between the base year and first follow-up, whereupon clusters of students were subsampled to reduce, for cost reasons, the number of high schools to be included in the study. The ability to successfully trace individuals was based upon extensive locating information collected in the base year from both students and parents. This locating information included name, address and telephone number for the student, each parent, and the family's closest relative or friend who did not live in the household. Eighth grade students were also asked to indicate what school they expected to be attending two years later. Tracing was carried out at two levels: first, it was ascertained if the sample member was at the expected school. If not, household information was used to locate the individual. In order to find base year nonrespondents (about 7 percent of the sample did not complete a 1988 student questionnaire and hence did not provide locating information), in addition to conventional survey locating sources, information about the schools matriculated to by the eighth grader's classmates was also utilized. Tracing procedures were repeated in the second follow-up, though between tenth and twelfth grade there is less dispersion to new schools and it was not necessary to further subsample students.

TABLE 2A

NELS:88 Second Follow-Up student survey results for Base Year– First Follow-Up panel participants

	Student/Dropout questionnaire (BY, F1 and F2) Completion rates		Student/Dropout cognitive test[a] (BY, F1 and F2) Completion rates	
	Weighted	Unweighted	Weighted	Unweighted
Total				
Participated	94.7	95.1	69.6	72.2
Selected	16,489[b]		11,902	
School type[c]	17,337		16,489	
Public				
Catholic	94.3	94.7	69.0	71.4
Other private	97.9	97.0	74.1	78.6
Urbanicity[c]	97.4	97.0	73.0	73.7
Urban				
Suburban	93.5	95.1	64.3	69.5
Rural	95.5	95.3	69.1	70.1
Region[c]	94.8	94.9	74.6	77.2
Northeast				
South	94.8	95.1	70.3	71.3
Midwest	94.1	94.5	68.2	73.1
West	95.7	96.0	74.9	76.4
Ethnicity	94.6	95.1	63.7	65.7
Asian/PI				
Hispanic	93.3	95.0	71.5	71.9
Black	93.1	94.4	63.9	65.5
White	92.4	92.6	59.6	67.0
Am. Indian	95.5	95.7	72.1	74.2
Refused/Missing[d]	94.1	91.3	64.8	64.0
Minority schools[c]	81.1	75.0	38.3	55.6
Schools with more than 19% minority students	92.2	93.5	55.1	59.3
Schools with less than 19% minority students	95.0	95.3	71.0	73.5

[a] Cognitive test coverage rate for each sample member who has completed a BY student questionnaire, F1 and F2 student/dropout questionnaire.

[b] Sample members who participated in the BY, F1 and F2.

[c] Refers to 8th grade schools.

[d] Refused/Missing refers only to the status of a sample member's ethnicity. It does not refer to student/dropout nonparticipants.

TABLE 2B

NELS:88 Second Follow-Up data collection results for Base Year – First Follow-Up panel participants

	Student questionnaire (BY, F1 and F2) Completion rates		School questionnaire[a] (BY, F1 and F2) Completion rates	
	Weighted	Unweighted	Weighted	Unweighted
Total	95.7	96.1	95.5	95.6
Participated	14,674[b]		13,182	
Selected	15,269		13,783	
School type[c]				
Public	95.4	95.8	95.8	95.7
Catholic	98.2	97.3	94.3	94.8
Other private	97.5	97.1	93.5	95.8
Urbanicity[c]				
Urban	94.4	96.4	93.7	94.7
Suburban	96.2	96.1	94.4	94.3
Rural	95.8	95.9	98.4	98.2
Region[c]				
Northeast	95.2	95.5	94.9	94.6
South	95.8	96.2	95.6	95.9
Midwest	96.2	96.5	97.3	97.8
West	95.5	96.0	93.1	93.2
Ethnicity				
Asian/PI	94.9	95.8	90.2	93.9
Hispanic	94.2	95.8	89.8	91.3
Black	94.3	95.0	95.1	95.3
White	96.2	96.4	96.5	96.5
Am. Indian	93.8	90.9	97.6	97.3
Refused/Missing[d]	74.2	72.7	100.0	100.0
Minority schools[c]				
Schools with more than 19% minority students	92.5	96.3	90.7	90.0
Schools with less than 19% minority students	96.0	94.4	96.0	96.2

[a] School questionnaire coverage rate for each student who completed a BY, F1, and F2 student questionnaire.

[b] Panel *students* only.

[c] Refers to 8th grade schools.

[d] Refused/Missing refers only to the status of a sample member's ethnicity. It does not refer to student nonparticipants.

In order to survey students, contractor (NORC) staff administered the survey forms at a date agreeable to the school. Make-up sessions were conducted for students who missed the initial survey session. Dropouts and chronic absentees were pursued outside school. Such individuals were invited to group sessions and provided reimbursement for their travel expenses, or were interviewed in their households, over the telephone or in person.

In rare instances, NELS:88 has made use of respondent fees. For example, some dropouts received a monetary incentive, as did some high burden teachers (teachers who had to rate an unusually high number of NELS:88 students such that their burden of questionnaire completion might be two hours or more). School coordinators were given a modest honorarium (normally $25) for assisting with survey activities (for example, supplying annotated rosters, arranging space, and so on), but neither schools nor students were ever paid for their participation

3. Other Sources of Survey Error

When all is said and done, it is the total variable error and bias of a survey estimate that is critical (see Kish, 1965; Andersen, Kasper, and Frankel, 1979; Groves, 1989). From the point of view of total survey error, our discussion thus far is incomplete. It may be useful to identify additional sources of survey error, though space limitations do not permit us to address them.

There are various "repeated measurement" problems in longitudinal surveys. One of these problems is that of panel effects.[12] We do not believe that problems associated with repeated measurements (such as remembering past responses to individual items) are likely to be a difficulty, both because of the sheer number of test and questionnaire items asked, and the two year intervals between data collections. However, participation in a longitudinal study in theory may influence the survey member's subsequent behavior or attitudes.

There are many sources of measurement error. The validity of responses to the NELS:88 eighth grade questionnaire items has been examined in Kaufman, Rasinski, Lee and West (1991), which compares parent and student reports. Transcript and student reports were compared for the HS&B data by Fetters, Stowe and Owings (1984). Psychometric issues in the base year tests are addressed in Rock and Pollack (1991) and in a forthcoming second follow-up psychometric report.

Our earlier discussion dealt with unit nonresponse as a problem of maintaining individual participation across rounds. However, school nonresponse in the base year, and item nonresponse across the survey instruments, also are important nonresponse issues. To the extent that students at noncooperating base year schools may have differed from students at cooperating schools, student level bias is introduced that persists through subsequent waves of observation. Base year school nonresponse is documented and analyzed in the *NELS:88 Base Year Sample Design Report*.

Item nonresponse rates and patterns are documented in the various NELS:88 user's manuals. In general, missing data have not been imputed in the NELS:88 dataset. Although item response rates in NELS:88 are generally high, item nonresponse propensities vary with student characteristics (e.g., race, gender, test quartile), and hence may be a source of bias.

Finally, our discussion has not dealt with the important consideration of sampling error. Design effects for NELS:88 are documented in the various user's manuals. In this respect, dispersion of the student sample after eighth grade has been both a blessing and a curse for NELS:88. The high costs of following dispersed students required that we subsample students in the first follow-up; subsampling increases design effects. At the same time, the general tendency in a longitudinal study is for design effects to decrease over time, as dispersion reduces the original clustering.

Bibliography

Andersen, R., Kasper, J., and Frankel, M.R. (1979). ***Total Survey Error.*** San Francisco: Jossey-Bass.

Babbie, E.R. (1973). ***Survey Research Methods.*** Belmont, CA: Wadsworth.

Bradburn, N.M., Rips, L.J., and Shevell, S.K. (1987). "Answering Autobiographical Questions: The impact of Memory and Inference on Surveys." ***Science***, 239:157-161.

CBO (Congress of the United States, Congressional Budget Office). (1990). ***Services of Support for Adolescent Mothers.*** Washington, D.C.: Author.

Duncan, G.J., Hill, M.S., and Hoffman, S.D. (1988). "Welfare Dependence Within and Across Generations." ***Science***. 239:467-471.

Fetters, W.B, Stowe, P., and Owings, J.A. (1984). ***Quality of Responses of High School Students to Questionnaire Items.*** Washington, D.C.: National Center for Education Statistics.

Groves, R.M. (1989). ***Survey Errors and Survey Costs***. New York: Wiley.

Ingels, S.J., et al. ***NELS:88 Data File User's Manuals***. Base Year: Student, Parent, Teacher, School (NCES, 1990); First Follow Up: Student, Dropout, Teacher, School (NCES 1992); Second Follow-Up: Student, Dropout, Parent, Teacher, School, Transcript (NCES 1994).

Kasprzyk, D., Duncan, G., Kalton, G., & Singh, M.P., eds. (1989). ***Panel Surveys***. New York: Wiley.

Kaufman, P., Rasinski, K.A., Lee, R., and West, J. (1991). ***Quality of Responses of Eighth-Grade Students in NELS:88.*** Washington, D.C.: National Center for Education Statistics.

Kish, L. (1965). ***Survey Sampling***. New York: Wiley.

Kish, L. (1987). ***Statistical Design for Research.*** New York: Wiley.

Lee, V.E., and Loeb, S. (1994). "Where Do Head Start Attendees End Up? One Reason Why Preschool Effects Are Not Sustained." Paper presented to the American Educational Research Association, Annual Meeting, New Orleans, Louisiana, April 7, 1994.

McGrew, K.S., Thurlow, M.L., and Spiegel, A.N. (1993). "An Investigation of the Exclusion of Students With Disabilities in National Data Collection Programs." vol.15. ***Educational Evaluation and Policy Analysis.***

Pearson, R. W. (1989). "The Advantages and Disadvantages of Longitudinal Surveys." ***Research in the Sociology of Education and Socialization***. vol. 8.

Rock, D.A. and Pollack, J.M. (1991). ***Psychometric Report for the NELS:88 Base Year Test Battery.*** Washington, D.C.: National Center for Education Statistics.

Rosenzweig, M.R., and Wolpin, K.I. (1993). "Maternal Expectations and Ex Post Rationalizations: The Usefulness of Survey Information on the Wantedness of Children." ***Journal of Human Resources***, vol. XXVIII.

Spencer, B.D., Frankel, M.R., Ingels, S.J., Rasinski, K.A., and Tourangeau, R. (1990). ***NELS:88 Base Year Sample Design Report.*** Washington, D.C.: National Center for Education Statistics.

Spencer, B.D. (1991). "Eligibility/Exclusion Issues in the 1990 Trial State Assessment." In G. Bohrnstedt et al., ***Assessing Student Achievement in the States: Background Studies for the Evaluation of the NAEP Trial State Assessment.*** National Academy of Education. Stanford, California: Stanford University.

Spencer, B.D., and Foran, W. (1991). "Sampling Probabilities for Aggregations, With Application to NELS:88 and Other Education Longitudinal Surveys." ***Journal of Educational Statistics***, 16(1).

END NOTES

1Longitudinal studies are prospective, in that they record new events for *individual* units under observation more or less as they occur. A repeated cross-sectional study can also be prospective, and measure change over time at the *group* level. A single (not repeated) cross-sectional study can measure change in *individual* units over time by assuming a retrospective focus--for example, by relying on individual memories to reconstruct an historical record of events and statuses. While retrospective studies are appropriate for many purposes, when used for other purposes there may be significant reliability problems. For a concise summary of issues concerning the reliability of retrospective reports, see Bradburn, Rips and Shevell (1987). For a useful comparison of prospective and retrospective studies, see Kish (1987) pp. 178-181. For a recent example (*ex post facto* reports of wantedness of children) of an analysis of the degree to which retrospectively-obtained survey data provides unbiased estimates, see Rosenzweig and Wolpin (1993).

2We take a representative sample to be a probability sample drawn, with known selection probabilities for sample units, from the target population.

3Groves (1989, chapter 3) provides a useful discussion of coverage error and its consequences.

4There are many ways to characterize sample designs that measure change over time, and the term "longitudinal" has both strict and looser usages. Kish (1987, Chapter 6) presents a useful typology of designs for covering time spans across populations, and Babbie (1973, pp.62-66) some standard terminology.

5For individuals interested in conducting trend analyses of NLS-72, HS&B and NELS:88 data, further information on content and design similarities and differences between these three studies is presented in the second follow-up student component data file user's manual. Comparison of sophomore cohort dropouts across HS&B and NELS:88 is discussed in the dropout component user's manual, while high school transcript comparisons (HS&B, NAEP 1987, NAEP 1990, NELS:88) are discussed in the transcript user's manual.

6For each cohort, the timing of the last follow-up assumes that the tentatively scheduled date for the fourth follow-up -- 1997 -- will hold.

7Repeated cross-sections compound sampling error. This is the case because a repeated cross-section is drawn two or more times; change measurement must contend with the fact that differences in multiple sample means will in part be a function of the sampling errors associated with each independent sample. In contrast, a longitudinal sample is drawn but once. However, for a freshened cohort study such as NELS:88, some sampling error may be associated with the freshening process. Hence when NELS:88 data are analyzed in the aggregate as a rolling cross-section, some of this advantage of a longitudinal design is lost.

8The 1988-90 dropout rate for the expanded (eligible + ineligible) NELS:88 eighth grade cohort was 6.8 percent for 1988-90. Excluding students who dropped out between 1988 and 1990 (or left the country), the dropout rate between 1990 and 1992 was 7.6 percent. However, the proportion of 1988 eighth graders who were dropouts in the spring of 1992 was 11.6 percent. (Of course, the number of sample members experiencing brief duration dropout spells or dropout *events* is even further undercounted by virtue of using a cohort status [spring to spring across two years] measurement.)

9On NLSY (the BLS National Longitudinal Survey of Youth which began in 1979), see CBO, 1990. On PSID (Panel Study of Income Dynamics, a nationally-representative sample of families, begun in 1968) results, see Duncan, Hill, and Hoffman, 1988.

10Researchers (see Lee and Loeb, 1994) have used the response to this retrospective item in conjunction with NELS:88 measures of school quality to inquire into whether Head Start participants are more likely than their peers to attend lower quality elementary/middle schools, a possibility that could in part explain why academic gains from Head Start may fade out over time.

11For further details of school-level exclusion, see Spencer, Frankel, Ingels, Rasinski, & Tourangeau, 1990, p.10.

12Recent investigations of the extent to which students with disabilities are allowed to participate in major national data collection programs suggest that 40-50 percent of students with disabilities are typically excluded from major assessments, though students with disabilities are included to a greater degree in data collections that do not require the completion of cognitive tests (McGrew, Thurlow, & Spiegel, 1993). Additional numbers of students are excluded from assessments or other state and national education data collection programs owing to language barriers to participation. For a parallel discussion based on the NAEP trial state assessments, see Spencer in Bohrnstedt, ed., 1991.

13According to Office of Special Education figures reported in the *Digest of Education Statistics, 1992*, Table 51, 5.5 percent of special education students receive services in separate schools or residential facilities, while .8 percent are in a homebound or hospital environment. Not all of these individuals are in graded programs. Separate facilities tend in particular to be available for comparatively rare populations such as individuals with severe visual or hearing impairments, and for emotionally disturbed students whose presence might impede regular classroom activities. Most students who are doubly physically disabled by being both deaf and blind are educated in special facilities.

14Some of these freshened students were dropped in the subsampling process either because they themselves were not included in the subsample or because the base year student to whom they were linked was not included. Some 1,043 students selected through the freshening procedure remained in the final first follow-up sample. In the second follow-up (1992), 244 students were added through freshening.

15See Kish (1965) for a discussion of the half-open interval procedure that underpins this approach.

16A strategy for estimating a school's selection probabilities under these circumstances is sketched in Spencer and Foran, 1991.

17Again, however, while weights can compensate for nonresponse by correcting errors in the population estimates for particular subgroups, they do not correct nonresponse *bias* within subgroups. For example, weighting can adjust for the fact that male eighth graders responded to NELS:88 at a lower rate than did their female classmates, but do not address bias that may be present if male responders and nonresponders differed in the very characteristics inquired into by the base year student questionnaire.

18Discussions of longitudinal conditioning or panel effects (also known as "time in sample bias" or "panel conditioning")--for example, whether strong effects potentially exist or could affect data quality--may be found in Kasprzyk, D., Duncan, G., Kalton, G., & Singh, M.P., eds. *Panel Surveys*, 1989 (New York: Wiley). See especially contributions by B. Bailar; D. Cantor; D. Holt; A. Silberstein and C. Jacobs; L. Corder and D. Horvitz; and J. Waterton and D. Lievesley.

DISCUSSION

Gary M. Shapiro
Abt Associates

The two papers in this session discuss the methodology of two of the most important longitudinal surveys ever conducted, and thus are of considerable interest. This session is on longitudinal surveys, but in fact the papers relate to two topics that are in the Statistical Policy Working Paper Series in addition to longitudinal surveys: nonresponse and survey coverage.

Both of these surveys have been very well designed. The sample design issues for the surveys have been carefully investigated, and are well documented both here and in earlier publications. Specific comments on each paper follow.

I. National Medical Expenditure Survey Paper by Steven Cohen

I will separately discuss the main topics covered in the paper. The first topic was the screening procedure and differential sample selection by demographic characteristics. The interviewers returned to screened-in addresses rather than follow individual persons. The paper provides a careful, detailed discussion of the consequences of following addresses, and contains a good discussion of the pro's and con's of following addresses vs. persons.

I have a couple of side observation on this topic. First, there were 722 units that were occupied at time of screening that became vacant by time of interview. However, there were only 448 vacant units that became occupied by time of interview. In theory, one would expect these two figures to be equal. The difference is probably statistically significant. This most likely signifies some coverage loss - either some of the 722 were not really vacant, or more than 448 of the vacants actually became occupied. I do not find this discrepancy to be alarmingly large, but it is evidence of a minor problem.

My second observation is on the use of screening. There were about 28,000 occupied addresses that were screened, with about 15,000 interviewed the next year. The smallest sampling rate in any demographic group was about 40%. This makes me wonder how cost effective the screening was. Fewer demographic categories and a simpler differential sampling scheme would quite possibly have been about as effective. It might have been preferable to undertake a very simple sub-sampling scheme at the time of interview instead of having entirely separate screening interviews. This has obvious implications for NMES-3.

I have no comments on the paper's second major topic, the return visits to households that were nonresponse for screener interviews, in an effort to improve response rates.

The third topic dealt with persons who were nonrespondents for only some of the waves. The paper discusses the characteristics of these people and estimation considerations for them. About 6.2 % of all participants missed at least one wave but not all waves, and about half of them were missing more than 1/3 of the data. The original paper states "When the level of partial response is low, it is often preferable to treat respondents as complete nonrespondents", as opposed to using imputation for missing waves. I invited audience members to discuss under what circumstances nonresponse adjustment should be done as opposed to imputation. One or two audience members expressed the view that imputation might have been preferable in this survey, and indicated some level of disagreement with the quotation above.

The fourth major topic was on the estimation for the institutional sample, when people move in and out of institutions during the year. I have no significant comments on this topic.

II. National Education Longitudinal Study of 1988(NELS:88) Paper by Steven Ingels and Jeffrey Owings

There are three main topics in this paper, two dealing with coverage issues and one dealing with nonresponse. I'm extremely pleased to see the emphasis here on coverage. Coverage is a major problem in many surveys, but it usually receives much less attention than nonresponse and many other less important topics.

The paper first discussed students that were excluded from the sample, with emphasis on disabled and non-English speaking student exclusions. These are reasonable exclusions that I think many survey practitioners would not have worried at all about excluding - I applaud the authors for their concern about the effects of the exclusions. I very much liked the paper's discussion of the effects of this undercoverage on survey results - more analyses of this type are needed.

The decision to follow students who were ineligible in 1988 in the 1990 and 1992 interviews is exemplary. I think most survey practitioners would probably not have undertaken the expense. It's unfortunate that except for the language exclusion category, the follow-up was not very effective.

The second topic of the paper was on sample freshening, to assure that all tenth graders in 1990 were covered in the survey. This again is a coverage issue that many surveys would not have worried about, and it is a real pleasure to see this level of care in survey implementation. As a side note, the survey estimated the number of schools attended by tenth graders in 1990 using complex methodology in Spencer and Foran(1991). I believe a better approach to this type of estimation issue is one used in the Survey of Income and Program Participation for estimating the number of households when following individuals (see Huang,1984, and Ernst,1989.)

The third topic dealt with two nonresponse issues. I have no comments on this section.

The last section of the paper briefly mentions several other sources of error, including a statement that item nonresponse was rarely imputed for. I note that even when there is not explicit imputation in a published table, there is implicit imputation in the sense that a reader generally assumes that nonrespondents are distributed like respondents. Thus, even a very crude explicit imputation is usually better than no imputation at all.

REFERENCES

Ernst, L. (1989), "Weighting for Longitudinal Household and Family Estimates", in "Panel Surveys", edited by D. Kasprzyk, G. Duncan, G. Kalton, and M.P. Singh, pp. 139-159, Wiley and Sons.

Huang, H. (1984), "Obtaining Cross-sectional Estimates for a Longitudinal Survey: Experience of ISDP", Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 670-675.

Spencer, B. and Foran, W. (1991), "Sampling Probabilities for Aggregations, With Application to NES:88 and Other Educational Longitudinal Surveys", Journal of Educational Statistics, Vol. 16, No.1, pp. 21-33.

DISCUSSION

James M. Lepkowski
University of Michigan

Panel surveys are used to address many scientific and policy issues. The primary focus of many panel surveys is to assess change over time and examine causal mechanisms in various phenomenon. A wide range of designs are employed in these panel surveys, reflecting the diverse uses to which panel survey data may be put.

The presentation of papers on the design of two panel surveys in this session provides an opportunity to compare and contrast features of panel survey design. The two panel surveys described in these papers have very different topics, populations, sampling frames, and sources of survey error. The National Education Longitudinal Study of 1988 (NELS 88) is designed to provide data on education for a cohort of eighth graders followed over a six year period, sampling students from schools. The National Medical Expenditure Survey (NMES) is a survey of health and health care utilization and coverage among the U.S. civilian non-institutionalized population and those residing in nursing homes. The NMES uses a national area sample frame supplemented with a national nursing home list. There are similarities in design features between these surveys, but there are also many interesting differences. Comparing and contrasting these surveys illustrates the value and flexibility of panel survey design.

The two surveys, for example, have different statistical estimation purposes. The NELS 88 provides data that can be used to create cross-sectional estimates at each data collection time period, inter-cohort comparisons over time, and individual change through the repeated measurement of the same individuals. These are typical goals of many panel survey designs, attempting to answer questions about a single time period through cross-sectional estimates and about changes over time in a group or among individuals of the group. The NMES uses the panel survey design for a different purpose, to provide precise and accurate measurement of health and health care utilization and coverage through the accumulation of retrospective reports from a panel of respondents. By limiting the time between observations, NMES reduces the size of recall errors. There is little if, any interest, in assessing change at an individual or group level in the NMES.

Different purposes have lead to different design features. NELS 88 uses a longer period between interviews, following the cohort of eighth graders every two years. Characteristics measured in the NELS 88 are not expected to change substantially in a shorter interval, and thus more frequent data collection does not appear necessary. NMES interviews reporting units (essentially households) every three or four months over a 16 month period.

This frequent interviewing schedule allows NMES to measure continuously the rapidly changing health characteristics of the sample. The longer period between interviews would pose a larger problem for the NELS 88 if it were not for the fact that a large share of the NELS 88 sample each interview is students in school being followed. Information on contact persons is collected to assist in the tracking of students who leave school or move to another location and school. The NMES also collects contact information to assist tracking efforts, but NMES reporting unit members appear to be more difficult to track than the NELS 88 student sample.

An important methodological difference between the two surveys is the need for screening and sampling at varying rates subgroups of the population. NELS 88 does not need to sample any particular subgroups at higher rates than others. NMES attempted to sample poorer persons at higher rates through an interview conducted before the primary data collection began. A substantial share of the NMES paper is devoted to an examination of the screening procedures and their effectiveness.

The two surveys address the issue of the changing nature of the population that the sample represents quite differently. In panel survey design, "representation" of the population by the sample is hindered by loss due to non-response and failure to recruit into the sample persons who are recently joined the population such as immigrants. The NELS 88 sample is "refreshed" every interview period by a selection of additional students who could not have been selected at any of the previous rounds of data collection because they were not members of the U.S. eighth grade student population in 1988. The refreshing process is designed to improve the coverage of the NELS 88 sample of students who were eighth graders in 1988 at any given future interview round. The NMES covers a much shorter time period during which fewer changes to the population that could affect the quality of the NMES data occur. For instance, immigration changes to the U.S. civilian population is not expected to be an important departure from complete coverage during the time period of the NMES.

The NELS 88 and NMES are both subject to non-response, and both surveys make considerable effort to reduce non-response rates. The NELS 88 faced difficulty recruiting school districts and schools, the primary sampling units in the selection, into the sample. When a sample school district or school refused to participate, NELS 88 substituted a school district or school as similar as possible to the non-participating school. Student participation within cooperating schools was quite high, with very low non-response at the initial round and all subsequent rounds of data collection. High response rates were maintained despite the difficulties of tracking school drop-outs. NMES did not lose any primary sampling units to non-response in the household sample portion of the sample, although it did experience losses of reporting units and reporting unit members at first and subsequent interviews. NMES also experienced loss of primary sampling units

in the nursing home survey portion of the sample, as well as difficulty obtaining interview data for persons who were extremely ill.

Many panel surveys compensate for non-response losses through weighting for unit non-response and imputation for item non-response. The NELS 88 analyzed first and subsequent round non-response, but it is not clear from the paper if non-response weights are employed. No imputation has been done to compensate for item non-response rates, which are reported to be low. The NMES discards persons from the sample who provided reports for less than one-third of the target year. Non-response compensation weights are employed to adjust for the these and all other persons who failed to respond to the survey adequately. As in the NELS 88, imputation is not employed to compensate for item non-response. As a result, both surveys implicitly impute for missing items by using available data, effectively an imputation of the mean values of the responses for the missing items.

The NELS 88 and NMES are examples of well-designed and carefully executed panel surveys. Since both have different goals and topics, the panel designs employed by each are quite distinct. Those interested in understanding more fully the features and complexities of panel survey design can learn much from a reading of these two papers.

Reports Available in the Statistical Policy Working Paper Series

1. Report on Statistics for Allocation of Funds (Available through NTIS Document Sales, PB86-211521/AS)
2. Report on Statistical Disclosure and Disclosure-Avoidance Techniques (NTIS Document sales, PB86-211539/AS)
3. An Error Profile: Employment as Measures by the Current Population Survey (NTIS Document Sales PB86-214269/AS)
4. Glossary of Nonsampling Error Terms: An Illustration of a Semantic Problem in Statistics (NTIS Document Sales, PB86-211547/AS)
5. Report on Exact and Statistical Matching Techniques (NTIS Document Sales, PB86-215829/AS)
6. Report on Statistical Uses of Administrative Records (NTIS Document Sales, PB86-214285/AS)
7. An Interagency review of Time-Series Revision Policies (NTIS Document Sales, PB86-232451/AS)
8. Statistical Interagency Agreements (NTIS Documents Sales, PB86-230570/AS)
9. Contracting for Surveys (NTIS Documents Sales, PB83-233148)
10. Approaches to Developing Questionnaires (NTIS Document Sales, PB84-105055/AS)
11. A Review of Industry Coding Systems (NTIS Document Sales, PB84-135276)
12. The Role of Telephone Data Collection in Federal Statistics (NTIS Document Sales, PB85-105971)
13. Federal Longitudinal Surveys (NTIS Documents Sales, PB86-139730)
14. Workshop on Statistical Uses of Microcomputers in Federal Agencies (NTIS Document Sales, PB87-166393)
15. Quality on Establishment Surveys (NTIS Document Sales, PB88-232921)
16. A Comparative Study of Reporting Units in Selected Employer Data Systems (NTIS Document Sales, PB90-205238)
17. Survey Coverage (NTIS Document Sales, PB90-205246)
18. Data Editing in Federal Statistical Agencies (NTIS Document Sales, PB90-205253)
19. Computer Assisted Survey Information Collection (NTIS Document Sales, PB90-205261)
20. Seminar on the Quality of Federal Data (NTIS Document Sales, PB91-142414)
21. Indirect Estimators in Federal Programs (NTIS Document Sales, PB93-209294)
22. Report on Statistical Disclosure Limitation Methodology (NTIS Document Sales, PB94-165305)
23. Seminar on New Directions in Statistical Methodology (NTIS Document Sales, PB95-182978)

Copies of these working papers may be ordered from NTIS Document Sales, 5285 Port Royal Road, Springfield, VA 22161 (703)487-4650