NCHSR

RESEARCH PROCEEDINGS
SERIES

Health
Survey
Research
Methods

Second
Biennial
Conference

Williamsburg, Virginia
May 4 - 6, 1977

Editor: Leo G. Reeder
          School of Public Health
          University of California
          Los Angeles

CONTRIBUTORS:

Lu Ann Aday
Ronald Anderson
Thomas Bice
Norman Bradburn
Gary Bridge
Charles Cannell
Joseph Duncan
Jack Elinson
Bradford Gray
Bernard Greenberg
Daniel Horvitz
William Kalsbeek
Judith Lessler
Kent Marquis
Lois Monteiro
Leo Reeder
Naomi Rothwell
Seymour Sudman
John Ware

CONTENTS

FOREWORD

The National Center for Health Services Research (NCHSR) and the National Center for Health Statistics (NCHS) support research in survey methods in order to increase the validity and reliability of information obtained concerning measures of health, the availability of health services, and the use of health services. In 1975, the two Centers jointly sponsored the first National Invitational Conference on Health Survey Research Methods, bringing together leading researchers in health survey methodology to address the problematic issue of the dissemination of the state of the art to the general health services research community. The proceedings were published as "Advances in Health Survey Research Methods," DHEW publication no. (HRA) 77-3154.

Successful as that conference was in the judgments of those attending and its sponsors, it was not long enough to address all the pertinent subjects. One of the outcomes was the realization that insufficient time and attention had been devoted to the complexities and importance of total survey design; another concerned the stimulation of additional research, which generated a number of unanswered research questions. These, together with the technical progress that has been made in the refinement of health survey methods and measures, and the recommendation of the first conference that the two Centers periodically sponsor meetings to include the review of the state of the art and analytical techniques and processes pertinent to recent approaches to survey research in health, prompted the original Planning Committee to seek support for this second conference.

Recognizing these continuing needs, the two Centers, in cooperation with the Veterans' Administration, sponsored this second conference through the School of Public Health, University of California at Los Angeles. All of the topics discussed herein are highly relevant to the effective execution of research, and the publication of this report is believed to be a contribution toward the fullest implementation of NCHSR's responsibility for the communication and use of research findings. The ultimate success of these efforts depends upon the users of these methods, and it is hoped that the additional visibility given health survey research methods through these conferences and their proceedings will result in a perceptible improvement in the design of health services research, in data quality, and in the impact of research results on national health policy.

Gerald Rosenthal, Ph.D.
Director
National Center for
    Health Services Research

Dorothy Rice
Director
National Center for
    Health Statistics

May 1978

PLANNING COMMITTEE

ACKNOWLEDGEMENTS

Charles F. Cannell, University of Michigan
Bernard G. Greenberg, University of North Carolina
Daniel G. Horvitz, Research Triangle Institute
William M. Kitching, NCHSR
Leo G. Reeder, Chairman, University of
    California at Los Angeles
Seymour Sudman, University of Illinois at
    Urbana-Champaign
Elijah L. White, NCHS

INTRODUCTION

Leo G. Reeder, University of California at
Los Angeles.

This conference is the second in what is planned as a series of symposia to synthesize the state of the art of survey techniques relevant to health surveys. The first conference utilized a semi-structured format; no papers were prepared for the meeting (NCHSR, 1976). In planning for the second conference, the Planning Committee decided to formalize the structure of the meeting somewhat by commissioning several papers on selected topics. These topics appeared to be most salient at the time for examination by an invited group of survey research experts. They are by no means comprehensive of survey research methodology. Thus, in addition to including such traditional methodological issues as response rate, the Planning Committee decided to recognize the increasing criticism of surveys as an intrusive technique by commissioning a paper on "Respondent Burden." In addition, another type of intrusion, the invasion of individual privacy, was also recognized as a critical issue to be discussed at the conference.

In addition to commissioned brief papers that summarized the state of knowledge about the subject matter, each session permitted open discussion among the conference participants who had been mailed the papers in advance of the conference. A recorder, working in collaboration with the chair of each session, summarized the discussion. It is worth noting that the conference participants came from multiple disciplines and professional backgrounds. This feature of the conference enhanced its character.

The objectives of this conference remained the same as the 1975 meeting:

1. To identify the critical methodological issues or problem areas for health survey research and the state of the art or knowledge with respect to these problems.
2. What types of research problems need to be given high priority for research funding.
3. To identify policy issues that can be addressed by survey research methods.
4. To communicate the results, recommendations, and implications of this conference to:
   a. the broader community of health researchers who use survey methods;
   b. relevant Government agencies and individuals;
   c. other potential users of these results of this conference.

This conference, like the first, was co-sponsored by the National Center for Health Services Research and the National Center for Health Statistics, with the cooperation of the Veterans Administration. The conference was funded by a grant from NCHSR (HS 2746). Both of the sponsoring agencies have supported considerable research aimed at improving the quality of health surveys. Together with the National Science Foundation, these agencies have made important contributions to advancing the body of knowledge concerning survey methodology in recent years. But there is still much to be done in a field undergoing rapid technological and substantive changes.

New techniques using the telephone and the computer-console-telephone to collect data from respondents, sampling strategies, analytical techniques for survey data, and issues concerning privacy and confidentiality, suggest the need for continuing methodological research to improve this traditional and standard technique. Assessing people's views, attitudes, and behaviors using surveys and interviews has limitations that are only too well-known to the sensitive researcher in this field. There are alternatives to the survey method and some of these were discussed at another recent conference (Sinaiko and Broedling, 1976). Every method has associated with it some type of degree of error and hence, no one method is totally and completely satisfactory. Until something better comes along to take its place, it is the better part of wisdom to continue to make efforts to improve survey methods since it is the method on which an overwhelming proportion of research in health, social science, and other fields depends. Nevertheless, survey researchers need to be alert to approaches and measurements that are supplementary or complementary to surveys, particularly to provide confirmatory or additional data. Use of experimental designs in survey research, improvements in interviewer training, and the elaboration of the conventional notion of survey research to include extensions and variations (such as

use of vignettes to enhance the content of the interview beyond the limitations imposed by experiences of the respondent) will, among other developments, further improve the effectiveness of the survey method. Although surveys have drawbacks and limitations, they are still remarkably good tools for answering the questions of policy makers as well as scientific researchers.

The second Biennial Conference on Health Survey Research Methods repeated one topic that had been a major focus of the initial conference, i.e., Total Survey Design. In part, this was because of the interest expressed by those in attendance at the first conference who felt that more time was needed to deal with the complexities involved. Also, the Planning Committee felt that the subject matter needed further elaboration and visibility. The position paper by Kalsbeek and Lessler demonstrated the need and application of Total Survey Design by presenting examples of survey biases from two actual large-scale studies.

The second paper, on "Response Rate," by Marquis, addressed an issue that has been discussed by survey researchers for many years but its salience at this time is related to an underlying thread at this conference, namely, the intrusiveness of surveys, complaints about too many surveys, etc. This same feature of surveys today was responsible for the commissioning of the paper on "Respondent Burden," by Bradburn. Both papers appear to allay some of the fears concerning the public's responsiveness to surveys. Similarly, Bradford Gray addressed the issues of privacy, confidentiality, and the rights of human research subjects, as viewed by the experiences of the National Commission for the Protection of Human Subjects.

The final commissioned paper, by Andersen and Aday, describes some of the advantages and problems associated with the attempt to standardize certain commonly-used survey items or measures. They raise several critical questions that need to be asked with respect to standardizing variables in the health field.

Finally, the interface between researchers and the Office of Management and Budget (for those engaged in contract research especially) was addressed by Duncan from his unique perspective. Of primary interest was the process of clearance of survey forms and instruments.

As noted above, the session chair and the recorder have synthesized the discussions that took place during each session. Theirs was a most difficult task indeed! But it is this writer's opinion that they succeeded very well. The next question we might ask is: How far have we come?

SUMMARY OF RECOMMENDATIONS

Assessing Prior Recommendations: Although only a short period of time has passed, we might try to assess the recommendations made at the Airlie House Conference before going on to summarize the Williamsburg Conference.

- In the conference at Airlie House, a number of policy issues emerged concerning the survey method of collecting data. It was observed that there was insufficient investment in methodological research to improve the quality of data obtained through surveys, despite the reliance upon these data by policy-makers. The need for increased investment in the methodological aspects of helath surveys, particularly non-sampling problems, continues to be a problem. Some progress is being made in this direction, but the general problem still looms large. The issue can be materially advanced by encouraging greater investment and collaboration between NIH, NCHS, and NCHSR.
- The Airlie House Conference had also suggested that sponsoring agencies establish guidelines of "good practices" in survey methods and procedures. Applicants for contracts and grants would routinely obtain a set of such guidelines. This suggestion has yet to be implemented.
- Both conferences had sessions devoted to the issues involved in privacy and protection of respondents. Clearly, considerable progress has been made in illuminating the implications of the recent federal legislation for researchers by the actions of a variety of professional organizations and agencies.
- The inclusion of Total Survey Design (TSD) in this Conference implemented a recommendation of the Airlie House Conference to familiarize the research community with this approach.
- Development of an information system containing data on various error components was suggested at the Airlie House Conference. This suggestion has yet to be implemented and is discussed further in the recommendations evolving from this Conference.
- Finally, the first conference at Airlie House suggested that the two sponsoring agencies (NCHS and NCHSR) initiate a Summer Seminar reviewing the state of the art in health survey research techniques, methods, and processes. This suggestion has not yet been implemented although alternative mechanisms are being utilized for such purposes.

In sum, there is movement toward achieving the recommendations and suggestions made at the first conference at Airlie House. Assuredly, not all of these recommendations need to be achieved. There may be compelling reasons for not implementing a recommendation. At any rate, some recommendations have been implemented, others are being actively considered, while still others remain to be acted upon. This is quite satisfactory progress for a two-year interval.

Summary of recommendations made at this Conference: At a number of points during the conference, recommendations were made for improving research methodology in survey-type research. These recommendations are assembled here as follows:

- Research support for certain high-priority problems including:
  a. Development of a unified total survey error model;
  b. A series of studies on the problems of non-response and measurement bias.
- The available data suggests that response rates are under the potential control of the field designer. Resources and design skills need to be applied to this task to achieve satisfactory response levels.
- There is a need to standardize definitions in surveys—particularly salient is the issue of response rates with particular emphasis on the denominator (number of potential respondents in the Universe). This is especially true in the case of the random digit telephone survey. Another concept in need of clarification is respondent burden. How is it to be defined? How is it measured? What are its effects?
- A number of techniques were proposed as devices for reducing respondents' actual burden including the use of instrumented sampling and interviewing; using respondent *estimates* of relatively frequent events rather than very pre-

cise, time-consuming responses; and using abbreviated attitude scales rather than lengthier scales. It was also observed that psychometrically "weaker" scales may be adequate for observing between-group differences.

- A variety of techniques to reduce the burden of repeated interviewing on small, interesting populations were described, including guarantees of not being surveyed more than once within a given time period. Subjective motivation for improving respondent cooperation was highlighted by several techniques. Improving question content and sequencing to arouse and maintain interest in the interview was recommended.

- With respect to the issues generated by the privacy legislation, the conference was reminded that there are unequal benefits to privacy in our multi-group society. There is a need for research about privacy as a value, who wants it and why, and under what circumstances people will trade some privacy for some other benefit.

- More research is needed on standardization of measures commonly employed in health surveys, such as: utilization, morbidity, insurance coverage, and health attitudes. What types of encounters constitute a physician visit? Are symptom check lists in need of more or less specificity? How do we measure intensity or severity of disability days? Other areas of research needs are health attitudes and orientations as conditioning variables.

- Finally, and perhaps most importantly, this conference reiterated the recommendation made at Airlie House for an information matrix for health survey research. The conference recommends that a feasibility study be initiated as early as possible on the likely costs and benefits involved in the development of such an information matrix. It was suggested that NCHS and NCHSR set up a joint committee, together with outside members, to facilitate such a study. The proposed information matrix would include a variety of data: items used, mode of administration, nature of sample, general purpose of the questionnaire, location of items, item reliability, validity, and temportal stability, and errors in the survey.

These recommendations constitute an agenda for research and action. There is clear evidence that there is greater interest in many of these issues and other methodological survey research questions. This is manifested in such diverse ways as the development of a new section of the American Statistical Association concerned with survey research methods, new journals emphasizing methodological research, official and quasi-official committees that address many of the issues discussed in these conferences, and other activities. All of this is beneficial to the further development of this enterprise. As all who engage in this type of activity know, survey research is costly, as are attempts to unravel complex methodological questions.

Hopefully, government agencies and foundations will recognize these realities and provide greater opportunities for investigators to tackle the compelling issues aired at these conferences.

SURVEY RESPONSE RATES: SOME TRENDS, CAUSES, AND CORRELATES*

Kent Marquis, The Rand Corporation

INTRODUCTION

What has happened to survey response rates? Why? and what can be done about it? Social scientists are currently concerned about trends in survey response rates as illustrated by the following recent statements:

"Pollsters are increasingly concerned about the growing reluctance of the public to be interviewed. The refusal rate has increased. It costs more to find respondents. The same pattern is evident in questionnaire surveys." (Lipset, 1976).

"There is a pronounced secular decline in the response rates from personal interviews: it is increasingly difficult to maintain response rates at reasonable levels, and despite increased effort and cost designed to maintain response rates, the decline has persisted." (Juster, 1976).

The paper is organized to explore the social changes hypothesized to underlie current response rate problems, to examine the nature of the trends based on data furnished by major survey organizations, and then to consider possible causes of the observed differences. Attention is focused mainly on the topic of in-person interviews. The response rates cited here are computed as the number of completed interviews with eligible sample units divided by the total number of eligible sample units. Bailar and Lanphier (1977) provide a discussion of current practices in defining response rates and illustrate possible problems with actual data from 36 contemporary survey research projects. The definition used here is the approach recommended by Bailar and Lanphier.

HYPOTHESIZED SOCIAL CHANGE CHARACTERISTICS CAUSING RESPONSE RATES TO DECLINE

The conventional wisdom asserts that survey response rates have been declining over recent years because of changes in society, the effects of which the survey practitioner can neither control nor overcome. The categories of change most often mentioned are of three types: availability, privacy, and physical security.

Availability. Increasing participation in activities outside the home makes it difficult to locate respondents for an interview. For example, in 1960 the U.S. labor force participation rate for women 18-64 years old was 42 percent. By 1974 it had increased to 53 percent (U.S. Bureau of the Census, 1975, based on Table 559). This trend should show up in increasing noninterview rates over time (especially the not-at-home component of household interview studies). This hypothesis is examined in the next section. Contrary to expectations, such a trend is not found in the studies examined.

Privacy. Persons are said to refuse interviews because their answers may be misused. Advances in data handling technology make the personal privacy issue a real one. It is now possible to create, link, and access large banks of information organized by individual identifiers such as the social security number. The media have pointed out unsuspected use of stored personal information by government officials, credit bureaus, and other institutions. Recent concern has focused on the lack of safeguards to protect medical information in data banks (e.g., Westin (1976) and the 1975 CHSS Workshop on Privacy and Confidentiality).

Concern about privacy should manifest itself in increasing survey refusal rates. Yet the available refusal rate trend data, shown in Tables 2-4 of the next section, suggest little or no increase in the last 10-15 years.

If privacy concerns are increasing, they may represent future problems rather than adversely affecting past survey efforts. The Privacy Act of 1974 and emerging state legislation now require many of us to give extra information to potential respondents about the voluntary nature of participation and the potential

uses of volunteered personal information. If privacy concerns were not salient to the public before, these required statements may make them salient now. Two recently completed studies have addressed privacy issues.* One study, conducted by the Bureau of the Census, experimentally varied what the respondent was told about how long his answers would remain confidential (forever, 75 years, 25 years, immediately released, not mentioned). The dependent variables are interview refusal rates before and after the introductory statement. Eleanor Singer and the National Opinion Research Center have completed a national survey in which interview and item response rates are studied in relation to 3 experimental dimensions: the amount of information given to a potential respondent about the survey, the strength of the assurance of confidentiality (absolute, "except as required by law", not mentioned) and a request that the respondent sign a consent form (no request, request before interview, request after interview). Results of the studies had not been released at the time this paper was written.

Physical Security. It is felt that the increase in crime rates affects survey response rates in at least two ways: directly by causing reluctance to answer the door when an interviewer calls and indirectly by causing people to live in buildings with security arrangements that exclude both interviewers and crooks. Trend data relevant to prevalence of security buildings and reluctance to answer the door do not appear to exist. The problem of nonfederal interviewer access to buildings with guards does exist. For the SRC economic surveys, Juster (1976) notes a recent 64 percent response rate for respondents in multi-unit structures, a 55 percent rate for residents in structures with more than 9 units and rates between 39-55 percent for units with such entry barriers as "guard dogs, locks (and) doormen." Nevertheless, the data in the next section suggest the overall nonrefusal component of nonresponse rates has not declined. If security building problems are increasing, it appears that other availability problems may be decreasing.

Other Social Change Issues. Other hypothesized changes in society are sometimes mentioned as causes of difficulties obtaining interviews. These include the idea that individuals are being over surveyed, that salespersons posing as interviewers are increasing, and respondent attitudes toward surveys and the intended uses of survey data may be changing.

A collaborative effort between the Census Bureau and Michigan's Survey Research Center includes a detailed investigation of respondent experiences with surveys along with their knowledge, opinions, and attitudes about surveys. Some of the questionnaire items are similar to ones used in earlier NORC and SRC surveys so time trends can be inferred. Until

these results are available, the only "hard" data available come from a couple of recent telephone surveys by Walker Research, Inc. (1975, 1976).* The data are not (unbiased) national population estimates.

In both surveys about 50 percent said they had been interviewed previously in the last 12 months (half of these by phone, 25 percent by mail and the remainder in the home or at a shopping center). Forty percent said they had experienced a sales pitch disguised as a survey and this group had slightly less favorable attitudes toward surveys.

Answers to questions about privacy and exploitation (along with the questions and response distributions eliciting the most and least favorable replies) are shown in Table 1.

For those who believe there is a casual link between prior attitudes and behavior, these results are cause for concern. Response rates (behavior) can be expected to suffer if half the telephone-owning households in a market area have been interviewed in the last year and at least 20 percent of all respondents have unfavorable attitudes about surveys. Fortunately, perhaps, the link between responses to attitude survey questions and behavior is not well established.

In the next section, the other part of the conventional wisdom is examined: Have response rates declined over the years and where are they now?

TIME TRENDS FOR IN-PERSON RESPONSE RATES

In this section, response and refusal rates for national studies conducted by federal and university-based survey organizations are examined. The data do not support the conventional wisdom: while response rates in the 1950's may have been higher than they are now, there is no definitive declining trend over the past 10-15 years. There are differences between organizations which have persisted over time and the understanding of these differences may shed more light on determinants of response rates than the previous discussion of changes in society.

One-time, In-person, Population Surveys. Trends in overall response rates and refusal rates for selected studies may be seen in Table 2. The Health Interview Survey employs a national sample of 35-40 thousand households per year. The fieldwork is done by the Census Bureau and each unit is interviewed once. One adult may respond for all household members although self-response is encouraged when possible. Interviews cover health problems and use of health

Table 1. *Percentage Agreement to Selected Survey Image Statements: 1974, 1976*

STATEMENT	PERCENT AGREE (STRONGLY, SOMEWHAT)	
	1974	1976
The research industry serves a useful purpose	87	83
The information obtained in polls or research surveys helps manufacturers sell consumers products they don't want or need	40	42
Polls or research surveys are an invasion of privacy	29	27
Answering questions in polls or research surveys is a waste of time	19	22

(Source: Walker Research, Inc., 1975, 1976.)

5

services, lasting from 30 minutes to over an hour. The Michigan economic surveys are conducted periodically by the Survey Research Center (SRC) at Michigan. The sample sizes range between 1,500 and 2,500 households in the coterminous United States. The average length of the interview has ranged over the years from 45 minutes to an hour-and-a half. Before 1972, the respondent was the head of the household. Specific persons are currently designated as respondents. The National Opinion Research Center (NORC) data are from selected national studies with interview lengths from 1 to 2½ hours. Subject matter of the interviews varies from study to study as do the respondent rules.

The data, taken as a whole, do not indicate uniform trends toward massive respondent noncooperation over time. The overall response rates for the Health Interview Survey are high, stable, and possibly getting better over the 15 year period shown. The data from the University-based organizations suggest that during most of the past 10 years, response rates were in the 70's-to-low 80's with the SRC economic surveys getting higher rates 15 to 20 years ago compared to recent years. The decline in SRC rates beginning in the early 70's is attributed to a change in respondent rules. The SRC refusal rates exhibit a possible increasing trend over time and a similar trend could be occurring in the Health Interview Survey. As footnote *h* points out, a recent large-scale NORC national survey of access to medical care achieved response rates in the 80's or 90's, indicating that such rates are still possible to achieve currently by nonfederal organizations for national samples.

Trend data were not readily available from the large number of private organizations who conduct survey research and polls. These organizations claim to be experiencing problems as the following excerpt from the 1973 Conference on Surveys of Human Populations (American Statistical Association, 1974) shows:

"... spokesmen for a number of private survey organizations, large and small, who were queried by one of the conference participants, all report that their completion rates on general population samples now average approximately 60 to 65 percent, in spite of three or four callbacks. This recent experience is in contrast to a completion figure of 80 to 85 percent for the same firms in the decade of the sixties."

Public disclosure of detailed response rate data from private organizations might aid in reconciling the apparent differences between such claims and inferences drawn from the federal and university-based organization experiences.

Repeated Interview Surveys. Possibly major declines in response rates can be found in the short term panel surveys which seek continued respondent cooperation. Time trends in response rates for three studies of this type, all conducted by the Bureau of the Census, are examined next. The data indicate the Census Bureau has been able to maintain very high response rates over the years but may be experiencing a slight contemporary increase in refusals.

The Current Population Survey has a current annual sample size of 48,000 households. A household respondent is interviewed about labor force and other data monthly for four months, not interviewed for eight months, and interviewed again monthly for another four months. In-person interviews are attempted during a designated three of the eight interview months. Telephone interviews may be

Table 2. Response[1] and Refusal Rate[2] Ranges for One-Time, In-Person Interview Studies: 1955-Present

YEAR	HEALTH INTERVIEW SURVEY[3]		MICHIGAN ECONOMIC SURVEYS[4]		SELECTED NATIONAL STUDIES[5]	
	Response	Refusal	Response	Refusal	Response	Refusal
1955–1959			83-88	5-8		
1960–1964	95-96	Not Available	79-85	6-12	76-83	13-17
1965–1969	95-96	1.2–1.3	78-83	9-14	75-76	15-18
1970–1974	96-97	1.1–1.5	78-81[7] (72-76)+5	12-14 14-16	74-82	14-17
1975[6]	97	1.4	(72-74)+5	15-16	[8]	

[1]The response rate is the number of completed interviews per 100 eligible households (or other sampling units).

[2]The refusal rate is only one component of the overall noninterview rate for eligible units. Other components (which include no one home, seasonal absence, language barrier, mental or physical problems precluding interview, etc. are omitted). The Health Interview Survey refusal rate is shown to one decimal place because rounding error would have obscured what some have seen as a trend toward increasing refusal rates over time.

[3]1968-1975 data adapted from Love and Turner (1975). 1960-1965 estimates derived from average noninterview rates per interviewer (Koons, 1973) assuming no correlation between an interviewer's total assignment size and her noninterview rate. Estimates for 1966-67 are not included. The Health Interview Survey uses a household respondent while the others are based on randomly selected respondents.

[4]Data adapted from Scott (1971 and 1976).

[5]Data courtesy of Martin Frankel, technical director, NORC.

[6]Recent data are provisional and not always for complete calendar years.

[7]First range is for 1970-1971. Thereafter, a different respondent selection and interviewing procedure was instituted causing an immediate 5 percent drop in response rates. (Juster, 1976, Fig. 1).

[8]Person response rates (Z rates) for the 1976 medical access study, involving all 4 samples of between 1200-4700 households each ranged from 82 to 98 percent (M. Frankel, personal communication).

substituted during the other months. The Current Population Survey has maintained consistently high response rates and low refusal rates for the last 10 years as shown in Table 3. If there are any trends, they are for the overall response rates to be increasing slightly over time and possibly for refusal rates to be increasing slightly. This implies that if refusals have increased, there is a more than compensating decreasing trend in the other components of nonresponse.

The Current Medicare Survey is also a short term panel study conducted by the Census Bureau. It involves 15 monthly personal (in-person or telephone) interviews which last about 10 minutes with 6,000 Medicare enrollees. Respondents report expenditures for medical goods and services incurred over the preceding month. According to Greene (1976) average response rates over entire panel periods (15 months each) have averaged about 97 percent in 1974, 1975, and 1976.*

Quarterly Consumer Expenditure Panels were conducted with a sample of 11,000 households in 1972-73 by the Census Bureau. For the interview component, households engaged in interviews lasting 2-3 hours in each for five consecutive quarters. Although not able to maintain response rates over 95 percent as in its other studies, the Census Bureau was able to maintain a very impressive level of cooperation throughout all waves. These data (reproduced from Greene, 1976) are in Table 4.

Nonavailability. To complete the picture, trends in the nonavailability component of nonresponse are

6

Table 3. Response and Refusal Rates Ranges for the Current Population Survey: 1965-Present[1]

Year	Response[2]	Refusal[2]
1965-1969	95-95	1.2-1.3
1970-1974	96-96	1.1-1.5
1975-1976	97	1.4

[1]1965-1974 data adapted from Love and Turner (1975) with 1975 and the first 11 months of 1976 data from "Current Population Survey Sample Size and Noninterview Rates—November 1976" compiled by the Reinterview Design and Analysis Section, Statistical Methods Division, U.S. Bureau of the Census, 8 December 1976.

[2]These are average rates over all interviewing waves. Within recent panels, there is a tendency for refusals to increase in later waves and other nonresponse components (e.g., not-at-home) to decrease.

Table 4. Response and Refusal Rates for the Interview Component of the 1972-73 Consumer Expenditures Survey by Year and Wave

Wave	Response Rates 1972	Response Rates 1973	Refusal Rates 1972	Refusal Rates 1973
1	95	94	3.3	4.4
2	91	90	6.9	7.6
3	88	89	8.6	9.1
4	88	89	10.0	9.9
5	88	89	10.8	10.0

(Source: Greene, 1976).

shown in Table 5. The numbers are the difference between the total nonresponse rate and the refusal rate shown in previous tables. The largest component of this residual are elements classified variously as not-at-home, unable to contact, or respondent absent.

These rates also fail to show large changes over time. The studies in which the Census Bureau or NORC did the fieldwork do not show increasing availability problems. The SRC data do indicate a recent increase but this may be due to the change in respondent rule.

The lack of trends may only reflect successful efforts made by these organizations to overcome availability problems, for example, by increasing the number of calls made per completed interview. Call data are not generally available to the research community (they are internal organizational information not necessary to the interpretation of published survey results) so it is not possible to test the hypothesis. However, Juster (1976) notes that the number of contacts per house-

hold in a 1973 SRC economic study was considerably lower than in a 1976 study.

Urban Area Response Rates. Survey response rates are often said to be lower in urban areas than in smaller or less densely populated areas. Available data lend only equivocal support to this hypothesis.

Scott's (1971, 1976) data for the SRC Economic Surveys (see Table 6) show that response rates are lower and have declined more rapidly in the large, self-representing SMSAs than in other areas. The recent (1975-76) rates have been especially low, close to those mentioned for private organizations at the ASA Conference on Surveys of Human Populations (American Statistical Association, 1974). However, at least part of the reason for low contemporary rates is due to a shift to more demanding respondent rules.

Recent Census Bureau experience (Table 7) comparing rates obtained in separate urban area studies with national rates for the Annual Housing Survey and The National Crime Survey do not exhibit the expected discrepancy. As the source articles point out, the SMSA sample field periods are considerably longer than those for national sample assignments, providing extra time to locate respondents and convert refusals. Walsh's (1976) analysis of expenditure diary response rates (Table 8) shows the central city and SMSA rates to be lower than for other areas in 1972 but the differences narrowed appreciably in the following year. As a whole these data suggest that urban area response rates can present problems, but ones that at least the Census Bureau has been able to overcome.

Finally, Eve Fielder at UCLA was kind enough to provide response rate data for the Los Angeles Metropolitan Area Survey which has conducted 11 urban area studies since 1970. There is a good deal of variance in the overall nonresponse rates and all eleven points do not produce a linear time trend significantly different from zero. The refusal rates, on the other hand, do exhibit a statistically significant increase over time averaging 3/4 of 1 percent per year.

CONCLUSIONS ABOUT TIME TRENDS IN RESPONSE RATES

The available data do not show a major decline in response rates over the past 10-15 years. The conclusion to be drawn, however, is *not* that high response rates are as easy to achieve now as in the past. This kind of conclusion would require an analysis of cost data or other indicators of field effort and these data are not generally available. Also response rate data from a more representative sample of survey studies would be required.

The available data are sufficient to suggest two things:

1. Response rates even in urban areas are still under the potential control of the field designer. The changes in society which have taken place

have not doomed survey research to failure in the future.

2. High contemporary response rates can be achieved, at least by some organizations, if sufficient resources and design skill are applied to the task.

In the next section, variables other than changes in society which may affect response rates are considered.

CAUSES AND CORRELATES OF RESPONSE RATE FOR IN-PERSON INTERVIEWS

If there isn't a universal decline in survey response rates we should turn our attention away from external causes toward variables under the control of the survey designer. Some recent literature is reviewed briefly in this section followed by a discussion of other things that may account for observed variation in response rates among organizations.

Auspices. Since the Census Bureau is relatively more successful than other survey organizations, one wonders if there is some magic in the name which over-comes nonresponse problems. As Love and Turner (1975) point out, the Bureau has a strong "brand name", citizens may feel an obligation to cooperate with their government, and the uses of census data, such as estimates of the unemployment rate, are well known. The name of the data collecting agency cannot affect not-at-home rates but it can convert potential refusals and chronically broken appointments.

Sudman and Ferber (1974) split their sample of Chicago area households into those approached in the name of the Census Bureau and those approached in the name of the Illinois Survey Research Laboratory. Respondents were asked to grant an initial interview and to report household expenditures for two weeks either by telephone or using a diary. Initial interview response rates were low (around 60 percent) with the Bureau auspices producing nonsignificantly more cooperation in the suburban sample. Full or partial cooperation in reporting expenditures over the next two weeks showed similar, nonsignificant trends.

The effect of auspices has been tested again in a joint Census-SRC study of confidentiality and attitudes.

Table 5. *Ranges of Nonresponse Rates Exclusive of Refusals[1] from Selected Surveys by Year*

YEAR	HEALTH INTERVIEW SURVEY[2]	MICHIGAN ECONOMIC SURVEYS[3]	SELECTED NATIONAL STUDIES[4]	CURRENT POPULATION SURVEY[2]
1955–1959		7-10		
1960–1964		7-12	4-8	
1965–1969	3-3[5]	6-11	7-9	3-3
1970–1974	2-3	7-8[6]	4-6	3-3
		9-12		
1975[7]	2	10-12	9	2

[1] This is an indirect measure of respondent availability problems. The published data use different definitions and degrees of disaggregation for components of nonresponse not due to refusals. The residual rate which is shown is the difference between the total nonresponse rate and the refusal rate. It includes noninterviews because of language or health problems in addition to the nonavailability components mentioned in the text.

[2] From Love and Turner (1975).

[3] From Scott (1971, 1976).

[4] Furnished by Martin Frankel, NORC.

[5] Years 1968 and 1969 only.

[6] 1970–71 and 1972–74 data shown separately due to a change in respondent rules in the latter years. The revised respondent rule applies to the 1975–76 data also. The apparent time trend is confounded with the effects of the altered respondent rule.

[7] Recent data are provisional and not always for complete calendar years.

Table 6. Average Response Rates for Michigan Economic Surveys
by Large SMSA and Other Areas: 1960–1976

YEAR	LARGE SMSA	OTHER AREAS
1960–1964	75	82
1965–1969	74	83
1970–1974	71	79
1975–1976	61	78

Source: Adapted from Scott (1971, 1976).

Table 7. National Sample and Urban Sample Average Response Rates for the Annual
Housing Survey and National (Local) Crime Survey

YEAR	ANNUAL HOUSING SURVEY		CRIME SURVEY	
	SMSA Studies	National Sample (longitudinal)	Urban Area Studies	National Sample (Panels)
1972			95	95
1973		97	95	96
1974	96	97	97	96
1975		96		96

Source: Adapted from Love and Turner (1976) and Greene (1976).

Table 8. Response Rates for the Diary Component of the National
Consumer Expenditures Survey by Type of Place

FISCAL YEAR	Central City Within SMSA	Other SMSA	Not SMSA
1972	75	81	82
1973	88	89	92

Source: Walsh (1976, Table 3).

Each organization was assigned half of each cluster of households for interviewing. Results are to be available in the 1977 *Proceedings* of the Social Statistics Section, American Statistical Association.

Seasonality. Available data indicate that nonresponse rates (or the not available component) increase slightly during the summer. Scott (1971) examined 29 economic surveys over a 15 year period and found summer nonresponse rates about 6/10 of one percent above spring and fall rates on the average. Interviewer nonresponse rates on the Health Interview Survey (Koons, 1973) were uniformly higher during the summer quarter (July-September) for the years examined (1958-1964). The Current Population Survey temporarily absent component of nonresponse averaged 1.2 percent in June-August 1976 compared to about 0.6 percent in other months of 1976. For most purposes, the magnitude of the summer seasonality bias would not appear great enough to warrant special concern. Palmer (unpublished), however, shows that the seasonal increase in CPS nonresponse and the 1965 nonresponse imputation procedures results in a statistically significant bias in published estimates of the labor force category "with a job, not at work."

Callbacks. It is worth remembering that effort in the field is positively correlated with response rates. Sudman (1967) and Kish (1965) show the productivity of each additional callback on response rates for the 1950's and 1960's. Scott (1976) shows that the marginal productivity of callbacks has not changed in SRC economic studies in the 1973-76 period, e.g., the third callback (4th call) yields a 62-66 percent interview rate. Health Interview Survey data indicate the 4th call yields a 91 percent rate. After the second call the response rate is 74 percent. (Sprately, 1975). Scott's (1971) data indicate that limiting callbacks to three in 1961 reduced response rates at least 5-8 percent compared to rates achieved in similar studies without the restriction in adjacent years. Within the published ranges, it is clear that increasing callbacks results in increasing the number of completed interviews with eligible sample units. (This is one of the few causal relationships we "know for sure" in survey design). The Census Bureau's recent paper on best times to call (Weber, 1973) should help increase the efficiency of calls in the field for a variety of studies using different respondent rules.

The effectiveness of a callback at the margin may also depend on the definition of a call (e.g., a visit while the interviewer is in the neighborhood vs. a deliberate attempt at a different time and day), the length of the field period, the interviewer's workload, and the amount of clustering in the sample. Within the total survey design context, it is theoretically possible to set the number of calls to be made per unit, cluster size, length of field period, number of interviews, and size of assignments so as to optimize costs and the resulting magnitude of nonresponse bias affecting inferences from the data.

Interview Length and Other Burdens. Interview length may have some effects on response rates, especially if length is very short or very long. NORC data, furnished by Frankel, include two, 20-minute national sample studies conducted in 1962 and 1966. Response rates were 90 and 84 percent compared to rates closer to 75 percent for studies using interviews lasting an hour or more in similar years. The SRC economic studies contain a narrower range of variation in average interview length (45-90 minutes). Scott (1971) reports the correlation between response rate and length to be zero or "slight" depending upon assumptions used to correct for secular time trends. The studies conducted by the Census Bureau cited earlier range from a few minutes to several hours in average interview length. Response rates drop to 90 percent only for the extremely long interviews but this may be due to the subject matter (expenditures). Dillman (1977) shows mail rates drop if length exceeds 12 pages.

The Market Research Society in Great Britain (MRS, 1976) points out that sources of burden (on both interviewers and respondents) other than interview length have increased over time. Examples cited include use of batteries of "semantic scales" and intricate questionnaire designs to meet optical scanning processing requirements. Such effects are seen as indirectly affecting response rates by interfering with rapport and causing respondents to be more reluctant to participate in future surveys.

The Walker Research, Inc. study (1975, 1976) indicates a minority of previously surveyed respondents report dissatisfaction with length (17-25 percent), overly personal questions (16-19 percent) or overly difficult questions (8-9 percent). Dissatisfaction with length was clearly related to perceived (recalled) length. Almost half of those who were interviewed for 11-20 minutes objected to the length and 63 percent of those interviewed for more than 20 minutes felt the survey was too long. It would be unwise to conclude, however, that interview length influences response rates via respondent attitudes. If length affects interview rates, it is probably because of the indirect effects on interviewer workload and constraints on the ability to schedule interviews.

Advance Letters, Brochures, Prior Appointments. The Census Bureau has prepared a recent bibliography on this subject for those who wish to pursue the topic further. (Survey Methodology Information System, Undated-a). My interpretation of the literature is that effects of activities prior to the at-the-door contact are equivocal. The advance material does provide information to the respondent upon which he may base his cooperation decision. This is fine if the decision to respond is positive but the other outcome is not unlikely and the reluctance encountered at the door is now much firmer than it might have been. Advance material and appointments can reduce interviewer travel and salary costs, possibly enough to warrant a lower overall response rate. The use of advance contact is certainly a dimension to be considered in the total survey design.

Incentives. Paying respondents to cooperate was discussed at the last conference (NCHSR and NCHS,

10

1975, pp.16-17). At this point, a simple generalization about the effects of compensation on response rates isn't possible. The literature indicates incentives do increase response rates in mail surveys but the upper ranges (payments or gifts over $1 haven't been tested adequately). The improvements noted are increasing otherwise low return rates to moderate levels, (Kanuk and Berenson, 1975). Payments appear to induce more people to take a health examination although apparently they aren't large enough to overcome the fear of embarrassment which some women experience (Bryant et al., 1975). Compensation may help sustain cooperation in diary panel studies lasting more than two weeks (Ferber and Sudman, 1974). Payments are not cost effective in shorter diary panels (Walsh, 1976). The use of monetary incentives to increase response rates for in-person interviews has received little experimental attention. Dohrenwend (1971) reports no effect on a $5 honorarium on initial and repeat interview response rates in an urban area. On the other hand, Chromy and Horvitz (1974) show that a $5 to $20 payment can motivate more young adults in a national household survey to complete knowledge tests compared to a no-incentive condition. The latter research indicates that varying the amount paid by the number of test packages completed by the respondent is cost effective. This condition has been adopted in succeeding surveys.

More information about incentives may be obtained by consulting the Census Bureau bibliography (Survey Methodolgy Information Service, Undated-b), and Ferber and Sudman (1974).

Interviewers. Some interviewers are more successful than others at getting completed interviews. Barbara Bailar has done extensive work in this area. Her report (Bailar and Lanphier, 1977) comments on the important amount of variance in response rates contributed by interviewers. For a review of empirical studies, see Inderfurth (1972).

Respondent Rules. In many surveys it is possible to consider accepting responses from informants about sample person characteristics as a method of reducing costs and nonresponse. Kovar and Wright (1973) report the results of the Health Interview Study experiment requiring 100 percent self response. For the experiment, interviewers were asked to contact households initially at usual times. There was a possible slight variation in initial time of call (22 percent after 6 p.m. to self-responding households vs. 19 percent to other units) but it produced 74 percent of adults at home on the initial call (vs. 63 percent in the household respondent treatment). In both groups the household response rate was the same 96 percent. The person response rate (Z-rate) dropped from 99.8 percent to 98.7 percent in the self response treatment. 96 percent of the adults responded for themselves in the condition requiring it whereas 67 percent self responded when it was not required. Costs increased 17 percent (probably an upper bound estimate since interviewers were not allowed flexibility in timing initial calls).

When SRC switched its respondent designation procedures, response rates declined. Juster (1976) notes a 5 percent reduction when the designated respondent was changed from the household head (or spouse if necessary) to a specifically selected household member over age 17 (with proxy responses not permitted). In Census Bureau studies requiring self-respondents, the effect on response rates is to decrease them 1-2 percent (see for example, Love and Turner's discussion of the National Crime Survey, 1976). The difference between household response rates and the overall person response rates in the recent NORC medical access study ranged from 2 to 6 percent (M. Frankel, personal communication).

Organizational Features and Quality Control. The most parsimonious hypothesis accounting for differences in response rates is that some organizations structure field activities more effectively than others. If society has changed, these organizations have been able to adapt without a great change in efficiency. It is well beyond the scope of this paper to present comparative management analyses of the various organizations and projects kind enough to furnish response rate problems and some key organization features which may differ from the norm for nonfederal survey groups.

A major feature is that the bulk of Census Bureau sample research is with continuous, large scale studies. With ongoing programs it is possible to identify problems, try out alternative solutions, and implement them directly. Continuous studies enable the organization to maintain a large, permanent field staff who become experts in one particular kind of interview. Expertise cumulates so that successful procedures can be taught to the entire staff in refresher training sessions. Organizations serving the one-time, customized survey market (often not requiring national samples or requiring more staff in one of the PSU's than in normally available) are less able to maintain a large, permanent field staff and find it difficult to correct problems during short field periods. Studies are different enough so that it is not always possible to generalize effective procedures from one to another.

1. Noninterview Standards and Quality Control. The Census Bureau sets very high standards for response rates and features these requirements prominently in training sessions and quarterly reviews of individual interviewer performance. Performance reviews are based on observations of fieldwork, results of reinterviews, and a tabulation of questionnaire entry error rates (Greene, 1976). Interviewer productivity (and other performance) is rewarded by cash awards, promotion, salary increases and (if not full time) additional work opportunities. Poor performance results both in additional monitoring and being placed on probation. Probation time does not count toward within-grade tenure needed for automatic salary increases (Greene, 1976).

2. Work Facilitation by the Organization.* Periodic refresher training sessions are held to discuss reasons given by reluctant respondents and methods of responding to them. This procedure, of course, is maximally effective in continuous studies. The Bureau has the advantage of accumulated experience with ongoing surveys and can set (and enforce) realistic assignment sizes and completion deadlines for each sample area. It is not as easy to do this for one-time special studies whose field requirements and problems vary across surveys. The Bureau provides interviewers with information on the best times to make initial calls to locate particular kinds of respondents. On some studies, the interviewer is required to contact the entire assignment early in the field period. For most studies, the office is notified of potential refusals and takes specific steps to persuade the respondent to cooperate when the interviewer calls again. Donny Rothwell (personal communication) points out that the local office can invoke a law prohibiting doormen (etc.) from denying the access of a census interviewer to units in a building.

Walsh presents the fieldwork case history of the Census Bureau's consumer expenditure diary survey. Households made daily entries into an expense diary over a two week period. The pretest in the Chicago area suggested cooperation rate problems (around 50 percent). It also demonstrated the impracticality of placing diaries in households on specific days. For the first part of the main study, more intensive interviewer training was given and an experiment paying respondents conducted. During the first 3 months, total noninterview rate was over 25 percent with refusals at 11 percent and a large percentage (up to 8 percent) of households not contacted during the 6-day field periods. The incentive treatment ($5, $10) had small, positive but not statistically significant effects on the interview rate and was discontinued part way through the quarter. The usual exhortations to field offices and interviews to improve performance were made and it was hoped that cooperation would improve on the basis of increased staff experience and dropping the complexities of the remuneration experiment.

Second quarter rates, however, showed very little improvement. The only major change was a reduction in the temporarily absent (vacations, etc.) rate, reflecting, in part, the fact that fewer households are on vacation in the fall. Being a continuous (2-year) survey,

opportunities for action to improve performance were available. Steps taken included provision of additional information by the office to locate poorly defined sample addresses, increasing the field staff, extending the field period by one day, requiring the interviewer to contact someone in each sample household early in the field period, reporting refusals to the supervisor early for further action, and retraining interviewers (emphasizing refusal conversion strategies). Nonresponse dropped from 24 percent to 18 percent in the next quarter with the major declines in the nonrefusal (temporarily absent, unable to contact, other) components. Improvement continued over the succeeding 15 months with diary response rates around 90 percent and refusals in the 6-7 percent range. Central city response rates averaged about 88 percent.

CONCLUSIONS ABOUT CORRELATES OF RESPONSE RATES

If in-person interview response rates are not affected importantly by the current social changes, can we say how they are determined? From the foregoing discussion, it appears that seasonality and interview length don't have the uniformly strong, negative effects that some suspect. Hypothesized positive forces such as Census Bureau auspices, using advance letters, and offering payment aren't as powerful as we might hope. Callback strategies can account for a meaningful amount of variance and respondent rules have some effect on rates (e.g., 1-6 percent) and on costs (e.g., a household respondent rule can save as much as 17 percent over a complete self-response rule).

Interorganizational differences are large and not attributable entirely to differences on the above dimensions. Three additional dimensions may be responsible for the relative success of the Census Bureau: the long term nature of the studies it undertakes (creating the potential for improvement over time), things done by the regional office to facilitate fieldwork, and the extensive quality control activities which monitor a wide range of important performance variables for individual interviewers and provide both immediate reinforcement (positve and negative) and corrective action (e.g., retraining) as appropriate. These are applications of the most powerful principles offered by psychological theories of motivation and performance. While it will be much harder for nonfederal survey organizations to apply these principles, the investment may have a substantial payoff.

12

*Discussion based on Greene, 1976; Love and Turner, 1975; and Walsh, 1976.

DISCUSSION OF RESPONSE RATES

Charles Cannell, University of Michigan, Chair
Lois Monteiro, Brown University, Recorder

In the past few years there has been considerable speculation and public statement to the effect that survey response rates have been decreasing significantly (especially that refusals have increased) and that as a result the survey is in some serious danger as a method for producing high quality data. The paper which Marquis has prepared is most striking in its rejection of this conclusion. While response rates have shown some decline and the effort (and costs) required to maintain an acceptable rate has increased the general finding is that the problem is not as critical as many of us had been led to believe.

Response rates are an important indicator of the quality of a survey operation. A low rate increases the danger of bias in the survey data because non-respondents are likely to differ from those who do respond in characteristics important to the analysis, but although an adequate survey response rate is important, the rate should not be taken as the sole indicator of the survey's quality. Since response rate is the most readily quantified statistic there is a tendency to over-generalize from it to overall survey quality. In fact, there is evidence that while extreme efforts may achieve rate increase, it may, in some cases, be at the expense of response validity (Haberman, Fowler, and Cannell).

While refusal to participate in a survey may be the most overt indication of rejection of the respondent role, others who grant the interview may in fact reject the role by covertly refusing to answer questions truthfully, or by responding truthfully only when the task is easy, but failing to exert effort necessary to produce valid responses to more complex questions. Good survey performance requires not only a high response rate, but also conscientious task performance and complete accurate responses. We should be attending to both of these factors in evaluating survey quality.

DEFINITIONS OF RESPONSE RATES

There are considerable differences between investigators as to the precise definitions of components of response rates. The problems are apparent in personal and mail interviews and become horrendous when considering telephone surveys. Even in personal interviews alternate definitions result in substantial differences in rates. Bailer finds that definitional differences may affect the rate by as much as 25 percentage points.

The simple formula for calculating a response rate for a probability sample is:

$$\frac{\text{number of completed interviews}}{\text{number of potential respondents in the Universe}}$$

Questions arise as to when an address should be considered a non-sample and when a non-response; when is a house vacant and when a not-at-home; what about people in hospitals, on extended vacations; mentally incompetent, location of dwellings supposedly in the sample, etc. Questions arise also as to definition of what constitutes a completed interview. For some studies, missing critical variables result in the classification of the interview as a non-response. For others the rules are more relaxed and the interview is classified as complete with missing data for some items.

The major conclusion from this discussion is the recommendation that standard definitions be developed. The response rate differences observed in Marquis' paper can reflect differences in the establishment of eligible respondents. For some surveys the selected household respondent can be any responsible adult present at the time, whereas for others only one person in the household is eligible to be considered the respondent. Higher response rates are more easily obtained in the former situation.

RECENT EXPERIENCES WITH RESPONSE RATES

The massive number of refusals and the lowered response rates that were predicted to arise from the 1973 Privacy Act did not materialize, as Marquis' paper attests. However, there are some indications of greater difficulty in the field process and the need for increased field efforts to attain adequate response. A number of participants, among them Sudman and

Reeder, underscored the feeling that there have been no indications of an increase in the refusal rate, and no evidence that the public is tired of being surveyed, although non-located and not-at-homes have in the recent past presented difficulties in making contact with respondents. Some effect that was seen immediately after the disclosure to respondents of the right to refuse, was counteracted after interviewers were made aware of the importance of encouraging participation even though the respondent was informed of the option of non-response (Gerson).

Scharff presented detailed evidence of some national changes in response rates between 1969 and 1976 that show an increase in refusals over the period. The refusal rate in the current Medicare Survey for the aged sample panel interviewed from October 1969 to December 1970 ranged from 1.4 to 2.1 percent, whereas in 1975-1976 it was between 2.4 and 3.7 percent for the elderly. The rate would probably have been greater if it were not for a vigorous follow-up on refusals. It also appears that in the Current Medicare survey—1976 Series—the refusal rate among the elderly (2.4 to 3.7 percent) is higher than the refusal rate among the disabled population (1.9 to 2.6 percent). The lower refusal rates among disabled are offset, however, by a larger component indicating mobility. Table A summarizes the survey expenses for three years.

Dalenius indicated that in Western Europe the experience has shown a marked increase in non-response during recent years, and in some instances there has been a four-fold increase. He also cited Swedish experience, and told of a Norwegian experience in which a survey was halted because of response problems.

Eckerman questioned to some extent the generally positive findings on the response rate issue presented in the paper by Kent Marquis. He pointed out the evidence is not all in. In fact, one or two rather intensive studies of this issue, for which final results were not yet available, were referenced in the Marquis paper. Eckerman also cited a recent report on response rates for the Detroit Area Study over a 10-year span of time. These findings clearly indicated a gradual decline from response rates in the mid eighties to those averaging in the low seventies (Hawkins, 1977). In one specialized sample involving inner city respondents the rate actually dropped to 55 percent. Such findings do not contribute to a feeling of complacency about response rates. In addition, there may be decreasing response rates among certain U.S. sub-populations. The Alameda County surveys show only an approximate 5 percent decrease in total response rates between 1963 and 1973 but, while the response for the total sample was 88 percent in 1973, the response rate for Mexican Americans was 70 percent (Roberts, unpublished data).

CORRELATES OF NON-RESPONSE

A number of correlates of non-response were pointed out, and the group discussed possible strategies to offset some of these factors. Individual characteristics related to non-response have been reviewed extensively in market studies (Bridge). These show that age, marital status, and social class are related to non-response. Intelligence is also related as demonstrated by an armed forces release study (Bridge, 1974), that showed a difference in response rates among soldiers with different I.Q. levels, such that of persons with the highest I.Q.'s, 81.9 percent responded after one follow-up, but of those with lower I.Q.'s, 67.9 percent responded. One year after the discharge, 2.3 percent of those with high I.Q.'s could not be located, but of the low I.Q. group, 9.8 percent could not be found.

The Survey Research Center studies have had experience with a non-response form to separate refusals from other non-responses. They find no difference in response rate by race or sex, but do find slight income differences, such that the middle income groups ($7,000-$9,000) have a poorer response rate (Scott). Age is also related, particularly with older respondents in central cities who are hesitant to allow entry to strangers. Education is related to response rate, but is related to age also. Only rarely in the SRC experience, does the subject matter of the interview lead to a wide fluctuation in response rate. Marshall concurred that on a broad sample in the Denver area, non-response was related to age, marital status, number of children, and income.

Differences in response rates are also related to the location and type residence. In four national health surveys with the same field effort such differentials were persistent from 1953 to 1970 (Andersen, et al., 1976; Andersen and Anderson, 1967; Anderson and Feldman, 1956). Also, within cities, multiple and single unit structures yield different response rates, with multi-unit structures having a lower response rate. There is approximately a 33 percent decrease in rate if the residence has a restricted entry (Scott). In central city clustered samples there may be a possibility of lowered rates by word of mouth reports of the interviewing among units in the cluster. Elderly women in central cities yielded an especially low response rate, about 48 percent (Horvitz). That probably can be attributed to respondent fear of the consequences of allowing entry. Prescreening by means of prior phone contact can be of some help in gaining entry with this group.

The differential response rates between urban and rural areas has importance in national surveys (Cannell), for if rural areas are high in response, e.g. 95 percent rate in rural Kansas, while central city response is very low, the overall rate may mask the latter rate. In

14

Table A. Non Response in Current Medicare Survey (In Percent of Total Sample)

Panel	Total			Refusal			Moved		
	Oct.	Jun.	Dec.	Oct.	Jun.	Dec.	Oct.	Jun.	Dec.
Aged Oct. 1969 to Dec. 1970	5.8%	7.1%	7.7%	1.4%	2.2%	2.1%	2.6%	3.2%	3.6%
Aged Oct. 1970 to Dec. 1971	6.3%	7.2%	8.2%	1.5%	2.2%	2.5%	2.7%	2.9%	3.3%
Aged Oct. 1975 to Dec. 1976	7.4%	8.8%	9.5%	2.4%	3.7%	3.7%	3.0%	3.5%	3.8%
Disabled Oct. 1975 to Dec. 1976	8.8%	10.0%	11.1%	1.9%	2.6%	2.6%	4.4%	5.4%	6.0%

one housing survey conducted in both an urban and rural area, Greenbay, Wisconsin and South Bend, Indiana, the procedures were the same but there was a 10 to 15 percent response rate difference between the sites (Hensler).

In a 7-month panel survey to collect data on medical utilization and expenditures, the initial rates were 86 percent in largely rural Washington County, MD and 72 percent in the Baltimore City area. The portion of households completing all interviews during the seven month period was 77 percent in Washington County and 60 percent in the Baltimore area. Those differences were found between the two areas, despite a greater effort to get interviews in the Baltimore area (Shapiro, et al., 1976). Palit noted that in the Wisconsin area central city rates are about 69 percent whereas in the rural areas of the state the rate is in the low 80's. No difference in urban-rural rates were found in a study in which there were strict limits in field time (Marshall). In that effort, a response rate of 65 percent was obtained in both types of areas.

FACTORS INFLUENCING RATES

Experiences show that response rates may be affected by administrative factors such as the timing of interviews, or number of call-backs, as well as by factors such as the sponsorship of the survey, the saliency of the topic to respondent and the target population. Since, in addition to refusals, non-located or not-at-homes represent a major factor in non-response rates, the group discussed the apparent increase in respondents whom interviewers are unable to contact. Speculations about the sources of this phenomena included changes in family life styles—more meals eaten out, more women in the labor force, fewer people at home on weekends.

The major suggestions to counteract these were to shift interview times to conduct interviews after 3:00 P.M. and to allow for more call-backs and longer field times even though the last two procedures increase

field costs and the evening interviewing may make it more difficult to hire interviewers (Marshall). Eckerman suggested, as a means of perhaps reducing costs while maintaining both high response rates and high quality data, that multiple modes of interviewing be considered as a regular feature of large scale surveys. Typically grants or contracts involving surveys are undertaken via a single interviewing mode, e.g., personal face-to-face interviews, mail questionnaires or phone interviews. An optimal approach might be developed in many surveys by combining personal interviews, phone interviews, and mail questionnaires for subsets of the population under study, assuming there is sufficient sampling frame information available. This approach would also permit direct comparisons of the nature, quantity, and quality of information which proves to be collectable through each mode.

Sponsorship by a "legitimizing" agency was attributed to be the source of high response rates among Kaiser enrollees (Pope). In addition, de la Puente noted the success of a survey of pathologists when conducted under the auspices of the Cytopathology Association, and he attributed the success to this sponsorship.

Health care seems to continue to be a salient topic, even to the central city respondent, according to Shapiro, who remarked that health surveys in Baltimore had an 85-90 percent response rate. He added that the saliency must be made clear at the entry point of the interview with a clear and rapid explanation of the purpose, for some purposes are very important to people and will influence the willingness to respond.

The National Health Examination Survey presents some evidence on the importance of saliency to the attainment of high response rate (Bryant). In that survey the response rate for adults was 87 percent, for children 6-11 years old the rate was 96 percent, and for adolescents 12-17 years old the rate was 90 percent. During 1971-1973, a health and nutrition survey was conducted with a 28,000 person sample that included respondents aged 1 to 74 years. The response rate on

the 19 primary sampling units was only 68 percent. An experimental sub-survey of 600 persons conducted in San Antonio promised payment for the examination and the response was increased by about 12 percentage points in the experiment. The remuneration policy was instituted for the remainder of the sample, but the overall survey response rate still did not surpass 75 percent.

Target populations tend to differ in response. When agencies are the respondents, e.g., hospitals or agencies, delivering specific services, one must resolve special issues with respect to response. Scott (1974) reported a response of 90 percent when the hospital was the selected respondent and another M.D. who was qualified to report for the unit could respond if the originally designated M.D. was unavailable. Research Triangle Institute reports that in its hospital record abstract study 80-85 percent of the selected hospitals participated. The percent of records abstracted varied within hospitals, however, because of informed consent requirements for physicians and patients (Kalsbeek). On the negative side, there have been some problems in obtaining consent from some agencies. Some hospital associations are beginning to act as screening agents for their constituent members and one may need to work in cooperation with these associations to get responses from hospitals (Bryant). To some extent this issue is related to the respondent burden issue—overly burdened groups, such as hospitals or physicians have lowered response rates. Some physican groups may be overburdened as for example, physicians in geographic areas such as central cities where their numbers are few, and with such physicians response rates may be as low as 43 to 64 percent (Waksberg). Other physician groups have been found to be very cooperative (Colombotos, 1975). In a 1973 national probability sample of physicians, using a telephone interview, there was an overall response rate of 75-80 percent. The rate for hospital house staff was over 75 percent, and for medical students with a mail follow-up, the rate was 84 percent. In a longitudinal study of physicians in New York State (calls lasting 45 minutes to 1 hour) the initial rate was 80 percent (1964) with about 80 percent response obtained from those sample members remaining in successive waves of interviews. Thus, generally for this group, with a telephone mode, a 75 to 80 percent rate is achieved. Similar findings among general practice dentists and pedodontists were obtained in a 1975 survey (Roberts, forthcoming) that achieved a 75 percent response rate by mail with telephone follow-up from pedodontists and 69 percent from general practitioners.

In a Health Care Financing Survey of the costs of physicians' office administration, there was a similar total response rate (70 percent of eligible physicians) but of those responding there was a significant amount of non-response on such key data items as salaries for office personnel (Jabine). This type of non-response was discussed and concern was voiced about the quality of so-called completed interviews when data is missing on crucial questions within the interview. Particular item refusals may differ by 15 percent from total response rates (Springer, 1976). Waksberg noted that for the Current Population Survey there is a 95 percent response rate for many items but an 80-85 percent response for income, with many refusals for that particular item. The solution to such problems may be to eliminate the incomplete items.

RESPONSE RATES IN TELEPHONE SURVEYS

This topic was considered separately as a special problem for response rate calculation. For telephone surveys using list samples of special groups, the calculation of response rates is relatively uncomplicated, but the problem becomes great when the sample is based on random digit dialing, for in that case one cannot determine when the unanswered phone is an assigned number. Since there are differences in various areas in what happens with unassigned numbers, it is difficult, costly, and sometimes impossible to determine whether an unanswered ring is because no one is at home or because the phone is not a working number. The former would be included in the denominator of the response rate equation, while the latter would not. As with response rates for personal interviews and mail surveys, there is a great need for standard definitions and methods of calculating the rate.

Bradburn reported on a complex telephone survey which combined telephone and mail methods, and used a telephone screening process to determine persons in scope, members of 5 certain ethnic groups. The screening process of 50,429 numbers yielded 6.1 percent who were in scope. The overall disposition rate for the screening process was 99.3 percent. Those found to be in scope were sent a mailed questionnaire, to be completed by an adult and a child in the household. Of those who on the telephone agreed to participate, 88.7 percent returned both questionnaires. The success of this mailed portion was attributed to a decision to compensate both the child and adult respondent by sending a new two dollar bill to each of them along with the questionnaire, for a pretest without the money compensation yielded only a 39 percent completion. There was no difficulty in getting street addresses from the persons once they were contacted by random dialing. Bradburn reported similar results in obtaining addresses for respondents when a screening was done by random dialing. When asked what percent of those contacted by phone did not give their name and address, Bradburn reported that in the latter study 4

percent refused to give either names or enough information for a complete screen and 18 persons out of 144 gave names but refused to answer the screening questions on the telephone. Discussion on the use of a screening telephone contact was continued by Waksberg who mentioned that in a random digit study that involved a screen followed by an interview there was a 10 to 1 ratio of screening calls to eligible interviews. But another study (Eisinger) involving a random digit dial screening with telephone follow-up, conducted for the Department of Defense required 5,000 calls to get 200 completed interviews. Eisinger noted that a further problem with such a method is that if the respondent refused on the random digit call there is no further recourse, because of no address, to try to contact the person by another route.

One method to make random digit dialing more efficient is to block the numbers (Waksberg). It is then possible to determine if a number is an active residential number. The telephone company has cooperated somewhat on cases in which there were mysterious sounds by confirming which of these were working numbers. This method cuts down considerably on the unknown, no response cases. Nevertheless, the group generally agreed that the issues of response rates for telephone interviews need further refinement and experimentation.

SUMMARY: MAJOR ISSUES CONSIDERED IN THE SESSION

Cannell, in his opening remarks to the group, suggested that the focus for the session should be a state of the art discussion on the correlates of non-response, the basis for non-response and the strategies to counteract non-response. All of these topics received some attention during the session and additional major issues were discussed. These included the problem of standardization of the calculation of the response rate with particular emphasis on the components of the denominator, expecially in random digit telephone surveys; the issue of cost effectiveness in maintaining response rates in the high 80's or 90's during field interview surveys; the need for better information on sources of non-response and strategies to counteract the non-response tendency. The group at several points expressed concern at the need to maintain the *quality* of the data and to avoid allowing the response rate to mask lower quality data.

The group generally agreed with Marquis' findings that adequate response rates have been maintained over recent years, but nevertheless expressed the need for continued watchfulness and reaction to the possibility of a gradual slippage in response quality and rate in the future.

TOTAL SURVEY DESIGN: EFFECT OF NONRESPONSE BIAS AND PROCEDURES FOR CONTROLLING MEASUREMENT ERRORS

William Kalsbeek, Research Triangle Institute
Judith Lessler, Research Triangle Institute

INTRODUCTION

Total Survey Design (TSD) conceptually refers to an allocation of survey resources so that known components of survey error can be quantified and acknowledged in a survey protocol which collectively minimizes the Total Survey Error (TSE) of estimates. In addition to the usual survey activities (e.g., data collection, processing, and analysis), applying the TSD concept requires a rational allocation of often fixed total survey resources to several other activities including: (1) construction of reasonable TSE and related cost models for principal survey estimates, (2) quantification of error components and cost model components, and (3) development of a TSE-minimizing survey protocol.

Models recognizing the various components of TSE have existed for some time (see, for example, Kish, 1965). Although these models have commonly isolated components attributable to sampling error, nonresponse bias, measurement variance, measurement bias and others, most attention in the theory of design and estimation from sample surveys has focused on the sampling error component alone. Quite obviously this view can be tolerated if nonsampling error components are insignificant relative to sampling error. If they are not insignificant, the usefulness of analyses from surveys may be seriously undermined.

This paper has two objectives: (1) to illustrate the need for practicing TSD and (2) to illustrate how TSD can be applied. Specific attention is given to nonresponse bias and measurement bias components of TSE. In the following two sections these are discussed in separate examples from existing surveys.

AN ASSESSMENT OF THE IMPACT OF NONRESPONSE BIAS

In this section of the paper we intend to demonstrate the importance of monitoring sources of TSE other than the component attributable to sampling error. In particular, nonresponse bias is discussed, estimators for several nonresponse bias measures are developed, and some findings from an existing nonresponse study are presented.

Prominent among recognized components of TSE is nonresponse bias. Viewed simply, it is defined as the difference between the expectation of an estimator when applied only to respondents and the "true" value of the parameter of interest over the total population of respondents and nonrespondents. If a population mean $\bar{Y}_0$ is estimated by $\bar{y}_1$, then the bias of the estimator can be represented as

$$\text{Bias}(\bar{y}_1) = E(\bar{y}_1) - \bar{Y}_0 \qquad (2.1.1)$$

The population mean $\bar{Y}_0$ can be expressed as

$$\bar{Y}_0 = P_1 \bar{Y}_1 + P_2 \bar{Y}_2, \qquad (2.1.2)$$

where P_1 is the proportion of population respondents, P_2 is the proportion of population nonrespondents (so that $P_1 + P_2 = 1$), $\bar{Y}_1$ is the true mean for all respondents, and $\bar{Y}_2$ is the true mean for all nonrespondents. Thus,

$$\text{Bias}(\bar{y}_1) = P_2 \Delta + \lambda, \qquad (2.1.3)$$

where $\Delta = \bar{Y}_1 - \bar{Y}_2$ and $\lambda = E(\bar{y}_1) - \bar{Y}_1$.

Equation (2.1.3) indicates that the magnitude of nonresponse bias under these circumstances is directly related to the proportion of nonrespondents, the difference of population means between respondents and nonrespondents, and the difference between the expected value of the estimated mean for respondents and the true mean for respondents.

It has become standard practice for better surveys to report levels of nonresponse. This information can be of some use to both analyst and user of survey data when the effect of nonresponse on analyses must be reconciled. If the level of nonresponse is high, one might anticipate a potentially significant contribution to TSE from nonresponse bias. This is, however, only a partial indication of the magnitude of nonresponse bias. To better quantify the effect of nonresponse on TSE, an accounting of the size of other components of the bias attributable to nonresponse must be made.

One might expect that Δ from equation (2.1.3) varies among surveys and that its size is a function of

the survey population addressed and the survey protocol employed. For example, if a survey population is heterogeneous with respect to the random variable Y and the association between Y and the response inclinations of members of the survey population is high, then one might expect Δ to be large. If, on the other hand, the survey population exhibits little variability with respect to Y, the size of Δ would probably be relatively smaller. The size of λ largely depends upon the imputation procedure that is used to adjust for nonresponse in producing $\bar{y}_1$.

The remainder of the discussion in this first part of the paper centers on estimating nonresponse bias in an existing survey. A model is developed which combines all components of equation (2.1.3). Findings from this special study indicate that the effect of nonresponse bias on interval estimates of proportions can be substantial.

Quantifying Nonresponse Bias in the National Assessment of Education Progress Study

Background. Since its inception in 1969 the National Assessment of Educational Progress (NAEP) study has conducted an assessment of learning retention by school-age children and young adults. Each year since that time, standardized subject matter exercises have been administered to a sample of individuals in the 9-year-old, 13-year-old, 17-year-old, and 26-to-35-year-old age groups. The NAEP sample is a three-stage probability sample consisting of two parts. The first part is a sample of students enrolled in elementary, or secondary schools, termed the in-school sample, and the second part is a household sample, termed the out-of-school sample.

The nonresponse study described below involved administration of exercises to a subsample of certain in-school sample nonrespondents or "No-Shows" from Year 04 (fourth year of assessment covering the 1972-1973 school year). More specifically, the subsample of No-Shows was from those 17-year-olds who were selected for the Year 04 "regular" NAEP assessment but who failed to appear for exercise administration. The No-Show study was limited to the 17-year-old age group since it had recorded the highest level of nonresponse (26 percent) in Year 03. The principal motivation for this No-Show study was a desire to view and to quantify the impact of nonresponse bias upon regular assessment analyses from this age group so that adjustments to regular assessment estimates could be made.

During any NAEP assessment year, two or three different subject matter areas are assessed. The subject matter areas for Year 04 were mathematics and science. Exercises are grouped together into packages some of which are administered in a group and some of which are administered individually. Every package in Year 04 contained a mixture of mathematics and science exercises. From the set of Year 04 packages for 17-year-olds, three group-administered packages and

one individually administered package were arbitrarily selected for the No-Show study. The group packages were numbered as 01, 03, 09; the individual package was numbered 13. It is by means of these packages that the nonresponse bias in the NAEP-reporting of student performance was assessed.

In order to understand the method by which the nonrespondent sample was selected, it is necessary to briefly explain the Year 04 sampling design. Year 04 Primary Sampling Units (PSU's) were composed of counties or groups of contiguous counties. PSU's were first stratified by region, size of community, and socioeconomic characteristics and then selected using probabilities proportional to the population of the sampling unit. A total of 118 PSU's were selected by this procedure. The secondary sampling units consisted of public and private schools within selected PSU's. Stratification of the secondary units by income characteristics and size of school took place before selection. Schools were selected using probabilities proportional to the estimated number of eligibles in each school. The tertiary sampling units were students who were enrolled in sample schools, who met certain age requirements, and who were not ineligible for any other reason.[1]

Year 04 PSU's were classified into two heterogeneous clusters. The clusters were constructed so as to be well-balanced with respect to region, size of community, and socioeconomic characteristics. One cluster was then randomly selected for the No-Show study using equal probabilities. The No-Show primary sample was thereby composed of 57 PSU's. Since the desirable condition of having two PSU selections per stratum did not hold by this arrangement, 27 pseudo-strata were formed by sequentially pairing the No-Show PSU's according to region and size. Finally, since 57 is an odd number, one pseudo-stratum was assigned three PSU's.

Eligible schools in the No-Show secondary stage sample consisted of all 17-year-old sample schools in No-Show PSU's in which at least one of the No-Show packages had been administered during regular National Assessment. Within selected schools, students eligible for the nonresponse study were selected for a particular No-Show package on a matched-sample basis. That is, all students who were originally selected for a group-administered package but who had not appeared for assessment were eligible for any of the No-Show group packages administered in the school during 17-year-old assessment. Similarly, any student who was selected for an individually administered package but who had failed to appear for regular assessment was eligible for the No-Show individual package, provided that the same package had been administered in the school during 17-year-old assessment. This matched sampling procedure was adopted so that the

[1]*Individuals who are emotionally or mentally retarded, functionally disabled, non-English speaking, or nonreaders, are excluded from the NAEP sample.*

analysis of differences between respondents and non-respondents could be viewed on a within-school basis. Eligible students were selected for specific No-Show packages using cyclic systematic sampling. These procedures produced a subsample of 2,771 students from the 7,725 17-year-old Year 04 No-Shows.

Attempts were made to contact selected individuals in school over a 3-week period following the regular assessment. Of the 2,771 students selected for the in-school portion of the study, 34 were determined to be ineligible; a total of 1,990 students out of the 2,737 who were eligible and selected were assessed; thus, the response rate for the in-school portion of the nonresponse study was 72.7 percent. At the end of the 3-week period, the names and addresses of all individuals who had not been contacted were requested from the schools. Several schools refused to release this type of information; however, names and addresses were obtained for 598 of the 747 eligible in-school nonrespondents. A systematic subsample of 130 of No-Show study nonrespondents was selected for the out-of-school portion of the No-Show study. During the out-of-school phase of the study, selected individuals were encouraged to take all four No-Show packages and were given an incentive payment of five dollars for each package which they completed. Ten of the individuals selected for the out-of-school portion of this study were determined to be ineligible. The total number of out-of-school respondents was 102; thus, a response rate of 85 percent was achieved during the out-of-school portion of the No-Show study.

Estimation of Bias Measures. Data collected in the No-Show study could be used to estimate several bias measures. Two of these measures were the bias as defined in equation (2.1.1) and the relative bias (or Rel-Bias) where for $\bar{y}_1$

$$\text{Rel-Bias } (\bar{y}_1) = \text{Bias } (\bar{y}_1) / \bar{Y}_o . \quad (2.2.1)$$

The methods used to estimate Bias $(\bar{y}_1)$, Rel-Bias $(\bar{y}_1)$, and associated measures of precision are generally stated below although a more detailed discussion of these methods is presented in a following section, Estimation of Bias and Rel-Bias.

Because of a suspected difference between in-school and out-of-school No-Shows, separate estimators involving all No-Shows and only in-school No-Shows were developed for bias and rel-Bias. For both types of No-Shows procedures for estimating the bias and rel-bias were similar and can be summarized by the following steps:

1. The quantities F_a and C_a were defined as

$$F_a = \sum_{j\epsilon\Omega} E_{oj} P_{aj} \bar{Y}_{aj} ; \quad (j=1,2)$$

and

$$C_a = \sum_{j\epsilon\Omega} E_{oj} P_{2j} \bar{Y}_{aj} ; \quad (j=1,2)$$

where E is the number of eligible students, P is the population proportion of eligible regular assessment students, $\bar{Y}$ is the proportion of exercises answered correctly, Ω is the set of all eligible schools, the subscript j refers to schools, and the subscript α refers to the total population (o), regular assessment respondents (1), or No-Shows (2).

2. The bias of equation (2.2.1) was defined as a function of C_1, C_2, and E_o where the latter is the total number of eligible students in all eligible schools.

3. The rel-bias of equation (2.2.1) was defined as a function of F_α and C_α ($\alpha=1,2$).

4. Using regular assessment and No-Show data, E_o, F_α, and C_α were estimated from which the estimators bias $(\bar{y}_1)$ and rel-bias $(\bar{y}_1)$ were derived.

Approximate variances for bias $(\bar{y}_1)$ and rel-bias $(\bar{y}_1)$ were formed by the so-called jackknife technique since both estimators were nonlinear. For both variance approximations this involved computation of PSU "contributions." The final form of var{bias$(\bar{y}_1)$} and var{rel-bias$(\bar{y}_1)$} was then expressed as a function of squared paired differences of the PSU contributions.

To assess the significance of the bias and rel-bias estimators, it was assumed that

$$T = \frac{\text{bias } (\bar{y}_1)}{[\text{var\{bias } (\bar{y}_1) \}]^{1/2}}$$

and

$$T' = \frac{\text{rel-bias } (\bar{y}_1)}{[\text{var\{ rel-bias } (\bar{y}_1) \}]^{1/2}}$$

are distributed as a "Student's" t-statistic with 29 degrees of freedom. Under this assumption, a significance level of 0.05 is indicated when $|T| \geq 2.045$ or $|T'| \geq 2.045$.

Two other bias measures were established and estimated. One was a so-called "bias-ratio" of $\bar{y}_1$ which was defined as

$$\text{BR } (\bar{y}_1) = \text{Bias } (\bar{y}_1) / \text{SE } (\bar{y}_1) , \quad (2.2.2)$$

where $\text{SE } (\bar{y}_1) = [\text{Var } (\bar{y}_1)]^{1/2}$

The entry of equation (2.2.2) was estimated by

$$\text{br } (\bar{y}_1) = \text{bias } (\bar{y}_1) / \text{se } (\bar{y}_1) , \quad (2.2.3)$$

where $\text{se } (\bar{y}_1) = [\text{var } (\bar{y}_1)]^{1/2}$

The estimator, bias $(\bar{y}_1)$, was taken from those bias estimates which involved all No-Shows. Values of var $(\bar{y}_1)$ were those separately derived for Year 04 regular assessment exercise data.

A final bias measure was the effective confidence level of interval estimates of $\bar{Y}_o$ when $\bar{y}_1$ (i.e., regular assessment data) is used. Interval estimates from regular assessment data take the form

$$\left[\bar{y}_1 - t_{1-a/2}^{(n)} \text{ se } (\bar{y}_1) , \bar{y}_1 + t_{1-a/2}^{(n)} \text{ se } (\bar{y}_1) \right] \quad (2.2.4)$$

where $t_{1-a/2}^{(n)}$

is a "Student's" t-statistic with n degrees of freedom and $(1-\alpha)$ is the resulting assumed confidence level of the interval described in equation (2.2.4). Assuming that $\bar{y}_1$ follows a "Student's t distribution with n degrees of freedom, it can be stated that

$$\Pr\left\{\left[\bar{y}_1 - t_{1-a/2}^{(n)} \; SE(\bar{y}_1)\right] \le E(\bar{y}_1) \le \left[\bar{y}_1 \right.\right.$$

$$\left.\left. + t_{1-a/2}^{(n)} \; SE(\bar{y}_1)\right]\right\} = 1 - a . \quad (2.2.5)$$

However, the stated intention was to estimate $\bar{Y}_0$ so that the effective confidence level of the interval defined by equation (2.2.4) was

$$\delta = \Pr\left\{[\bar{y}_1 - t_{1-a/2}^{(n)} \; SE(\bar{y}_1)] \le \bar{Y}_0 \quad (2.2.6)\right.$$

$$\left. \le [\bar{y}_1 + t_{1-a/2}^{(n)} \; SE(\bar{y}_1)]\right\}$$

$$= \Pr\left\{[-t_{1-a/2}^{(n)} - BR(\bar{y}_1)] \le [\bar{y}_1 - E(\bar{y}_1)]\right.$$

$$\left. / \; SE(\bar{y}_1) \le [t_{1-a/2}^{(n)} - BR(\bar{y}_1)]\right\} \le 1 - a .$$

Since n associated with $\bar{y}_1$ was sufficiently large, the effective confidence level derived from equation (2.2.6) could be estimated as

$$\hat{\delta} = \Phi\{1.96 - br(\bar{y}_1)\} -$$

$$\Phi\{-1.96 - br(\bar{y}_1)\} , \quad (2.2.7)$$

where $\Phi\{\cdot\}$ is the cumulative distribution function for the standardized normal distribution.

Findings From the NAEP Study

Measures of bias as discussed previously were developed for individual exercises in NAEP packages 01, 03, and 09 which were administered to groups of students, and package 13, which was administered to individual students. These packages consisted of both mathematics and science exercises designed to test students' level of understanding in these areas. The number of exercises varied among packages and the nature of exercises was somewhat different with individually-administered package 13 as opposed to group packages 01, 03, and 09. Package 13 contained more multipart items with an assortment of skip patterns due to the one-to-one nature and more direct involvement of the exercise administrator. Table 1 indicates the number of exercises analyzed by package and subject matter.

Table 2 presents the percent of exercises with positive estimated biases. In most categories the majority of biases are positive indicating a general overestimation of $\bar{Y}_0$ by $\bar{y}_1$. The tendency to overestimate appears to be less common among group packages when only in-school No-Shows are used. The reverse of this tendency appears to be the case with individual package 13.

Table 3 presents the percent of significant biases according to the test for bias suggested in the preceding section, Estimation of Bias Measures. For group

Table 1. Number of Exercises Per Package by Subject Matter for NAEP No-Show Study Packages Administered to Students in the 17-Year-Old Age Group

Package	Exercise Subject Matter Type		Total
	Mathematics	Science	
01	16	23	39
03	19	13	32
09	19	28	47
13	22	7	29
Total	76	71	147

packages the percentages vary among packages and subject matter but are higher when all No-Shows are involved. For package 13 percentages are higher when only in-school No-Shows are involved. These findings seem to indicate that in-school No-Shows are more nearly similar to regular assessment respondents than are out-of-school No-Shows.

Table 4 presents the smallest and largest values of estimated rel-biases by package and subject matter. These data indicate the general magnitude and variability of biases relative to estimate of $\bar{y}_1$. Ranges tend to be somewhat greater when all No-Shows are involved as opposed to when only in-school No-Shows are involved.

In Table 5 median absolute bias-ratios are presented by subject matter and package. The magnitude of these data indicate a substantial impact of nonresponse bias relative to sampling errors particularly with group packages when all No-Shows are involved. As findings from Table 3 also indicate, this impact appears to be generally greater for group packages when only in-school No-Shows are involved.

Table 6 presents a comparison of the proportion of significant bias and median absolute bias-ratios by the difficulty of exercises across all packages. These data indicate some variability according to exercise difficulty but no new pattern is apparent.

Table 7 records the distribution of effective confidence levels for interval estimates of $\bar{Y}_0$ using $\bar{y}_1$ for assumed 95 percent confidence intervals from regular assessment. Clearly, resulting effective levels vary considerably, and the impact of the bias (i.e., lower effective confidence levels) appears to be somewhat greater among mathematics than science exercises.

A number of points are made concerning the findings discussed above. First, the magnitude and resulting impact of nonresponse bias (and other sources of nonsampling error) can be great and, in many circum-

Table 2. Percent of Exercises With Positive Biases by Subject Matter for NAEP No-Show Study Packages Administered to Students in the 17-Year-Old Age Group

No-Shows Involved:	Package	Exercise Subject Matter Type		Total
		Mathematics	Science	
All:				
	01	93.8	91.3	92.3
	03	100.0	100.0	100.0
	09	100.0	92.9	95.7
	13	81.8	57.1	75.9
	Total	93.4	90.1	91.8
In-School Only:				
	01	68.8	73.9	71.8
	03	89.5	76.9	84.4
	09	73.7	46.4	57.4
	13	90.9	71.4	86.2
	Total	81.6	63.4	72.8

stances, too great for the survey investigator to ignore. Second, the impact of nonresponse bias upon individual estimates from surveys is not completely predictable and may vary considerably within a given survey according to the "environment" in which the estimates are produced (e.g., with NAEP No-Show Study data: package, subject matter, size of the group in which administration occurs, and exercise difficulty). Third, contributions to the impact of nonresponse bias may come more heavily from certain categories of nonrespondents than others (e.g., with NAEP No-Show Study data: out-of-school No-Shows).

Results of the NAEP Year 04 No-Show Study were incorporated into recommendations for the survey protocol of subsequent years. Since No-Show data implied that a relatively heavier contribution to nonresponse was attributable to out-of-school No-Shows, an out-of-school followup of a subsample of regular assessment in-school nonrespondents was suggested.

This followup could not be implemented, but a followup of regular assessment in-school No-Shows was recommended and adopted. Procedurally, it involved an in-school next-day followup of selected 17-year-olds who were not present on the day of the regular assessment session. The next-day followup of No-Shows increased the 17-year-old response rate to over 80 percent.

Loss Incurred by Failure to Follow Up No-Shows

The importance of doing a followup study of nonrespondents can be illustrated by a simple example. Let us suppose that one of the following two alternative protocols must be followed:

Followup Alternative. Select a sample of m_0 students of which m_1 respond to regular assessment. Select a subsample of m_2 of the $m_0 - m_1$ nonrespondents to regular assessment and locate all of them for a followup assessment. Determine the bias of regular assessment estimates and adjust these estimates for the detected bias.

No-Followup Alternative. Select a sample of m_0^* students of which m_1^* respond to regular assessment. Produce biased regular assessment estimates but do not follow up nonrespondents.

The total field cost for the followup alternative can be expressed as

$$C = C_0 m_0 + C_1 m_1 + C_2 m_2 \qquad (2.4.1)$$

and for the no-followup alternative as

$$C^* = C_0 m_0^* + C_1 m_1^* \quad , \qquad (2.4.2)$$

where C and C^* are the total field costs for followup and no-followup alternatives, respectively, C_0 is the unit cost for all sampled students regardless of re-

Table 3. Percent of Exercises With Significant Biases by Subject Matter for NAEP No-Show Study Packages Administered to Students in the 17-Year-Old Age Group

| No-Shows Involved: Package | Exercise Subject Matter Type | | Total |
	Mathematics	Science	
All:			
01	75.0	47.8	59.0
03	94.7	76.9	87.5
09	15.8	10.7	12.8
13	4.5	0.0	3.4
Total	44.7	33.8	39.5
In-School Only:			
01	12.5	17.4	15.4
03	21.1	7.7	15.6
09	0.0	7.1	4.3
13	22.7	0.0	17.2
Total	14.5	9.9	12.2

sponse status, C_1 is the unit cost to complete and process regular assessment respondents, and C_2 is the unit cost to complete and process nonrespondents in the followup study.

If y_0 were an estimator for the population mean Y_0, if response status were known in advance of selection, and if m_1, m_2, and m_1^* were sizes for simple random samples, then neglecting finite population corrections, the mean squared error for y_0 for the followup alternative would be

$$MSE(\bar{y}_0) = P_1^2 S_1^2 / m_1 + (1-P_1)^2 S_2^2 / m_2 , \quad (2.4.3)$$

where P_1 is the proportion of respondents to regular assessment, $S_1^2 = Y_1 (1 - Y_1$ is the unit variance for respondents, and $S_2^2 = Y_2 (1 - Y_2)$ is the unit variance for nonrespondents. Kish (1965) has shown that the value of

$$k = m_2 / m_0 - m_1 ,$$

which minimizes equation (2.4.3) for fixed C, can be expressed as

$$k_{opt} = \left(\frac{S_1}{S_2}\right)\left(\frac{P_1 C_2}{C_0 + P_1 C_1}\right)^{1/2} \quad (2.4.4)$$

The mean squared error of y_0 for the no-followup alternative would be

$$MSE^*(\bar{y}_0) = S_1^2 / m_1^* + \left\{ \text{Bias}(\bar{y}_0)\right\}^2 . \quad (2.4.5)$$

To express the relationship between mean squared errors resulting from the followup and no-followup alternatives, a loss ratio could be computed as

$$LOSS = MSE^*(\bar{y}_0) / MSE(\bar{y}_0) . \quad (2.4.6)$$

LOSS values greater than one indicate an advantage to the followup alternative. LOSS values less than one indicate an advantage to the no-followup alternative.

Table 8 presents values of k_{opt}, m_0, m_1, m_2, and LOSS for different absolute values of Bias (y_0) given the following parameters for two hypothetical studies:

Parameter	High-Budget Study	Low-Budget Study
$C (=C^*)$	$200,000	$20,000
C_0	$2	$2
C_1	$10	$10
C_2	$50	$50
Y_1	0.50	0.50
P_1	0.75	0.75

Data from Table 8 indicate that, for the assumed parameter levels, LOSS values become dramatically large

Table 4. Smallest and Largest Absolute Values of Rel-Biases by Subject Matter for NAEP No-Show Study Packages Administered to Students in the 17-Year-Old Age Group

No-Shows Involved:	Package	Exercise Subject Matter Type			
		Mathematics		Science	
		Smallest	Largest	Smallest	Largest
		(Range indicated parenthetically)			
All:					
	01	0.0109	0.0869 (0.0760)	0.0022	0.0996 (0.0974)
	03	0.0275	0.1452 (0.1177)	0.0194	0.0965 (0.0771)
	09	0.0161	0.1126 (0.0965)	0.0001	0.0903 (0.0902)
	13	0.0049	0.0488 (0.0439)	0.0052	0.0627 (0.0575)
In-School Only:					
	01	0.0032	0.0626 (0.0594)	0.0073	0.0810 (0.0737)
	03	0.0009	0.0783 (0.0774)	0.0026	0.0474 (0.0448)
	09	0.0004	0.0561 (0.0557)	0.0010	0.0782 (0.0772)
	13	0.0001	0.0452 (0.0451)	0.0022	0.0253 (0.0231)

as the size of the bias increases despite the larger sample size with the no-followup alternative. This is particularly true for the high-budget study since the second term of equation (2.4.5) tends to dominate the LOSS value of equation (2.4.6) as sample sizes increases due to the availability of more money for field costs.

ESTIMATION OF BIAS AND REL-BIAS

Formulation in this section applies generally to individual exercises completed in the No-Show study. It is an adaptation of methodology for estimating bias associated with total package performance (Folsom, 1974). A symbol is intended to define an entity, while the attached subscript serves to determine its domain of applicability. A block symbol refers to a random variable, and a script symbol refers to a parameter. An upper case script symbol refers to a parameter for the population of all units, and the corresponding lower case script symbol refers to an estimate of the parameter associated with a sample of these units.

Specifically, we define

$$Y = \begin{cases} 1 & \text{if the individual exercise is answered} \\ & \text{correctly during administration} \\ 0 & \text{if otherwise} \end{cases}$$

$P(p)$ = population proportion of eligible regular assessment students,

$\bar{Y}(\bar{y})$ = proportion of exercises answered correctly,

$E(e)$ = number of eligible students.

The first-position subscript (α) associated with the above symbols refers to the total population (o), regular assessment respondents (1), and nonrespondents or No-Shows (2). Population totals F_α and C_α refer to the quantities

$$F_a = \sum_{j \in \Omega} E_{oj} P_{aj} \bar{Y}_{aj} \; ; \; (a=1,2)$$

$$C_a = \sum_{j \in \Omega} E_{oj} P_{2j} \bar{Y}_{aj} \; ; \; (a=1,2)$$

Table 5. Median Absolute Bias-Ratio by Subject Matter for NAEP No-Show Study Packages Administered to Students in the 17-Year-Old Age Group

No-Shows Involved:	Package	Exercise Subject Matter Type	
		Mathematics	Science
All:			
	01	1.64	1.38
	03	2.15	1.17
	09	1.29	1.61*
	13	0.59	–**
In-School Only:			
	01	0.71	0.70
	03	0.73	0.51
	09	0.39	0.55*
	13	0.85	–**

*The bias-ratio could not be estimated for one package 09 science exercise since $se(\bar{y}_1)$ data were unavailable.

**Bias-ratios could not be estimated since $se(\bar{y}_1)$ data were unavailable.

which are estimated by f_α and c_α respectively. These quantities will be combined to assess the magnitude of nonresponse bias in NAEP regular assessment statistics.

The following symbols are used in subscripts and as other notation:

- h = pseudo-stratum,
- i = PSU within pseudo-stratum,
- j = school within PSU,
- k = student within school,
- m = number of eligible sample students taking a package,
- w = package sample nonresponse adjusted weight (i.e., adjusted inverse of the probability of selection into the study),
- Ω = set of all eligible schools,
- ω = sample set of eligible schools,
- $+$ = summation over all possible subscript values.

Measures of bias and associated levels of precision were estimated using the following approach.

First, note that the "true" value of $\bar{Y}$ is

$$\bar{Y}_o = \frac{\sum_{j\epsilon\Omega} E_{oj} \bar{Y}_{oj}}{E_o}$$

$$= \frac{\sum_{j\epsilon\Omega} E_{oj} (P_{1j} \bar{Y}_{1j} + P_{2j} \bar{Y}_{2j})}{E_o} ;$$

$$E_o = \sum_{j\epsilon\Omega} E_{oj} .$$

If one lets

$$\bar{y}_1 = \frac{\sum_{j\epsilon\omega_1} \sum_{k=1}^{m_{1j}} w_{1jk} Y_{1jk}}{\sum_{j\epsilon\omega_1} \sum_{k=1}^{m_{1j}} w_{1jk}}$$

then the expectation of the estimator $\bar{y}_1$ can be expressed as

$$E(\bar{y}_1) = \frac{\sum_{j\epsilon\Omega} E_{oj} \bar{Y}_{1j}}{E_o} .$$

Thus,

$$\text{Bias}(\bar{y}_1) = E(\bar{y}_1) - \bar{Y}_o \qquad (A.1)$$

Table 6. *Percent of Exercises With Significant Biases and Median Absolute Bias-Ratio by Subject Matter According to Exercise Difficulty for Combined NAEP No-Show Study Package Exercises Administered to Students in the 17-Year-Old Age Group*

No-Shows Involved:	Exercise Difficulty*	Exercise Subject Matter Type			
		Mathematics		Science**	
		Percent Significant Biases	Median Absolute Bias-Ratio	Percent Significant Biases	Median Absolute Bias-Ratio
		(Number of Exercises in Parenthesis)			
All:					
	Easy	43.2 (44)	1.99 (44)	37.8 (37)	1.76 (37)
	Difficult	46.9 (32)	1.18 (32)	38.5 (26)	1.03 (26)
In-School Only:					
	Easy	13.6 (44)	0.71 (44)	16.2 (37)	0.70 (37)
	Difficult	15.6 (32)	0.58 (32)	3.8 (26)	0.56 (26)

*"Easy" exercises were those with $\bar{y}_1 \geqslant 0.5$; "difficult" exercises were those with $\bar{y}_1 < 0.5$.

**Figures exclude one package 09 and all package 13 science exercises since $se(\bar{y}_1)$ data were unavailable.

$$= \frac{C_1 - C_2}{E_0}$$

since $P_{2j} = 1 - P_{1j}$ by definition.
Similarly,

$$\text{Rel-Bias } (\bar{y}_1) = \frac{\text{Bias } (\bar{y}_1)}{\bar{Y}_0} \qquad \text{(A.2)}$$

$$= \frac{C_1 - C_2}{F_1 + F_2}$$

Ratio-type estimators were used to estimate values associated with equations (A.1) and (A.2):

$$\text{bias } (\bar{y}_1) = \frac{c_1 - c_2}{e_0} = \frac{\gamma}{\eta} \qquad \text{(A.3)}$$

$$\text{rel-bias } (\bar{y}_1) = \frac{c_1 - c_2}{f_1 + f_2} = \frac{\gamma}{\xi} \qquad \text{(A.4)}$$

where

$$e_0 = \sum_{j\epsilon\omega_1} \sum_{k=1}^{m_{1j}} w_{1jk}$$

$$c_1 = \sum_{j\epsilon\omega_1} p_{2j} \sum_{k=1}^{m_{1j}} w_{1jk} Y_{1jk}$$

$$f_1 = \sum_{j\epsilon\omega_1} p_{1j} \sum_{k=1}^{m_{1j}} w_{1jk} Y_{1jk}$$

$$c_2 \equiv f_2 = \sum_{j\epsilon\omega_2} \sum_{k=1}^{m_{2j}} w_{2jk} Y_{2jk}$$

Table 7. Percent Distribution of Effective Confidence Levels for Assumed 95 Percent Confidence Interval Estimates of $\bar{y}_0$ Using $\bar{y}_1$ for NAEP No-Show Study Packages and Involving All No-Shows in Bias Estimation

Effective Confidence Level	Exercise Subject Matter Type		Total
	Mathematics	Science	
95-80	36.9	38.2	37.4
80-65	17.1	20.7	18.7
65-50	10.5	9.5	10.1
50-35	15.8	19.0	17.3
35-20	3.9	6.3	5.0
20-0	15.8	6.3	11.5
Total	100.0	100.0	100.0
Number of Exercises	76	63*	

*Effective confidence levels could not be computed for one package 09 and all package 13 science exercises since se($\bar{y}_1$) data were unavailable.

The parameters P_{1j} and P_{2j} were estimated from school response rates during regular assessment. The estimate for the No-Show study group package in a school was the response rate to all group packages given in that school. Similarly, the No-Show study individual package response rate in a school was obtained from the response rate to all individual packages given in that school. The w_{1jk} are regular assessment weights adjusted for regular assessment nonresponse by a weighting class procedure. The w_{2jk} weights denote the reciprocals of No-Show selection probabilities adjusted for No-Show nonresponse.

Equations (A.3) and A.4) yield bias estimates involving in-school regular assessment respondents and all No-Show respondents. Another set of meaningful bias estimates involves in-school regular assessment respondents and in-school No-Show respondents. The definition changes indicated by the (*) were motivated by the attempt to form a matched school bias estimator based exclusively on in-school No-Shows. The set of schools ω_1^* is the subset of regular assessment ω_1 schools which provided in-school No-Show responses for the particular package in question. The deleted schools either had no cooperating in-school No-Show respondents for the package, or were subsampled out at the No-Show package assignment stage to control the package yield per PSU. The regular assessment respondent weights for the set of ω_1^* schools with in-school No-Show responses for the package were inflated to account for the deleted schools, hence the adjusted w_{1jk}^* weights.[A]

Regarding the components of equations (A.3) and (A.4),

$$\text{rel-bias}^*\ (\bar{y}_1) = \frac{c_1^* - c_2^*}{f_1^* + f_2^*} = \frac{\gamma^*}{\xi^*} \quad (A.5)$$

and

$$\text{bias}^*\ (\bar{y}_1) = \frac{c_1^* - c_2^*}{e_o^*} = \frac{\gamma^*}{\eta^*} \quad (A.6)$$

where

$$e_o^* = \sum_{j\epsilon\omega_1^*} \sum_{k=1}^{m_{1j}} w_{1jk}^*$$

$$c_1^* = \sum_{j\epsilon\omega_1^*} p_{2j} \sum_{k=1}^{m_{1j}} w_{1jk}^* Y_{1jk}$$

$$f_1^* = \sum_{j\epsilon\omega_1^*} p_{1j} \sum_{k=1}^{m_{1j}} w_{1jk}^* Y_{1jk}$$

$$c_2^* = f_2^* = \sum_{j\epsilon\omega_1^*} \frac{p_{2j}\ m_{1j}}{m_{2j}^*} \sum_{k=1}^{m_{2j}^*}$$

$$w_{1jk}^*\ Y_{2jk}^*$$

Table 8. Comparison of Followup and No-Followup Alternatives

$$C_0 = \$2 \qquad \bar{Y}_1 = 0.50$$
$$C_1 = 10 \qquad P_1 = 0.75$$
$$C_2 = 50$$

Absolute Value of Bias ($\bar{y}_0$) (In Percent)	k_{opt}	With Followup			LOSS
		m_0	m_1	m_2	
High-Budget Study:					
$C = C^* = \$200,000$					
$m_1^* = 15,790$					
0.0	1.99	12665	9499	1594	0.6
0.5	1.99	12669	9502	1593	1.7
1.0	1.99	12681	9510	1591	4.7
1.5	2.00	12701	9526	1586	9.8
2.0	2.01	12730	9548	1581	17.1
2.5	2.03	12767	9575	1574	26.5
3.0	2.04	12814	9611	1565	38.1
3.5	2.07	12870	9653	1555	52.1
4.0	2.10	12936	9702	1542	68.5
4.5	2.13	13013	9759	1528	87.5
5.0	2.17	13101	9825	1511	109.4
Low-Budget Study:					
$C = C^* = \$20,000$					
$m_1^* = 1,579$					
0.0	1.99	1267	950	159	0.6
0.5	1.99	1267	950	159	0.7
1.0	1.99	1268	951	159	1.1
1.5	2.00	1270	952	159	1.6
2.0	2.01	1273	955	158	2.3
2.5	2.03	1277	958	157	3.2
3.0	2.05	1281	961	157	4.4
3.5	2.07	1287	965	156	5.8
4.0	2.10	1293	970	154	7.5
4.5	2.13	1301	976	153	9.4
5.0	2.17	1310	983	151	11.6

With m_{2j}^* denoting the number of in-school No-Show responses from school $j\epsilon\omega_1^*$, the definition of the set of schools ω_1^* assures that $m_{2j}^* > 0$. Since the adjusted weights $w_{1jk}^* = w_{1j}^* E_{oj}/m_{1j}$ with w_{1j}^* denoting the non-response-adjusted school by package weight, one can recast C_1^* and C_2^* as follows:

$$c_1^* = \sum_{j\epsilon\omega_1^*} w_{1j}^* \; p_{2j} \; E_{oj} \; \bar{y}_{1j}^* \;\; ,$$

$$c_2^* = \sum_{j\epsilon\omega_1^*} w_{1j}^* \; p_{2j} \; E_{oj} \; \bar{y}_{2j}^* \;\; .$$

The numerator of equations (A.5) and (A.6) is therefore

$$(c_1^* - c_2^*) = \sum_{j\epsilon\omega_1^*} w_{1j}^* \; p_{2j} \; E_{oj} \qquad (A.7)$$

$$\left(\bar{y}_{1j}^* - \bar{y}_{2j}^* \right).$$

Overall estimators of bias and rel-bias can be viewed in terms of components defined at the PSU level. To do this, one must attach subscripts "hi" to $\gamma, \eta, \xi, \gamma^*, \eta^*$, and ξ^* of equations (A.3) through (A.6). Using

these terms associated with PSU-i within pseudo-stratum-h, one obtains overall bias and rel-bias estimators which involve all No-Shows as

$$\text{bias } (\bar{y}_1) = \gamma_{++} / \eta_{++} \qquad (A.8)$$

$$\text{rel-bias } (\bar{y}_1) = \gamma_{++} / \xi_{++} \qquad (A.9)$$

and which involve only in-school No-Shows as

$$\text{bias}^* (\bar{y}_1) = \gamma_{++}^* / \eta_{++}^* \qquad (A.10)$$

$$\text{rel-bias}^* (\bar{y}_1) = \gamma_{++}^* / \xi_{++}^* . \qquad (A.11)$$

30

The second-order estimators of variance for expressions (A.8) through (A.11) were based upon a form of the "jackknife" technique introduced by Quenouille (1949) and advanced for interval estimation by Tukey (1958). The exact form used here was presented by Frankel (1971). The procedure is described for estimates involving all No-Shows, although the procedure for estimates involving only in-school No-Shows is similar.

First, these definitions are given:

$$\beta_{h1} = \frac{2\gamma_{++}}{\eta_{++}} - \frac{\gamma_{++} + \gamma_{h1} - \gamma_{h2}}{\eta_{++} + \eta_{h1} - \eta_{h2}}$$

$$\beta_{h2} = \frac{2\gamma_{++}}{\eta_{++}} - \frac{\gamma_{++} + \gamma_{h2} - \gamma_{h1}}{\eta_{++} + \eta_{h2} - \eta_{h1}}$$

$$\zeta_{h1} = \frac{2\gamma_{++}}{\xi_{++}} - \frac{\gamma_{++} + \gamma_{h1} - \gamma_{h2}}{\xi_{++} + \xi_{h1} - \xi_{h2}}$$

$$\zeta_{h2} = \frac{2\gamma_{++}}{\xi_{++}} - \frac{\gamma_{++} + \gamma_{h2} - \gamma_{h1}}{\xi_{++} + \xi_{h2} - \xi_{h1}}$$

The associated jackknife estimators are

$$\text{var} \{\text{bias } (\bar{y}_1)\} = 1/4 \sum_{\substack{h=1 \\ h \neq h^o}}^{H} [\beta_{h1} - \beta_{h2}]^2 \qquad (A.12)$$

$$+ 1/18 \sum_{i=1}^{2} \sum_{j=i+1}^{3}$$

$$[\beta_{h^o_i} - \beta_{h^o_j}]^2$$

and

$$\text{var} \{\text{rel-bias } (\bar{y}_1)\} = 1/4 \sum_{\substack{h=1 \\ h \neq h^o}}^{H} \qquad (A.13)$$

$$[\zeta_{h1} - \zeta_{h2}]^2 + 1/18$$

$$\sum_{i=1}^{2} \sum_{j=i+1}^{3} [\zeta_{h^o_i} - \zeta_{h^o_j}]^2$$

where h^o is the pseudo-stratum assigned three PSU's.

A SURVEY DESIGN TO REDUCE THE EFFECT OF MEASUREMENT ERRORS IN COLLECTED DATA

In this portion of the paper a survey design which employs double sampling to reduce the effect of measurement errors on estimates from survey data is described.

Basic Measurement Error Concepts

The concepts employed are based upon the Census Bureau model (CBM) for measurement errors (Hansen, et. al. 1961; Koch, 1973) which may be briefly described as follows: There exists a population of N individuals. For a particular measurement process, the measurement obtained for the i-th individual at the t-th trial of the survey is Y_{it}. The subscript t indexes a series of repeatable trials of the measurement process (i.e., of the census or survey in question).

One can define the expected value over trials of the measurement for the i-th individual:

$$E_t \{Y_{it} \mid U_i = 1\} = Y_i , \qquad (3.1.1)$$

where U_i is an indicator random variable denoting presence in the sample.

If we denote the "true" or actual values for the i-th individual as X_i, then the expected measurement for the i-th individual may not be equal to the actual or "true" values for that individual. That is $X_i \neq Y_i$ and there may be a net bias in the measurement process. For example if one wishes to estimate the population mean

$$X = \frac{1}{N} \sum_{i=1}^{N} X_i ,$$

there is a net bias in the measurement process if $\bar{X} \neq \bar{Y}$, where

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i .$$

In the CBM the mean square error of an estimate y, say, is as follows,

$$MSE(y) = MV + SV + B + IV \ , \qquad (3.1.2)$$

where

$MV =$ measurement variance (or response variance) which arises due to the fact that the measurements obtained for a particular individual are not always the same for repeated trials of the measurement process.

$SV =$ sampling variance and arises because a sample is drawn. This is similar to the usual sampling errors or standard errors of estimates.

$B =$ bias term and arises if there are any systematic errors in the measurement process.

$IV =$ interaction variance which is due to the interaction of the sampling errors and the measurement errors and arises if the expected measurement for a particular individual in the sample depends upon the other individuals in the sample.

For the remainder of the discussion, we will assume (as is commonly done) that the interaction variance is zero. We will term a measurement process which measures Y_{it} as a faulty measurement process. For a simple random sample of size n, an estimate of the population mean using the faulty measurement process is

$$\bar{y}_t = \frac{1}{n} \sum_{i=1}^{n} Y_{it} \ . \qquad (3.1.3)$$

The mean square error is

$$MSE(\bar{y}_t) = \frac{1}{n} \{ (SMV) + (n-1)(CMV) \} \qquad (3.1.4)$$

$$+ \frac{1}{n} \left(\frac{N-n}{N} \right) \{ BV + BT + TV \}$$

$$+ B \ .$$

The first term is the measurement variance (or response variance), MV, and consists of two components, SMV, the simple measurement variance which is due to the trial to trial variability in measurement for a single individual and CMV, the correlated measurement variance which arises due to the between individual correlation of measurement errors. This latter component is often thought to be introduced by the presence of interviewers, abstractors, coders, etc.

The second term is the sampling variance, SV, and is due to the variability of the Y_i around $\bar{Y}$. The sampling

variance is composed of three components: BV, the sampling variance of the individual bias terms, $B_i = Y_i - X_i$ around the net bias $\bar{B}$; TV, the sampling variance of the true values; and BT, the interaction between the individual bias terms and the true values.

The third term B is due to the square of the bias in the measurement process. For a fuller definition of these components see Hansen, et. al., 1961; Koch, 1973; Lessler, 1974; 1976.

Two of the above components, the simple measurement variance and the sampling variance, are affected by the size of the sample. If these two components are large, the overall mean square error can be reduced by increasing the sample size. The two other components of the mean square error, CMV, the correlated measurement variance, and B, the square of the net bias, are not affected by the sample size; if these two components are large, large reductions in the mean square error cannot be achieved by increasing the sample size.[2] Thus, if CMV and B make a large contribution to the total MSE of a survey estimate some type of quality control procedure must be introduced to control their effect.

In a 1976 article, Bailar (1976) discusses the sources of the above error components and gives estimates of their effects on Census statistics. The following table (Table 9) is adapted from that article and shows what part each of the above components contributes to the total relative mean square error. Note that the combined contribution of the CMV and B terms ranges from 7.4 percent to 89.4 percent of the total mean square error indicating that, in many cases, it is very important to have available special procedures for reducing the effect of these two components.

Survey Design for Controlling Measurement Errors

The following describes a survey design which can be used to control the effect of B and CMV. Suppose one has available for use in the Census or survey two measurement processes: (1) a cheap-faulty measurement process and (2) an expensive-accurate measurement process. These two measurement processes may be employed in combination to reduce the effect of the bias and correlated measurement variance on the survey estimates which would be present if only the inexpensive but inaccurate measurement process was used. A double sampling scheme (Madow, 1965) is employed in which an initial sample is drawn and measurements are obtained using the cheap-faulty

---

[2] *The CMV is affected by the number of interviewers or, in a record survey, the number of abstractors.*

Table 9. Percentage Contribution of Various Error Components to the Total Relative
Mean Square Error of an Estimate.*

Characteristics	Number of Persons	Component Due to Sampling + Simple Measurement Variance	Component Due to Correlated Measurement Variance	Component Due to Square of Relative Bias
Language Spoken in Home	7,500			
English Only	5820	10.6	20.6	68.8
French	75	67.2	0.0	32.8
German	218	13.0	11.8	75.3
Polish	120	48.8	38.0	13.2
Yiddish	180	22.4	72.0	5.6
Italian	278	26.2	60.0	13.8
Spanish	142	76.0	4.6	19.4
Number of Children Ever Born, Females 14 and Over	2,870			
None	930	59.5	23.3	17.3
0 to 2	1972	51.1	27.8	21.1
1 to 3	1455	92.6	6.9	0.5
4	230	61.8	37.1	1.1
5 or more	255	70.5	4.9	24.6

* Adapted from Bailar (1976, p. 282) "Table 3. — Estimates of components of Mean-Square Errors for Two
Characteristics for an Area of 7,500 Persons: 1970 Census." The published table contains the actual estimates
of the relvariances, square of relative bias, and relative mean square error. The above percentages were
calculated from the estimates given.

measurement process. From this initial sample, a sub-sample is drawn and the accurate measurements are obtained using the expensive measurement process.

In the case of estimating the population mean, the estimator for a particular trial of the survey process is

$$\bar{w}_t = \bar{y}_t - (\bar{y}_{st} - \bar{x}_s) \quad , \qquad (3.2.1)$$

where $\bar{y}_t$ is the biased estimate of the mean from the original sample, $\bar{x}_s$ is the unbiased estimate for the subsample, and $\bar{y}_{st}$ is the mean of the elements in the original sample that are also in the subsample.

For the double sampling scheme (DSS), the variance is

$$V(\bar{w}_t) = \frac{n-n_1}{nn_1} \{ SMV - CMV \} \qquad (3.2.2)$$

$$+ \frac{n-n_1}{nn_1} \quad \{ BV \}$$

$$+ \frac{1}{n} \left(\frac{N-n}{N} \right) \{ TV \}$$

where n is the size of the original sample and n_1 the size of the subsample. Note that, in addition to the fact that the DSS estimator is unbiased, the correlated measurement variance makes a negative contribution to its overall variance.

Depending upon the relative sizes of the measurement error components (i.e., MV and B), and the cost of the cheap data relative to the expensive data, the best survey design may be, (1) a survey which uses only the cheap-faulty data with relatively large sample sizes, (2) a survey which uses only the expensive accurate data and small sample sizes, or (3) a survey which employs the DSS. In addition, the above general DSS scheme may be adapted to a variety of survey designs from simple random sampling to complex multistage designs (see Lessler, 1974; 1976).

**Example From the National Medical Care
Expenditure Survey**

In the remainder of the paper, the manner in which the above design may be employed for controlling the mean square error is illustrated for a specific survey. RTI is in the process of conducting a National Medical

Care Expenditure Survey (NMCES). The specifications of the survey call for conducting household interviews in which the utilization of health care facilities and the associated medical care expenditures of each member of the household are collected along with other data.

Recognizing that a household survey approach is essential to the survey objectives, a natural question to ask is: "What is the optimum way in which to spend the resources available to collect the medical care expenditure data? Should the non-sampling errors in the interview data be ignored and the entire data collection budget be allocated to the household sample or should some of the budget be spent to verify the household data by checking the records of the various medical care facilities that provided care to the household?" A pilot study investigating the biases associated with various methods for collecting household data was conducted by the Johns Hopkins Health Services Research and Development Center (Shapiro, et. al., 1976; Yaffe, et. al., 1977). Data concerning utilization, charges, and payments for health care was collected from three sources, households, providers and third party payors (TPP). The data from each of these sources were compared and a "Best Data" set was constructed to be used as a criterion for measuring the accuracy of the household data. These data may be used to illustrate the point at hand. A tape of the Johns Hopkins data sets was obtained by RTI. The household data and the best data were used to compare optimum DSS survey in which the best data would be collected only for a subsample of individuals to a survey design in which the best data would be collected for each individual.

A simplified survey design was assumed which ignored the stratified multistage cluster design of the actual survey. We assume the following model for the household or faulty measurements:

$$Y_{it} = X_i + B_{it} \quad , \qquad (3.3.1)$$

where

$\quad Y_{it} \quad = \quad$ household interview value;

$\quad X_i \quad = \quad$ accurate value (best);

$\quad B_{it} \quad = \quad$ bias.

The estimate of the mean using the DSS is $\bar{w}_t$ as defined above. Using equation (3.2.2) the variance of $\bar{w}_t$ may be written as

$$V(\bar{w}_t) = \frac{SMV - CMV + BV}{n_1} \qquad (3.3.2)$$

$$+ \frac{TV - SMV + CMV - BV}{n}$$

$$- \frac{TV}{N} \quad ,$$

where n and n_1 are the sample and subsample sizes.

Optimum sample and subsample sizes are chosen to minimize variance for fixed cost using the following cost function:

$$C = nC_1 + n_1 C_2 \quad , \qquad (3.3.3)$$

where

$\quad C \quad = \quad$ total fixed cost for survey;

$\quad C_1 \quad = \quad$ cost of obtaining expenditure (or utilization) information by household (HH) interview;

$\quad C_2 \quad = \quad$ additional cost of obtaining the best expenditures (or utilization) measure. This would include the cost of followup to the providers and TPP, cost of matching records and visits, and cost of determining the best values X_i.

For the survey procedure in which each individual is followed up to determine the best, the estimate of the mean would be

$$\bar{x} = \frac{1}{n_2} \sum_{i=1}^{n_2} X_i \quad , \qquad (3.3.4)$$

$$n_2 = \frac{C}{C_1 + C_2} \quad .$$

The preferred of these two designs is considered to be that which has the smaller variance for fixed cost. The Johns Hopkins data contained information from 2300 individuals some of which were associated with an area probability sample and some of which were drawn from a sample of provider and TPP records. The 1164 individuals in the area probability sample were used in this analysis. Table 10 shows the relationship between the best data and the household data for nineteen selected utilization and charge variables.

For the DSS individual optimum allocation for sample and subsample sizes were calculated for each of these variables. Overall joint optimum allocation were calculated using a procedure proposed by Kish (1974).

In the Kish procedure a loss function for a particular variable is defined as the ratio of the variance obtainable under the actual sample allocation to that obtainable under the optimum sample allocation. The overall loss function for the survey is the weighted sum of the individual loss functions where the weights have been chosen to reflect the relative importance of the different variables included in the survey. The joint optimum allocation is the sample allocation that minimizes the overall loss function.

Estimates for the components of variance were calculated using the Johns Hopkins data by forming the individual bias terms and calculating their sample

Table 10. *Average Household Values and Average Best Values for*
Selected Variables. Area Probability Sample Only.[K]

Variable	Household Value	Best Value	Percentage HH is of Best
Physician Outpatient Nonclinic			
Utilization*	1.5	1.6	94
Charges ψ	20.79	24.87	84
Physician Outpatient Clinic:			
Utilization *	0.30	0.42	71
Charges ψ	4.39	7.71	57
Emergency Room:			
Utilization *	0.13	0.13	95
Charges ψ	6.68	8.28	81
All Physician Outpatient:			
Utilization *	1.9	2.2	86
Charges ψ	31.86	40.85	78
Other Medical Providers:			
Utilization *	0.29	0.36	81
Charges ψ	5.62	6.51	86
Dentists:			
Utilization *	0.74	0.81	91
Charges ψ	20.96	21.90	96
Drug Prescriptions:			
Utilization δ	1.9	2.3	83
Charges ψ	10.87	12.80	85
Physician Inpatient:			
Utilization ϵ	0.15	0.16	94
Charges ψ	19.92	23.45	85
Hospital Inpatient:			
Utilization δ	0.54	0.58	93
Charges ψ	68.53	76.82	89
Total:			
Sum of all Charges ψ	157.54	182.10	87

[ε]Number of different physicians seen

*Number of visits per person

ψ Dollars per person

[δ]Number of prescriptions per person

[γ]Days of care per person

[K]Data for a subset of the Johns Hopkins data. The actual dollar values and percentages may be somewhat different than those for the full John Hopkins data set.

variance, s_B^2, as well as s_x^2 the sample variance of the best data. And, it can be shown that

$$E(s_B^2) = SMV - CMV + BV \quad ;$$

$$E(s_x^2) = TV \quad .$$

In the following, the results for several sets of cost components are illustrated, as well as, several sets of weights depicting the relative importance of the different variables. These components are chosen for illustrative purposes and do not necessarily reflect the actual costs of the survey. In fact, several other considerations govern the decision as to the nature of the followup sample to be used in the NMCES and these are not reflected in this analysis.

Results. Charges and utilization are considered separately. Tables 11 through 13 show the results for charges. The following points may be noted:

For the majority of the variables the DSS has considerable gain over the complete followup, particularly for those variables for which the net bias is relatively low. An exception is dentist charges for which the household figure is 96 percent of the best. In this case, however, the variance of the bias terms was very high.

The effect of changes in the relative cost is illustrated by looking at the results for a particular variable across the three tables. Preference for the DSS over complete followup declines as the cost ration C_1/C_2 increases. When the cost ratio is 6/10 (Table 13) the optimum DSS is complete followup for clinic charges which has a very high bias; the household value is only 57 percent of the best value. Additional cost ratios are not shown but were examined at RTI. The same pattern was manifested as the C_1/C_2 cost ratio was increased, i.e., the optimum subsample sizes increase relative to the optimum sample sizes. When $C_1 = C_2 = \$10$, clinic charges is still the only variable for which the optimum allocation is the equal allocation.

The minimum weighted overall losses for the DSS are very low. The lowest possible value for this quantity is one. This indicates that the joint optimums are fairly good for all variables considered and that there were not highly conflicting requirements implied by the individual optimum designs. Since the biases associated with the clinic charges are very large, there is evidence for treating this variable differently from others.

The Kish procedure is easily adapted to a multistage design in which separate optimum allocations for various strata and levels of the sampling design can be calculated. It is likely that high utilization of clinics would be concentrated in certain geographic areas which could be subsampled at higher rates than in the overall sample.

Examination of Tables 14 through 16 showing the results of the optimization procedure for utilization reveals that similar results are obtained for the utilization variables. Gains of the DSS relative to the complete followup are even more dramatic for utilization variables. This is due to the fact that the biases and variance of the bias terms are smaller for the utilization variables than for the charge variables.

The above clearly indicates that gains can be achieved by a double sampling procedure employing faulty and accurate measurement processes. The model shown here is very simple and a more complete model would need to be used. A more complete model would consider subsampling at all design levels, the nonresponse in both the initial and followup data, possible nonlinear cost functions, etc.

Relevance to Total Survey Design in General

The above illustrates several concepts that are important to Total Survey Design.

1. If we accept the Census Bureau model for measurement errors in collected data, it is clear that large samples may not be an effective way for controlling the MSE of survey estimates.
2. Special survey designs may, however, be employed to control the effect of the error components that are not reduced by increasing sample sizes. The DSS design described above is one such design.
3. Not only is the net bias associated with particular survey procedures important, but the variance of the individual bias terms is also important.
4. It is important that cost accounting be done in a manner that will allow one to associate costs with the various levels of the sampling design. This will permit one to rationally employ the overall optimization procedure.
5. In the example dealt with here the joint optimum allocations were close to the individual optimums in most cases. This does not always occur however. In an evaluation of the NCHS Hospital Discharge Survey conducted by RTI, the optimum designs for certain large overall domains were very different than those for small domains dealing with specific conditions. When this occurs the minimum overall loss is not close to one. In this case, instead of attempting to develop a single sampling plan for all characteristics measured in the survey, different designs for different subsets of the characteristics measured may be cost effective.

Table 11. Optimum DSS for Charges Variables, Loss for a Sample Design Which Gathers Best Data for Everyone, Weighted Overall Loss, Joint Optimum Allocation, and Minimum Weighted Overall Loss for Various Weight Sets.

Total Cost = $100,000; c_1 = $1 ; c_2 = $10

Variable	Optimum Sample Size	Optimum Subsample Size	Loss of Complete Followup Relative to Optimum (Relative Efficiency)
Phy. Outpatient—Nonclinic	33945	6605	1.75
Phy. Outpatient—Clinic	19480	8052	1.13
Phy. Emergency Room	29583	7042	1.51
All Outpatient Physician	28232	7177	1.44
Other Medical Providers	49807	5019	3.01
Dentists	31453	6855	1.61
Drugs	56616	4339	3.73
Physician Inpatient	41656	5834	2.28
Hospital Inpatient	44180	5582	2.49
Total All Charges	46518	5348	2.69

Joint Optimum Allocations

Weight Set	Weighted Overall Loss of Complete Followup	Joint Sample Size	Joint Subsample Size	Minimum Weighted Overall Loss
W.S.1: Equal weights all variables	2.16	38709	6129	1.05
W.S.2: Omits the two variables that are sums of other variables	2.19	38929	6107	1.05
W.S.3: Omits the three variables that sum to outpatient physician	2.46	42885	5711	1.03
W.S.4: Omits the components of outpatient physician charges, total charges, and drugs	2.17	39354	6065	1.03

Table 12. Optimum DSS for Charges Variables, Loss for a Sample Design Which Gathers Best Data for Everyone, Weighted Overall Loss, Joint Optimum Allocation, and Minimum Weighted Overall Loss for Various Weight Sets.

Total Cost = $100,000; $c_1 = \$3$; $c_2 = \$10$

Variable	Optimum Sample Size	Optimum Subsample Size	Loss of Complete Followup Relative to Optimum (Relative Efficiency)
Phy. Outpatient—Nonclinic	15697	5291	1.32
Phy. Outpatient—Clinic	9843	7047	1.02
Phy. Emergency Room	14039	5788	1.20
All Outpatient Physician	13508	5948	1.17
Other Medical Providers	21073	3678	1.91
Dentists	14761	5572	1.25
Drugs	23109	3067	2.20
Physician Inpatient	18430	4471	1.58
Hospital Inpatient	19274	4218	1.68
Total All Charges	20035	3990	1.77

Joint Optimum Allocations

Weight Set	Weighted Overall Loss of Complete Followup	Joint Sample Size	Joint Subsample Size	Minimum Weighted Overall Loss
W.S.1: Equal weights all variables	1.51	16961	4912	1.05
W.S.2: Omits the two variables that are sums of other variables	1.52	17009	4897	1.06
W.S.3: Omits the three variables that sum to outpatient physician	1.65	18531	4441	1.04
W.S.4: Omits the components of outpatient physician charges, total charges, and drugs	1.52	17388	4784	1.03

Table 13. *Optimum DSS for Charges Variables, Loss for a Sample Design Which Gathers Best Data for Everyone, Weighted Overall Loss, Joint Optimum Allocation, and Minimum Weighted Overall Loss for Various Weight Sets.*

Total Cost = $100,000; c_1 = $6 ; c_2 = $10

Variable	Optimum Sample Size	Optimum Subsample Size	Loss of Complete Followup Relative to Optimum (Relative Efficiency)
Phy. Outpatient—Nonclinic	9288	4427	1.14
Phy. Outpatient—Clinic	6250	6250	1.00
Phy. Emergency Room	8452	4928	1.07
All Outpatient Physician	8178	5093	1.06
Other Medical Providers	11809	2915	1.47
Dentists	8819	4708	1.10
Drugs	12695	2383	1.64
Physician Inpatient	10604	3638	1.29
Hospital Inpatient	10995	3403	1.35
Total All Charges	11343	3194	1.40

Joint Optimum Allocations

Weight Set	Weighted Overall Loss of Complete Followup	Joint Sample Size	Joint Subsample Size	Minimum Weighted Overall Loss
W.S.1: Equal weights all variables	1.25	9763	4142	1.05
W.S.2: Omits the two variables that are sums of other variables	1.26	9776	4134	1.06
W.S.3: Omits the three variables that sum to outpatient physician	1.33	10558	3665	1.03
W.S.4: Omits the components of outpatient physician charges, total charges, and drugs	1.25	10034	3980	1.03

Table 14. Optimum DSS for Utilization Variables, Loss for a Sample Design Which Gathers Best Data for Everyone, Weighted Overall Loss, Joint Optimum Allocation, and Minimum Weighted Overall Loss for Various Weight Sets.

Total Cost = $100,000; $\qquad$ c_1 = 1 ; $\qquad$ c_2 = 10

Variable	Optimum Sample Size	Optimum Subsample Size	Loss of Complete Followup Relative to Optimum (Relative Efficiency)
Phy. Outpatient—Nonclinic	48629	5137	2.89
Phy. Outpatient—Clinic	35664	6434	1.84
Phy. Emergency Room	58271	4173	3.93
All Outpatient Physician	46935	5306	2.73
Other Medical Providers	43535	5646	2.46
Dentists	47782	5222	2.81
Drugs	51427	4857	3.17
Physician Inpatient	43923	5610	2.47
Hospital Inpatient	59902	4010	4.12

Joint Optimum Allocations

Weight Set	Weighted Overall Loss of Complete Followup	Joint Sample Size	Joint Subsample Size	Minimum Weighted Overall Loss
W.S.1: Equal weights all variables	2.93	48480	5152	1.02
W.S.2: Omits the variable that is sum of other variables	2.96	48669	5133	1.02
W.S.3: Omits the three variables that sum to outpatient physician	2.96	48927	5107	1.01
W.S.4: Omits the components of outpatient physician charges, and drugs	2.91	48434	5157	1.01

*Table 15. Optimum DSS for Utilization Variables, Loss for a Sample Design Which Gathers Best Data for Everyone,
Weighted Overall Loss, Joint Optimum Allocation, and Minimum Weighted Overall Loss
for Various Weight Sets.*

Total Cost = \$100,000; c_1 = \$3 ; c_2 = \$10

Variable	Optimum Sample Size	Optimum Subsample Size	Loss of Complete Followup Relative to Optimum (Relative Efficiency)
Phy. Outpatient—Nonclinic	20705	3788	1.86
Phy. Outpatient—Clinic	16328	5102	1.38
Phy. Emergency Room	23582	2925	2.28
All Outpatient Physician	20168	3949	1.79
Other Medical Providers	19060	4282	1.66
Dentists	20438	3869	1.82
Drugs	21571	3529	1.98
Physician Inpatient	19182	4245	1.67
Hospital Inpatient	24042	2787	2.36

Joint Optimum Allocations

Weight Set	Weighted Overall Loss of Complete Followup	Joint Sample Size	Joint Subsample Size	Minimum Weighted Overall Loss
W.S.1: Equal weights all variables	1.86	20492	3852	1.02
W.S.2: Omits the variable that is sum of other variables	1.87	20532	3840	1.02
W.S.3: Omits the three variables that sum to outpatient physician	1.88	20699	3790	1.01
W.S.4: Omits the components of outpatient physician charges, and drugs	1.86	20530	3841	1.01

40

Table 16. Optimum DSS for Utilization Variables, Loss for a Sample Design Which Gathers Best Data for Everyone, Weighted Overall Loss, Joint Optimum Allocation, and Minimum Weighted Overall Loss for Various Weight Sets.

Total Cost = \$100,000; c_1 = \$6 ; c_2 = \$10

Variable	Optimum Sample Size	Optimum Subsample Size	Loss of Complete Followup Relative to Optimum (Relative Efficiency)
Phy. Outpatient—Nonclinic	11645	3013	1.45
Phy. Outpatient—Clinic	9598	4241	1.17
Phy. Emergency Room	12896	2262	1.68
All Outpatient Physician	11403	3158	1.41
Other Medical Providers	10897	3462	1.33
Dentists	11525	3085	1.43
Drugs	12029	2783	1.51
Physician Inpatient	10953	3428	1.34
Hospital Inpatient	13090	2146	1.72

Joint Optimum Allocations

Weight Set	Weighted Overall Loss of Complete Followup	Joint Sample Size	Joint Subsample Size	Minimum Weighted Overall Loss
W.S.1: Equal weights all variables	1.45	11506	3097	1.02
W.S.2: Omits the variable that is sum of other variables	1.45	11518	3089	1.02
W.S.3: Omits the three variables that sum to outpatient physician	1.46	11618	3029	1.01
W.S.4: Omits the components of outpatient physician charges, and drugs	1.45	11539	3077	1.01

DISCUSSION OF TOTAL SURVEY DESIGN

Daniel Horvitz, Research Triangle Institute, Chair
Judith Lessler, Research Triangle Institute, Recorder

A background paper summarizing the concept and state of the art of Total Survey Design (TSD) was not prepared for this session. Rather, in order to stimulate discussion, the position paper presented two examples which demonstrated both the need for and the application of TSD in practice. In the first hour of the session, highlights of the two examples discussed in detail in the paper were reviewed by the authors.

Kalsbeek showed the consequences of nonresponse bias on the estimates of p-values for 17-year olds in the National Assessment of Educational Progress (NAEP) (i.e. estimates of the proportions of 17-year olds correctly answering specific exercises in science and mathematics). The efficiency loss resulting from failure to follow-up at least some of the 17-year old nonrespondents was developed for two different survey budgets and various levels of nonresponse bias when the cost per "no-show" follow-up was five times the cost for the initial respondents.

Lessler focused upon measurement bias in household surveys concerned with utilization of and associated charges for medical care. A double sampling scheme for collecting personal interview data from a sample of households with record checks of medical providers for a subsample of the households was presented. The optimum sample and subsample sizes were illustrated for several cost models, and loss functions were computed which compared the optimum design to the limiting case in which all households in the sample are selected for record checks rather than just a subsample. Following this the chairman opened the floor to discussions of the issues at hand.

The review of the discussion period is organized in terms of the topics considered: nonresponse bias and adjustment procedures, measurement errors in the collected data, and TSD.

NONRESPONSE BIAS AND ADJUSTMENT PROCEDURES

The discussion of nonresponse focused around four basic points: (1) the effect of nonresponse adjustment procedures upon the bias due to nonresponse, (2) the differential nature of groups of nonrespondents, as well as the different levels of nonresponse bias for various subgroups of the population, (3) the effect of nonresponse bias when the aim of the survey is to estimate differences, either between years or between subgroups of the population, and (4) the procedures for using the information that is gathered in a "no-show" or nonresponse bias study to guide the development of new studies.

Considering the effect of various adjustment procedures for nonresponse upon the net nonresponse bias, it was noted that in the example dealt with in the position paper, a "weighting class" special adjustment procedure for nonresponse was used in the "no show" study. It was pointed out in the discussion that there are various sophisticated adjustment procedures available in which measures for nonrespondents are imputed that are believed to reduce the effect of the bias due to nonresponse. Thus, optimum values for subsampling fractions for nonresponse follow-up would differ with the adjustment procedure used. The chairman and Kalsbeek readily agreed to this assertion.

Several points of discussion were raised concerning the differential nature of nonrespondents. A question was raised as to how the nonrespondents differ from the respondents in the NAEP study on characteristics other than their scores on the various packages. Kalsbeek replied that information was available on this point, and that, in general, the nonrespondents had lower grade point averages, higher absences, and were less likely to be college bound. It was observed that the amount of bias that one would expect due to nonresponse is likely to be associated with the reason for nonresponse. For example, in the NAEP study, one would expect that there might be a difference between persons who were nonrespondents due to illness and those who were deliberate no shows. There appeared to be general agreement that this was so; however, Kalsbeek reported that there was data available on reasons for nonparticipation, but that it had not been related to the level of nonresponse bias (Kalsbeek, et. al., 1974).

It was emphasized during the discussion that nonresponse bias will not be the same for different subgroups of the population. In addition, the relative degree to which the bias effects the overall mean

square error of an estimate will depend upon the coefficient of variation of the subgroup for which the estimate is made. For example, for an estimate for the total population covered in a survey, in which case sample size is large, the bias may make a relatively large contribution to the overall mean square error. However, when one makes an estimate for a subdomain of the total population, the sample size will be small and the sampling variance will make a relatively greater contribution to the mean square error of the estimate. In addressing this point, Kalsbeek stated that the aim of the survey should be to obtain some overall joint optimum allocation which attempts to balance out the requirements for small specific domains and for large overall domains. This could be done using estimates of the bias for each domain, overall and specific, for which survey estimates will be made. A weighting procedure which reflected the relative importance of the various statistics to be produced by the survey in the context of the overall aims and goals of the survey should be developed and used to guide the development of the overall design. It was also noted that one is not limited to a single subsampling strategy for follow up of nonrespondents, and different subgroups of the population which are expected to have different nonresponse rates and different levels of nonresponse bias can be subsampled separately with appropriately chosen subsample sizes in the nonresponse follow up study.

The notion of nonresponse adjustment procedures and the notion of variability among subgroups in nonresponse bias were brought together when it was remarked that nonresponse imputation techniques should take into account what is known about the nonrespondents. Various known characteristics of a particular individual for whom a response was not obtained can be used to adjust for this missing value under a model for nonresponse bias. An old example from the 1940's, the Politz-Simmons technique (Politz and Simmons, 1949; 1950), was cited as a case in point. Under this technique one associates the value of an individual response with a probability that the individual is at home at the time the survey is made. This information is then used to predict the overall value for both the respondents and the not-at-homes (nonrespondents). Much additional research is needed on the development of models for adjusting for nonresponse and these models should consider response rates as well as the magnitude of nonresponse bias.

Are nonresponse biases serious in the case where the purpose of the survey is analytic and the aims are to report on differences between various subgroups of the population or, possibly, differences between two points in time? It was pointed out that, as long as the bias is the same for both groups used in the comparison, that the estimate of the difference will be a valid estimate. However, others noted that the bias is not always the same for various subgroups of the population or at different time points and that assuming so could distort one's picture of the actual differences. It was noted that in a longitudinal survey our interest is not only in the year to year changes but, also, in an assessment of the conditions in a particular year. Greenberg remarked that in survey research one should always trust Murphy's law which says that if anything can go wrong it will. Thus, if one did not use procedures which adjust for nonresponse bias but planned instead to trust that the nonresponse bias would be the same for the two comparison groups, under Murphy's law this would, most likely, not happen.

The manner in which information from a study such as the NAEP no-show study could be used to plan future surveys was discussed. The following scenario was outlined: At the beginning of the design of a new study or, in the case of the NAEP example, the next round of the study, one would decide that his survey would be conducted essentially in two phases. In the first phase, the usual survey procedures would be undertaken, and in the second phase, a sample of those who failed to respond to the usual survey procedures would be followed up, i.e., a new procedure would be instigated to obtain the responses for those who had not responded to the initial procedure. Knowing the level of resources available for the survey, the cost of using the initial procedure, and the cost of using the follow up procedure, and knowing information on the extent of nonresponse and the magnitude of the nonresponse bias, optimum sizes for the initial sample and the nonrespondent subsample would be obtained in advance of the survey. The statistics produced at the end of the survey would then be adjusted for nonresponse bias. Individual optimum sample-subsample allocations for each domain would be considered, as well as, an overall optimum allocation which attempted to balance out the requirements for the various domains relative to their importance to the overall aims of the survey.

PROCEDURES FOR CONTROLLING MEASUREMENT ERRORS IN COLLECTED DATA

Discussion related to the proposed procedures for treating measurement errors in the collected data centered around several points: (1) the adequacy of overall measurement error models, (2) special issues concerned with the correlated measurement variance component, (3) procedures for estimating the sizes of the various error components, and (4) how the sizes of the various error components affect the decision to use a particular survey scheme.

During the discussion several points concerning the adequacy of the Census Bureau Model for measurement errors were raised. The adequacy of the model in reference to the definition of the characteristic being measured was questioned. It was pointed out that often one has a definition for a characteristic which cannot be applied in actual practice because one feels that large biases would result in the attempt to apply this definition and some other related characteristic is

measured. Bias due to this is not reflected in the present model. In response to this point, Lessler agreed that the model, as it is presently formulated, does not adequately deal with these situations and suggested that the model needs to be expanded to deal with at least four identifiable cases. The first of these is the case in which there is an operational definition for the characteristic being measured and that operational definition is being carried out during the survey; second, the case in which a clear operational definition exists for the characteristic being measured but this operational definition cannot be carried out during the survey; third, the case in which clear operational definitions for the characteristic are not yet available; and finally, situations in which a true or actual value for the characteristic does not exist.

On the same line, doubt was expressed as to whether one could ever have an accurate measurement process in the sense of a process that would measure the true values for the characteristic in question. Assuming not, what one really uses in the double sampling scheme (DSS) is the difference between a faulty measurement process and a less faulty measurement process. In response to this, it should be noted that the more accurate measurement process can usefully be conceived of as the ideal or the best measurement process possible under the conditions existing for the survey. Thus, the more accurate measurement process is the standard at the point in time that the survey was conducted; however, as one increases one's expertise and ability to measure the characteristic, new standards will be developed for its measurement. Such an evolutionary process in which one continually refines measurement processes making them more accurate and precise is characteristic of all scientific endeavors.

Two questions were raised as to the sources of bias and variance in the overall mean square error. First, in a longitudinal survey the process of continually measuring an individual can affect his response such that additional biases are introduced by the measurement process at the later interviews. Do the current survey error models adequately cover this situation? It was agreed that the model does need to consider such situations; however, the net measurement bias can come from a variety of sources. For example, in a longitudinal or panel survey, the increasing bias in a particular measurement process as it is repeated from measurement point to measurement point could be reflected in the net measurement bias for each year. Second, an example of alternate forms of a questionnaire was given in which each questionnaire has an individual bias and these biases may not be the same for the different questionnaires, i.e., there is variability in the biases between questionnaires. Does one call this variance or bias in the overall mean square error model? The breakdown of the sampling variance shown in equation (3.1.4) of the position paper helps to clarify this point. In addition to an overall net bias in the measurement process, the sampling variance contains a component BV which is due to the sampling variance of the individual bias terms around the net bias.

The use of alternative estimators for the DSS proposed in the position paper was brought up and Dr. Lessler was asked whether or not a regression estimator had been considered instead of the simple difference estimator. Dr. Lessler replied that a regression estimator had not been considered initially, but that Wolter (1975) had considered such an estimator but only for the case when the correlated measurement variance component is 0. Further research is needed to determine the optimum estimator for a double sampling scheme. Ratio estimators, difference estimators, and regression estimators should all be considered and the situations as to when one is to be preferred over the other should be delineated.

45

The correlated component of the measurement variance received particular attention during the discussion. It was pointed out that the simple model formulated in the position paper assumed a single interviewer. In the Census Bureau Model for measurement errors, it is usually assumed that the correlated component of measurement variance is largely due to the interviewers, coders, supervisors and other personnel who handle the data. As sample sizes increase, the number of individuals handling the data will also increase, i.e., the number of interviewers, coders, etc. will increase with increasing sample size. If one assumes that correlated measurement variance is largely confined to interviewer effects, the correlated component of the measurement variance will be divided by the number of interviewers. In this sense one would expect the correlated measurement variance to be affected by the sample size because the number of interviewers employed would increase as the sample size increases. There was agreement that this would occur; however, it was pointed out that interviewers and other data handlers are not the only source of correlated measurement variance in surveys. Other sources are the form and wording of the questions, the order of questions or of categorized responses, and, in categorical data, the number of categories and the sizes of the intervals chosen. These sources can all be expected to contribute to the correlated measurement variance, so that even in a self-enumeration survey some correlated measurement variance remains (Andrews and Crandall, 1976).

The question was raised as to whether or not one could conclude that the most important component of the mean square error and the one for which one needed to obtain special estimates or to apply special measures for controlling its contribution to the mean square error is the bias component, particularly since the usual methods of estimating the variance of survey statistics accounted for some part of the measurement variance component. In response, it was noted that the usual estimates do contain a part of the measurement variance; however, the part that may be omitted from the usual variance estimation procedures is the correlated measurement variance which, in some cases, may

be the largest component of the mean square error. Horvitz (1952) gave an example from a morbidity survey conducted some 20 years ago in which it was found that the correlated measurement variance accounted for 80 percent of the overall variability of the estimated mean number of illnesses per household.

In the discussion dealing with the bias due to nonresponse, considerable thought was given to how a net bias would effect a difference estimator which compared measurements taken at two points in time. A similar query arose concerning how the measurement variance would be affected in such a situation. In response, it was noted that the measurement variance for a difference estimator would consist of a measurement variance component for the first time point plus a measurement variance component for the second time point, and, subtracted from that, a measurement variance component due to the correlation between the measurements obtained on the two occasions.

Several issues concerning use of the DSS as it is formulated in the position paper were raised. In response to a question as to how faulty the faulty measurement could be before it ceases to be useful, Lessler noted that she had examined this question for a variety of hypothetical data. The results depended upon the relative magnitude of the various measurement error components and the relative cost of the faulty-cheap measurement process and the expensive-accurate measurement process. In this investigation, use of three survey schemes was compared: a survey scheme which employed only the accurate measurement process, a survey scheme which employed only the faulty measurement process, and a double sampling survey scheme which employed the two measurement processes in combination. Table A is illustrative of the type of results that are obtained. By examining the table, we note that for the level of bias considered one uses the faulty measurement process alone when the errors are relatively small and the cost of the accurate measurement is relatively high. One uses the accurate measurement process alone when the measurement errors are very high or when the cost of the accurate relative to the faulty is relatively low. In the intermediate stage, when errors are moderate and relative costs are moderate, one uses the double sampling scheme. It was pointed out in the discussion that the results depicted above will vary with the relative frequency in the population of the characteristic being measured. This is indeed true. In fact, a key consideration as to how important certain percent biases will be to the mean square error of an estimate from a particular measurement process is the coefficient of variation in the population of the characteristic being measured. However, it should be noted, that for rare characteristics the opportunity for realizing large relative biases is present. For example, it is not unlikely that one could realize a relative bias of over 200 percent for a rare characteristic.

It was pointed out that procedures for adjusting for measurement errors in survey data are not often employed, and it was suggested that the reason for this was due to the difficulty in obtaining sources for validity data. It was generally agreed that it is difficult to obtain these sources and that survey researchers should be making a greater effort to obtain information on the measurement error components. Jabine reported that the Current Population Survey has some good estimates of components of error in reports of income data and, in addition, has information on the causes or correlates of these errors, i.e., the reasons which cause estimates to be in error. This information is used to adjust CPS data for use in the analysis of income maintenance and tax programs (Alvey and Cobleigh, 1975; Sailer and Vogel, 1975; Kilss and Millea, 1975; Herriot and Spiers, 1975; Alvey and Kilss, 1976; Vaughan and Yuskavage, 1976).
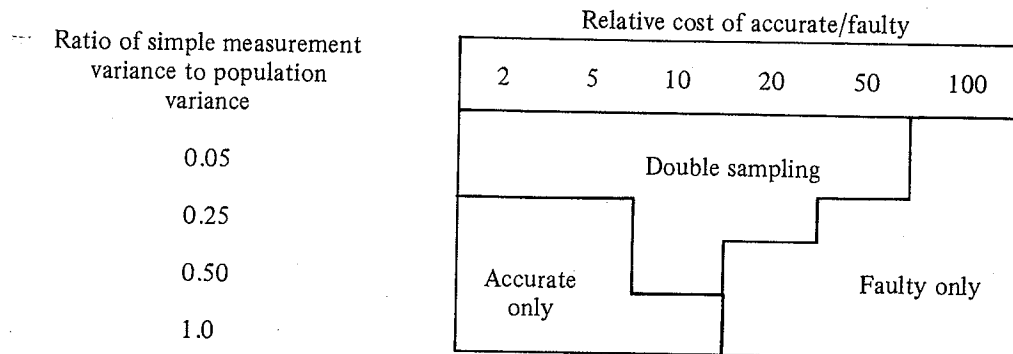
TOTAL SURVEY DESIGN

It was suggested that what had been discussed previously and what had been discussed in the position paper was Almost Total Survey Design in the sense that the two types of errors, those due to nonresponse and those in the collected data, had not been discussed simultaneously. The question was raised as to why this was not done. Lessler noted that there is not at present a unified model for considering the two simultaneously. Steps are just now beginning to be taken to formulate such a model. A recent paper by Platek, et. al., (1977) makes a step in this direction.

During a discussion as to how one would employ the concept of TSD at the planning stages of the survey, the chairman suggested that the method of applying the model should be similar to that used when deciding upon sample sizes for a survey. One obtains advance estimates or guesses as to the size of the various components of the mean square error, as well as, estimates of cost associated with various levels of a sampling design and those associated with procedures for adjusting for nonresponse and procedures for adjusting for measurement errors in collected data. One then applies overall cost and error models to obtain an optimum design under the values assumed. This would be done for each characteristic measured in the survey and for each domain or subpopulation for which estimates are to be made. A joint optimum allocation would then be constructed which reflected the relative importance of the various statistics to be produced from the survey. It was pointed out that, if conflicting requirements were implied by the various optimum allocations for the different characteristics and different subpopulation or domains, one may wish to employ different strategies for different domains. For example, poststratification could be used in which different nonresponse follow up strategies and different adjustment strategies for measurement errors would be employed among the various levels of poststratification.

Table A

Preferred survey scheme for measuring an attribute with the true proportion in the population = 0.2. Net bias in the faulty measurement process = 20%. Population variance, TV = 0.16; Sampling variance + simple measurement variance, SV + SMV = 0.1824, Bias variance, BV = 0.008; correlated measurement variance, CMV = 0. Preferred survey scheme is that which has the smallest mean square error for fixed cost. Total Cost = $1,000 C_1, where C_1 equals cost of faulty process.

Ratio of simple measurement variance to population variance	Relative cost of accurate/faulty					
	2	5	10	20	50	100
0.05						
0.25		Double sampling				
0.50	Accurate only				Faulty only	
1.0						

As to where one is to get valid estimates for the error components, the chairman suggested that what needs to be developed for the survey community is an information system in which the available estimates of the mean square error components for completed surveys are stored and made available for access by survey researchers who are in the process of designing new surveys. Continuous feedback to such a system as various surveys are conducted would allow one to begin to obtain a good picture of how certain characteristics or variables are affected by the various sources of error.

When considering the question of how to estimate the components of bias in TSD, Seymour Sudman came up with a novel solution for the case when one is attempting to obtain the validation information from an institution. In the Sudman procedure, the net bias including the bias due to measurement error and the bias due to nonresponse would be estimated by obtaining aggregate values for the characteristic in question from the institutions from which one sought the validating data. This would allow one to circumvent the necessity for getting individual permissions to obtain the validation information from all individuals in the sample. If one had data from cooperating individuals as well as aggregate values, it would be possible to obtain an estimate of the bias due to nonresponse by subtracting the bias due to measurement error from the total bias.

RECOMMENDATIONS

As a result of the discussion in the Total Survey Design session and also as a result of the subsequent discussions in the remaining day and a half of the conference, the following recommendations are made as to problems that deserve a high priority for research support:

1. Development of a unified total survey error model which combines the mean square error model for errors in collected data with models for nonresponse bias.

2. Development of procedures for using the estimator proposed by Sudman in which aggregate institutional data is used in conjunction with individual record check data to estimate the net bias and its components, that is nonresponse bias and measurement bias.

3. A comprehensive study of the effect of different imputation procedures on the bias due to nonresponse and delineation of the optimum subsampling or follow up nonresponse strategies under each of these imputation procedures.

4. Investigation of alternate procedures or techniques for adjusting data for measurement errors and measurement bias.

RESPONDENT BURDEN

Norman Bradburn, University of Chicago

The topic of respondent burden is not a neat, clearly defined topic about which there is an abundance of literature. In our review of the literature on response effects (Sudman and Bradburn, 1974), we did not explicitly code studies for variables that might indicate total respondent burden. Searches of relevant abstracts and other indexing systems do not show respondent burden as a category that is used to organize methodological studies. Therefore, this paper will be more in the way of a general discussion of the issues that one might consider under the general rubric of "respondent burden", than as an even partial literature review. In our discussion we shall take as the focus of attention the general question: What is the effect of different levels of respondent burden on data quality? By "data quality", I mean such things as response rates, breakoffs, accuracy, amount of missing data, etc.

Before turning to specific topics related to respondent burden, let me first sketch out our general way of thinking about survey research interviews so that we can fit the specific problems into a more general framework that can be used to study other types of variables that may produce response effects. We start with the notion that the research interview is a two-role social system governed by general norms about the behavior of the actors. The two roles are that of respondent and interviewer. The roles are joined by the common task of giving and obtaining information. In the most general sense, the quality of the data is the criterion by which to judge how effectively the task has been carried out.

We distinguish three sources of variation in the quality of the data, that stemming from the characteristics of the respondent, that from the interviewer's performance and that coming from the characteristics of the task itself. We have further divided the task characteristics into 3 broad classes having to do with the task structure, problems of self-presentation and the saliency of the information being obtained. In considering the question of respondent burden, we will be most concerned with task variables, particularly those related to what we have called task structure (see Sudman and Bradburn, 1974, Chapter 2, for full discussion).

Since it is the task that defines the relationship between the actors in the research interview, the notion of respondent burden is most naturally related to variations in the nature of this task. As the task becomes more difficult, *ceterus paribus*, the burden on the respondent increases. On the other hand, since the task is defined as obtaining information from the respondent and the demand characteristics of the situation (cf. Orne, 1969) are such as to require the respondent to give accurate information if he is to be a good respondent, more difficult tasks may be interpreted as more challenging and interesting and subjectively perceived as less burdensome. In discussing the variables that we tend to think of in connection with respondent burden, we should consider the conditions under which a particular type of task may be viewed as more or less burdensome. "Burdensomeness" is not to be an objective characteristic of the task, but is the product of an interaction between the nature of the task and the way in which it is perceived by the respondent.

The interview is a social encounter and is not immune from general considerations that are obtained when people voluntarily participate in social events. The researcher is asking the respondent to provide information, but, on the whole, we do not pay much attention to what motivates respondents to participate in an interview or to what we, as researchers, may do to increase or decrease that motivation, particularly the motivation to perform the respondent role well. In general, we stress contribution to knowledge and/or civic duty as reasons for participating in research. Such reasons appear to be fairly powerful motives as evidenced by the relatively high co-operation rates for serious studies.

But the interview may also be an enjoyable social event in its own right when conducted by trained interviewers who can put respondents at their ease and listen to them sympathetically. E. Noelle-Neumann (1976) has pointed out the importance of proper questionnaire construction for motivating the respondent to participate actively in the interview and to make the effort to give accurate data. Some questionnaires may be boring or tedious, and attention should be given in the design of questionnaires to creating an interesting and enjoyable experience for the respondents. In par-

ticular, the researcher's desire to get extra data fairly cheaply should not be allowed to add so much to a questionnaire that it puts off the respondent and reduces his willingness to participate fully in the research enterprise. If the task is not to be perceived as a burdensome one, attention must be paid to the needs of the respondent, as well as to those of the researcher.

In considering variables related to respondent burden, I shall divide the discussion into 4 main headings; (1) the length of the interview; (2) the amount of effort required of the respondent; (3) the amount of stress on the respondent; and (4) the frequency with which the respondent is interviewed.

LENGTH

Interviews and questionnaires differ greatly in their length as measured by the number of questions, number of words per question, number of pages or other measures of bulk and total length of time to complete the interview. Most investigators think of total length of time to complete the interview of questionnaire as the measure of length. It is typically this figure that is told to the respondent when his co-operation is solicited. Interviews may run from a few minutes to 3 hours or more. While I know of no data on the distribution of the length of interviews in the survey field, my guess is that the mean is around 1 hour with a standard deviation of about 15 minutes. The tail on the upper end is probably quite long. Of course, if one considers repeated interviews, the total length of time being given by the respondent can be much greater. The current longitudinal study of medical care expenditures will require more than 10 hours of interview time per respondent, but this is distributed over more than a year.

There does not appear to be any simple relationship between length of an individual interview and data quality. Within the range of 45 minutes to 1-½ hour, there does not seem to be a clear effect either on response rates or breakoffs, although systematic evidence on the matter is not easy to come by. Nor is there any belief that even substantially shorter interviews have a better completion rate. The experienced field workers I have spoken with believe that while length *per se* does not have much to do with completion rates, at least within these ranges, the longer the interview schedule, the more difficult it is to achieve a high completion rate; that is, length does have some relation to effort, and thus to costs, in getting a high completion rate.

When the amount of time a respondent is asked to give to the interview becomes large, the question of respondent compensation arises. How much time should respondents be asked to contribute free to research? For serious research efforts with some evidence of a contribution to the public good or to scientific understanding, there does not seem to be much difficulty in getting respondents to contribute an hour or 1-½ hours to an interview. When the interview begins to go

beyond that point, we begin to consider monetary of other compensations to the respondent. While there is some evidence that monetary compensation to respondents increases the response rate (e.g. NCHS, 1975) there is no agreement that its effect is truly cost effective; that is, does payment increase the response rate sufficiently to offset the added costs? It may be that the major effect of respondent compensation is not on the respondent at all, although it will have some effect, but on interviewer (or investigator) guilt about imposing such a burden on the respondents. If payment makes the researchers and interviewers feel better about pursuing the reluctant respondents, it may result in higher response rates and better interviewing. (For a more complete discussion of compensation, see Ferber and Sudman, 1974.)

The above comments on length refer to personal interviews. The situation may be different with other data collection techniques. There has been a general feeling that telephone interviewing imposes greater time limitations on the interview than does personal contact. The evidence for this belief, however, is not great. At the 1975 Airlie House conference, the consensus of the participants was that telephone interviews up to an average of 1 hour were quite possible without adverse effects on data quality. I am not sure that there is much experience with longer telephone interviews, but it is not immediately clear that longer ones are out of the question. It does seem likely that longer telephone interviews will need careful scheduling with the respondent so that he is not inconvenienced by tying up his telephone for a long time. Here again a longer interview that was perceived by the respondent as very important could very well result in high co-operation rates. I expect that it would take a higher level of justification to get respondent co-operation.

Intuitively, one would expect that the strongest relationship between length (at least apparent length) and response rate would be in self-administered questionnaires. I have heard several researchers maintain with great conviction that it is extremely important to make self-administered questionnaires not only short, but also to appear to be short. Operationally, this advice leads to reducing the number of pages in the questionnaire to an absolute minimum, even at the cost of crowding more onto a single page. Two studies (Champion and Sear, 1969; Sheth and Roscoe, 1975), however, provide evidence that there is no significant difference in response rate between short and long questionnaires, at least within the range of 3 to 9 pages. With longer questionnaires, more than 12 pages, however, one does find a significant effect (Dillman, publication date 1978).

Even though length might not affect completion rates on a particular study, it may have an effect on follow-up studies with the same respondents. It is difficult to come up with any good evidence one way or the other since most investigators who are planning longitudinal stufies worry about the follow-up rates and adjust their data collection aspirations with such

rates in mind. There is at least anecdotal evidence from one NORC study in which the original interviews were up to 3 hours in length. A 10-year follow-up study was conducted with a sub-sample of the respondents. The length of initial interview was still remembered by many respondents and may have played a role in some refusals for the follow-up study.

On the other hand, the Consumer Expenditure Survey, which is a very long questionnaire with repeated interviews, has a high completion rate (90 percent) and few respondents complain about the survey when reinterviewed. Respondents may be interviewed for two or more hours, five times a year. The survey covers detailed expenditures about sometimes unreasonable items (e.g., asking poor or elderly respondents about purchases of airplanes or snowmobiles) and asks respondents to refer to records and to prepare themselves for the follow-up interviews. The survey, however, is used to form the basis of the cost of living index which has significant income implications for large numbers of people. Both interviewers and respondents may consciously or unconsciously use this information to justify the expenditure of so much effort.

In sum, there is no clear evidence that interview length is in itself an important contribution to response rate, although it is fairly clear that longer interviews are more costly and that for really long interviews, the cost increases are non-linear.

But we should also consider the other side of the coin. Ordinarily when we are in a position to afford longer interviews, it is because the study has been judged of sufficient importance to justify a bigger budget. Whatever it is about the study that contributed to the judgment of importance may also work on the researchers and interviewers to increase their efforts to insure high completion rates and to influence the respondents so that they are willing to make a greater effort to contribute to the study. If length is correlated with importance and importance is correlated with higher completion rates, we might even find a mild positive correlation between length and response rate.

RESPONDENT EFFORT

As with length, the amount of effort required of the respondent in answering questions in a survey differs considerably. Respondents may be asked their opinion on matters with which they are familiar and to which they can respond without much thought. On the other hand, they may be asked for complicated and detailed information about finances (e.g., Housing Allowance Supply Experiment) or expenditures (e.g., Consumer Expenditure Study, Medical Care Cost Study). They may be asked to assemble records in their own homes or they may be asked to come into a central testing site to take tests or submit to a medical examination. To some extent differences in effort are correlated with length, but it is possible to have long interviews that do not require any greater effort on the part of the respon-

dent than a short interview, other than that entailed by the greater number of questions themselves. Since it takes time to assemble records or to go to a central examining location, it is almost always the case that studies requiring great effort on the part of the respondent will also take more time. I know of no studies that try to sort out the effects of total time from those of effort.

While the use of records has complicated effects of the level and accuracy of reporting (see Sudman and Bradburn, 1974, Chapter 3) and, properly used, can improve overall data quality, it is not clear that they have any effect on response rates. As with the case of length, the request to use records may indicate the greater importance attached to the study and thus emphasize the demand characteristics for the "good" respondent to co-operate and provide the most accurate data he can. I do not know of any evidence that asking the respondent to go to greater trouble in the form of consulting his records leads to a lower completion rate. In most instances, it improves the accuracy of reporting and may reduce telescoping. The use of records might thus be justified even if one had to pay some price in completion rate. John Lansing used to say that he would be willing to trade off a huge reduction in completion rates if he could be guaranteed that the economic data he was interested in were accurate. Perhaps the trade-off between data accuracy and sample bias due to low response rates should be explored more systematically.

Effort, as measured by coming into a central examining station, is also an important variable. High completion rates have been obtained even under conditions requiring respondents to make considerable expenditures of time and effort to come to an examining location, as for example with the National Health Examination Survey (HES) which requires respondents to come to a mobile testing station and undergo an extensive physical examination. Response rates on this study were high (between 87 and 96 percent) on the first 3 cycles.

In 1971, however, when the HES was expanded to include responsibility for measuring and monitoring the nutritional status of the U.S. population, the response rate dropped to around 64 percent. It is not clear what factors were responsible for this drop. One hypothesis is that the addition of the nutritional portion of the survey lowered the appeal of the study to the respondents, either because the study was now longer and/or because nutrition is deemed less important. The effect of the change in the HES could partially be offset by respondent remuneration, but it may be that some sort of threshold of effort may have been reached that began to have a serious effect on response rate.

From the fragmentary evidence, I would conclude that when greater effort is required by the respondent, particularly when it means going to some special location for testing, response rates may suffer somewhat and greater efforts on the part of the researchers will be needed to insure high completion rates. Again, as with

length, if respondents perceive the study as particularly important, they may be willing to expend greater effort and perform the role of a good respondent.

RESPONDENT STRESS

By respondent stress, I mean the amount of personal discomfort that a respondent undergoes during the course of the interview. Such discomfort may arise from the content of the questions, such as might result from embarrassing or ego-threatening questions or from those that might provoke emotionally laden responses, or from other activities such as mental or physical tests that might be part of the data collection operation. Other things being equal, one might expect that greater respondent stress would be associated with lower completion rates and/or lower validity of data.

The relationship between stress and completion rates is difficult to determine. It is difficult to know how much respondents are warned in advance about the potentially stressful nature of the material or, even when there are efforts to explain more fully the nature of the interview, how much the respondent actually takes in of what he is being told. With the increased concern for a workable definition of informed consent, some experimental work is underway to determine empirically the effects of differing levels of initial explanation about the content of interviews. Johnson and Delamater (1976) report on a study undertaken for the Commission on Obscenity and Pornography and on several experiments they conducted on response effects in sex surveys. They conclude that there is some differential effect on completion rates within demographic groups, but that co-operation is not obviously more problematic in sex surveys than in surveys on other topics.

Even if it were true that the sensitivity of topics had little effect on response rates, either for the interview as a whole or for specific questions, it still might be the case that respondents evade stressful questions by underreporting. Such a device might be particularly true for topics which have many subsidiary questions that are filtered through a general question of the type: "Have you ever done X?" If the respondent denies ever having done X, he then avoids a whole series of questions about frequency, amount, dates, etc. In a recent methodological experiment, we have some evidence that suggest such evasion of response is going on for those respondents who find particular topics anxiety provoking (Bradburn, et al, 1977). There are more ways to evade a question than outright refusal. Even with complete anonymity, as with the random response technique, we know there is still substantial under-reporting of threatening events (Locander, et al, 1976).

Respondent stress as a variable is more difficult to deal with than variables such as length or effort. While length and effort are fairly constant across all respondents, stress probably involves much more individual variance. While we think of some topics as more threatening or sensitive than others, e.g., illegal behavior, sex, drug use, there still seems to be substantial individual differences in sensitivity to topics. Thus the stratagems for coping with differences in respondent stress may have to depend on finer tuning or adjustment of data based on the data from the individual respondent rather than on some more general procedure that would apply to all respondents.

FREQUENCY OF BEING INTERVIEWED

I have already touched on the problem of repeated interviews under the discussion of length. Clearly, repeated interviews as part of a single longitudinal study pose problems of respondent burden. The difficulties in maintaining high completion rates in longitudinal studies are well known. Many of the difficulties come from locational problems with a mobile population and some come from maintaining co-operation. On the whole, however, the fact that a respondent has previously responded to an interview is the best predictor of subsequent participation, given that he can be located. After several waves of interviewing, one has probably gotten a sample of co-operative respondents who will continue to participate. By that time they know that they are in for it for some time, even if the exact number of waves was not known in the beginning.

There is another source of burden to some respondents about which more should be known. I mean here, the problem of being repeatedly drawn in samples of different and independent studies. As long as one is thinking about national probability samples, the probability of a household falling into two independently drawn samples is small. Survey research organizations, such as NORC, make sure that the same segment of households is not drawn more than once in 5 years. Even with the overlap of the major PSU's, overburdening the same households with interviews, does not yet seem to be a problem.

There are, however, classes of respondents who are more frequently selected in samples and for whom the burden may be perceived as high. When the population is relatively small, as for example, a single professional group, such as physicians or more particularly the specialties, or incumbents or a particular position, such as mayors of cities or members of Congress, the probability of falling into a sample for independent studies becomes fairly high. When the population is very small, as with chairmen of psychology departments, the temptation to do a census is overwhelming and thus one becomes a respondent for every study done of that population.

In the medical area, we may be reaching a point at which guidelines should be developed about the number of surveys a particular respondent should be asked to participate in over a given period of time. High response rates for physicians can still be obtained even when the length and amount of effort required is high, as for example in the National Ambulatory Medical Care Study (NAMC) which requires physicians to fill out a questionnaire for each of their outpatients for a week. With considerable effort and support by the relevant professional societies, response rates averaging 85 to 90 percent are maintained each week. One of the elements in maintaining that rate is the promise to the physician that he will not be asked to be a respondent in the NAMC study more than once in 2 years. As surveys of medical care practitioners become part of a routine monitoring of the medical care system, procedures to protect respondents against over-interviewing will have to be worked out. Otherwise, we run the risk of a major revolt from segments of the respondents that will undermine the entire data collection process.

A study is currently being conducted by the Survey Research Center in co-operation with the Bureau of the Census in which respondents were asked about frequency of receiving questionnaires in the mail, telephone interviews or request for personal interviews. Data from this study are not yet available, but we should have some idea soon about the "frequency" burden on the American population.

CONCLUSION

I have tried to outline some of the issues with regard to respondent burden that are of importance in enhancing the quality of data collected in surveys. The major theme throughout is that respondents seem to be willing to accept higher levels of burden if they are convinced that the data are important. In general, it seems to me the problem is not whether there is a burden level which respondents will not tolerate, but rather how to relate the level of burden to the importance of the data. To a considerable extent, this is controlled by the amount of funding available, since greater respondent burden usually requires more extensive efforts to insure high response rates and good quality data. Perhaps the most serious problem that is not easily related to budgetary control is the increasing use of surveys among specialized populations. In some respects these surveys may have high importance but become burdensome just because the population is so small and the probability of multiple interviews is high. Given the decentralized system of funding and conducting research, it is difficult to see how the overworking of some classes of respondents is to be prevented. But I think we must give some serious attention to the matter or it may be determined for us by others. The recent experience with the attempt to cut down the amount of data supplied by citizens does not suggest a welcome precedent.

53

DISCUSSION OF RESPONDENT BURDEN

Naomi Rothwell, Bureau of the Census,
   Chair
Gary Bridge, Columbia University, Recorder

Logical presentation of the subject would follow this sequence:
   How is respondent burden defined?
   How is it measured?
   What are its effects?
   What can be done to eliminate, alleviate, or counteract any negative effects?
   In this presentation, logic has been sacrificed in favor of reporting first on current developments and new ideas for reducing respondent burden. By starting at the end and working back to the beginning, it is possible to emphasize progress without pretense that the first questions, which contain much of the unaddressed agenda, have been asked or answered adequately.

RECOMMENDATIONS FOR REDUCING RESPONDENT BURDEN

In his paper, Bradburn noted that respondent burden is a *subjective* phenomenon. What is a burden to one respondent may be unimportant, unnoticed or even pleasurable to another. The assumption is that the greater the subjective burden, the lower the respondent's motivation to participate in the survey. Some of the participants, however, stressed objective difficulties of the tasks assigned to respondents. Cannell, for one, felt that some seemingly simple questions in the opinions of researchers are extremely difficult for respondents to answer and require that, in effect, they create their own mental questionnaires. He cited as examples questions about whether the respondent was hospitalized and, if so, how often. He described the mental processes required of respondents if they take seriously the task of replying accurately. Although no conclusions were reached about the relative importance of subjective or objective components of respondent burden, the prescriptions for reducing burden can often be classified according to perceptions of its nature—which is the way the discussion of the issues is outlined here.

REDUCING BURDEN BY OBJECTIVE MEANS

Two aspects of objective burden reduction were considered: One focused on the conduct or content of the interview itself and described ways of making the task easier for respondents in the sample already selected. The other was concerned with sample selection and involved techniques for spreading the burden more equitably among potential respondents.

1. A number of techniques—some familiar and others highly speculative—were proposed as devices for reducing the respondents' actual burden. These can be summarized as follows:

   a. Instrumented sampling and interviewing techniques may make it easier to collect data over time (Bridge). For example, one researcher placed Honeywell "extensor" machines on the desks of randomly selected high school principals. At randomly determined times, the machines beeped and ejected a questionnaire on a precoded data card. The respondent indicated what he or she was doing at the time, and inserted the completed data card back into the machine. Similar applications of this technique are easy to envision in surveys of medical clinics or doctor's offices.

   Bridge cited another example of instrumentation reducing burden and providing more accurate information In studies of hyperactive children's behavior on medication, teachers tended to overreport childrens' restlessness. Insertion of a transducer at the end of their chair legs more accurately signalled their movements while reducing the need for teacher observation. Another measure required teachers to wear an FM transmitter which they used to signal when a child exhibited hyperactive behavior. The signal activated a videotape recorder so that a permanent record of the event was obtained.

In addition to reducing respondent burden, then, the use of instrumentation provides more accurate reporting. It also permits people to provide more data, so that instead of trying to get many offices or respondents to provide a small amount of data each, a smaller sample can, with the help of instruments, provide large amounts of data per respondent. This, of course, calls for appropriate adjustments in statistics.

b. Instead of demanding very precise, but time consuming, responses to some questions, respondent *estimates* may satisfy research needs (Andersen). The alternative is particularly applicable to estimates of relatively frequent events. Reported frequency of hospitalization based on respondents' estimates for a one year period rather than on their efforts to remember every hospitalization produces results which are similar to NCHS frequency reports which ask respondents to remember each event.

c. Much of the bulk of long surveys is due to repetitive attitude scales. Jabine asked whether there has been research to learn whether abbreviated scales could serve the same purpose as those with 30 to 40 items. According to Ware, the Rand Corporation's research on health status perceptions shows that single item measures are often inadequate, especially with lower social class respondents, but long attitude scales may produce "psychometric overkill". Fuchsburg presented confirming evidence that a 14 item scale to measure depression captured 90 percent of those identified as depressed by a 45 item scale. Misclassifying some relatively small portion of the population may be acceptable under some conditions, depending on the survey objectives. Bradburn commented that the psychometric tradition, which is sometimes mindlessly adopted by survey researchers, grew out of a desire to measure individual differences in relatively homogeneous populations (e.g., college students). As a result, long scales were necessary to get reliable discriminations between individuals. Survey researchers, on the other hand, are more interested in between-group differences, and hence psychometrically "weaker" scales may be adequate.

d. Sometimes the use of single direct question provides better results than even short, simple scale items. Two examples were cited where this was the case. Haberman believed that asking a respondent if he has a drinking problem provides as good information as scales designed to learn whether he has.

Bradburn said that asking soldiers whether they would rather be sent to the artic or the tropics if they had their choice provided the only reliable discrimination between people who did well in one or the other place. He suggested that respondents have access to much more information about themselves than they can give researchers in an interview and will synthesize it better than researchers can in the special circumstances when there is possible payoff for telling the truth and no incentive to conceal, exaggerate or provide a socially acceptable but incorrect reply.

e. Attitude scales exemplify the tendency of some researchers to request more data than are necessary to meet survey objectives and thereby create undue respondent burden. Using income questions as his example, Waksberg recommended that detailed income questions be included only when there is an explicit reason for them. In the Current Population Survey conducted by the Bureau of the Census, information about income is obtained in two ways. One is by asking a single global question and another calls for a series of probing questions about every potential income source. The median income is 10 to 15 percent higher when the second method is used and probably is closer to the truth. For many survey purposes, however, the global question is adequate and the additional accuracy unnecessary.

Eckerman described a Research Triangle Institute study, sponsored by the United States Department of Agriculture, designed to measure the financial status of rural families. Very intensive measures of income and assets were made and sources other than the interview with the respondent were used (e.g., snapshots of the facilities and buildings, realtors' assessments). Presumably, these lightened the burden on respondents and provided valid estimates.

Speaking about collecting income information in telephone interviews, Sudman reported that respondents found it easier to answer a series of short alternative questions about different levels of income (i.e., "Is your income over or under $20,000?") than to answer a question which covered all levels simultaneously. Refusal rates dropped by about half, suggesting that respondent burden is eased by use of the easier question. Marcus, in an unpublished study confirmed that use of such cut-off income questions in the telephone sample portion

of a currently conducted UCLA study was preferred by respondents and interviewers.

f. In addition to techniques designed to make interviews easier, examples illustrated the possibilities of providing easier *access* for the respondent, thereby easing the conduct of the survey:

(1) Where the respondent must travel to a special site for the interview or physical examination, the location of the sites can affect respondent burden and hence completion rates. Fink reported that in two studies conducted in New York, both of which involved screening for breast cancer, a completion rate of 65 percent was obtained when respondents could go to any of 23 clinics, but the response fell to 46 percent when in a study 10 years later respondents were asked to go to one of only three central clinics. Past some threshold, convenience of access to the survey site appeared to have been an important component of respondent burden in these special surveys. Below that threshold, however, there was no relationship between the distance women had to travel to the original 23 sites and efforts required to induce them to come. The degree of effort was measured by the number of letters and telephone calls required to obtain participation.

(2) Ease of access is also an issue for studies now being conducted at Peter Bent Brigham and described by Monteiro. Ways in which the identical task (taking a hypertension pill and blood pressure readings daily) can be accomplished are varied. Some patients are paid for coming to the clinic; others are given non-monetary rewards; others are given the blood pressure apparatus to take home and record their own pressure there; others report to health units on the job. Thus, alternatives to ease respondent burden are being tested.

(3) In connection with post survey inquiries of respondents, Scott reported that the single most frequently made suggestion respondents make is that the interviews should be conducted at more convenient times. He described efforts to vary interviewing times to respondents' convenience as failures. Appointments have been costly in terms of dollars and response rates

have suffered in both telephone and personal visit studies.

2. The burden of repeated interviewing falls heavily upon some small, interesting populations (e.g., department heads in the academic disciplines, drug addicts), and more intelligent policies are needed to spread it among available respondents. Some techniques for accomplishing this were described:

a. Wolter noted that a Canadian plan assigned each area or unit a "survey burden," not unlike the tax burden that is assigned to particular civic units. Once people have been surveyed, and their survey burden is paid, they cannot be surveyed again until others have fully "paid up" their share of the survey burden.

b. Somewhat similar plans are used for some government surveys in the U. S. at this time. For example, the Social Security Administration assigns different sets of Social Security numbers to different surveys of beneficiary populations, so that people in a particular group will not be surveyed more often than others (Jabine).

c. The Public Health Service excludes physicians who are participants in the National Ambulatory Medical Care Study (NAMC) surveys from being sampled in other federally sponsored surveys (Eisinger).

d. Research Triangle Institute guarantees that schools will not be resurveyed in a National Assessment of Educational Progress (NAEP) within the four year period of the project's duration (Horvitz).

e. Professional organizations can act as screens and coordinators to weed out duplication. The American Psychological Association has taken some steps to do this (Bradburn).

REDUCING BURDEN BY SUBJECTIVE MEANS

Techniques for reducing the subjective burden without taking less of the respondent's time or physical effort call for improved respondent motivation. The following examples seem rather clear cut instances of efforts to improve motivation and hence reduce perceived respondent burden:

1. Cash incentives may motivate some respondents.

a. Horvitz described the results of paying people 26-35 years old to take a package of exercises requiring about 45 minutes to complete for the National Assessment of Educational Progress. Sampled respondents were not paid for taking one package of exercises, $10 for two, $15 for three, and $20 for four exercises. Currently, they are

57

signing up for an average of 3.9 packages and the completion rate is 83 percent. Before incentives were used, the acceptance rate was 68 percent.

  b. Greenberg suggested caution in making generalizations about the use of monetary or other incentives. Different techniques of appealing to respondents to cooperate must be tested because no one method will be applicable to everybody at all times. He cited the experience of a migration study in which blacks in rural North Carolina were pleased to participate in a survey. Their main incentives were the social rewards of participating and some were insulted by offers of money. After the same people moved to Washington, D.C., they were entirely different and practically refused to participate unless reimbursed for their time. What is motivating obviously varies with the social system in which the survey takes place.

  c. Cash incentives also seem to keep people working at diaries (Sudman), but again, they are needed only if the topic does not interest respondents. Thus, they help for expenditure diaries but do not seem to be needed for keeping health records.

2. Some experimental studies of interviewing (e.g., Freedman and Fraser, 1966) suggest that people will agree to a very difficult interview (e.g., a pantry survey) more readily if they have first been asked for a small favor (e.g., for a drink of water). Many interviewers consciously or unconsciously increase their completion rates by asking for or accepting small favors ("could you help me", "this is my last call today"; "yes, thank you, I would enjoy a cup of coffee") before asking for the large favor of granting a lengthy or difficult interview (Bridge).

3. Other seemingly small differences in interviewer behavior can motivate respondents to cooperate. Marshall cited an example from a Mathematica study of ex-drug addicts. Noticeably higher response rates were obtained in a community where interviewers expressed their own personal belief in the confidentiality and usefulness of results after the standard introduction had been given.

De la Puente and Dalenius also stressed the importance of selecting and training appropriate interviewers who can present surveys positively and motivate people to respond willingly.

4. Interviewers are not the only source of stimulation for respondent cooperation. Fink's report about breast cancer detection studies provided another example of motivating respondents through a sense of participating in a worthwhile or interesting program. He noted that there was a 20 percent increase in breast examinations during the three months following the well-publicized mastectomies of Mrs. Ford and Mrs. Rockefeller. He added that the women who came to the clinic during this period were identical in age, ethnicity, income and education to the populations served before the increase. The rate dropped back to the pre-publicity levels in the following three months.

Improved motivation does not necessarily require dramatic events or publicity similar to that described by Fink. Dalenius reported that the level of cooperation among Swedish people tends to improve when they are told the purposes and uses of a survey.

5. Defining the survey in socially acceptable terms can help motivate respondents. For example, calling a survey a "health study" rather than a survey about alcoholism makes the subject matter less threatening and raises the probability that the respondent will cooperate (Haberman). Woolsey recalled a longitudinal survey of accidental injuries in which reporting declined over time, where upon it was redefined as an accident prevention program.

6. Two kinds of questionnaire modifications or reforms were advocated. They dealt with *content* and *sequencing*. The hypotheses are that adding questions of interest to respondents and changing item sequences or vocabulary can encourage respondents' self esteem, engage their interest, and make the task appear easy or non-threatening. The discussion of specific modifications follows:

  a. The use of "interest getting" questions may motivate the respondent to work hard in the interview. Some times these are questions which are not necessary for meeting the survey objectives (e.g., asking mothers about their children) (Dillman). The practice of starting interviews with a question not necessarily related to study objectives but designed to put respondents at ease goes back at least to 1942 when it was employed by the Department of Agriculture (Rothwell).

  b. While some participants agreed that it is important to avoid placing income questions at the beginning of an interview, there is little empirical evidence about the effect of question position. Market researchers often put income questions at the very front of the interview, because they are used as screeners. Health researchers, on the other hand, almost always put income questions at the very end of the interview, although Hensler has found no difference in putting income queries in the middle instead of at the end. This is an issue about which there is much lore but very little reliable evidence.

c. Keeping respondents interested in the survey can relieve burdens created by boredom. A number of speakers including Bradburn, Bridge, Dalenius, Pope and Scharff suggested that more attention to survey introductions, conversational question wording, and smooth flow of questions could make the interview more understandable and engaging. Bradburn referred to the portion of his paper in which he suggested that German survey researchers have been more concerned than Americans with the problem of motivating respondents.

Ethical Considerations. While techniques for reducing subjective respondent burden without changing objective burden (as measured in terms of time or respondent effort) may increase completion rates and improve data quality, some have ethical implications. Is it fair, for example, to use preconditioning small favor requests to get respondents to agree to something that they would not have agreed to without the preconditioning request (Bridge)? Objectively, they have contributed time that they would not otherwise have given without preinterview manipulation, although subjectively they may end up enjoying the interview experience. In the rush to secure high completion rates, ethical implications of methods can be overlooked. Some data—albeit data with low external validity—show how the public views the use of deception in different kinds of social research studies (Wilson and Wilson, 1976). One danger of using questionnable manipulations to reduce subjective burden is that the pool of "naive" respondents may be used up, making it eventually impossible to do social research (Montiero). Social psychologists' deception experiments may be the worst violators, but some survey methods also violate ethical standards. Eisinger described a deceptive survey of physicians which was falsely introduced as if it were a study of attitudes toward government interventions in the health area—a subject of greater interest to respondents. Once started on that topic, respondents were willing to answer the questions which were less interesting to them. That approach, however, was unacceptable to OMB.

EFFECTS OF RESPONDENT BURDEN

Starting this report with a description of the ways of alleviating respondent burden, makes it appear as if there were agreement that burdening respondents is always a negative thing. A few examples were cited, however, of positive effects of burdening or upsetting respondents during an interview. Bridge reported about Yale University experiments conducted by Irving Janis. One involved interviewing obese women as they entered a weight loss program. Two interviews of equal lenth were employed. One asked only for factual information and the other was a very personal inter-

view probing the effects of obesity on the respondent's social and sexual relationships. After 11 weeks, women who had been subjected to the stressful interviews rated their interviewers as better and as caring more about them. They also tended to lose much more weight than those who had the routine clinical intake interview (Janis, 1972). In another study, stressful interviews on the subject of blood donations resulted in increasing the proportion who volunteered to give blood. In both cases, stressing the respondents resulted in positive effects. Most conference participants, however, cited negative effects of respondent burden, and the overwhelming opinion was that negative effects outweigh any positive ones.

In this section the detection and measurement of sources of burden are considered, and the results point to alternative procedures for dealing with respondent burden.

1. Respondent burden reduces response rates for the particular survey or interview, for the panel or follow up interview, or for subsequent surveys.

   a. The length of the interview or mailed questionnaire was blamed in some cases but others reported that this was not found to be a problem. Here are the specific cases cited:

   (1) A sixteen page booklet mail questionnaire produced a 10 percent decline in response rate as compared with a 12 page questionnaire. At or below 12 pages, however, no differences were observed in response rates based on length. Response rates for these shorter questionnaires has been between 75 and 80 percent. These conclusions are based on results from 48 surveys of varying lengths (Dillman).

   (2) A 10 page questionnaire with 60 questions mailed to dentists produced a low response. The form was reduced to a single page with 10 questions and mailed to nonrespondents including those who had explicitly refused. Half of those who had refused earlier mailed back the short form. The quality of the data was as good as that of the long form (Roberts).

   (3) The 1970 census results suggest that the auspices of the survey may be more important than the questionnaire length. The mail back rate (before deleting vacant and nonexistant addresses) from a single page form was 81 percent and for a 20 page form the rate was 79 percent (Jabine).

   (4) Longer questionnaires may actually be better, depending on the respondent population, according to Axelrod. With only one follow-up, he obtained a 42 percent response from a single page

questionnaire and more than 65 percent response from a 24 page form mailed to lawyers asking them to rate judges.

  b. Frequent interviewing negatively affected response rates in the following situations:

  (1) The Health Services Research and Development Center at Johns Hopkins conducted a pilot study of the NCHS National Medical Care Expenditure survey employing an experimental design in which *frequency* and *method of interviewing* were varied. The initial interview, regardless of subsequent method, was in person. Subsequent interviews were monthly by telephone, bimonthly by telephone, bimonthly in person, and monthly alternating telephone and personal visit. The attrition rate for the last panel was 14.5 percent as compared with about 6 percent for the other groups. When respondents were asked about monthly versus bimonthly interviews, 16 percent preferred monthly, 44 percent bimonthly and 38 percent had no preference (Yaffe).

  (2) In the breast cancer detection study described by Fink, the examinations in the initial study appeared to have a positive effect on participation in the follow up study. The evidence for this was that no differences in response rate to the followup study were found between those who had initially been classified as ready versus those who had been reluctant respondents in the first survey.

2. Inaccurate reporting may result from respondents' unwillingness to assume the burden imposed by long or difficult interviews. Discussion of this point included diverse contributions, as follows:

  a. In a study conducted by himself and Fowler, Cannell started with a sample of people drawn from hospital records. Thirteen percent of those reached in the first mailing, 15 percent of those who responded to the second, and 32 percent of those who required personal or telephone follow up failed to report their hospitalization (Cannell and Fowler, 1963). Cannell's explanation for the observed differences is that a higher proportion of those who responded late saw the survey as burdensome and did not report well.

  b. Telescoping (or misreporting the timing of events which are the subject of the inquiry) was discussed in the context of respondent burden. Cannell described it as resulting from respondents who are unwilling to

think hard enough to be precise. Some attributed the phenomenon to deliberate efforts to shorten an interview and others to inability to remember accurately. Waksberg distinguished between large expenditures (for residential alterations) and important events which are likely to be inaccurately reported because of telescoping and small expenditures or minor events which are more likely to be forgotten.

  c. Yaffe reported that there are data which show that people who have high medical expenditures tend to report less accurately than those who have lower expenditures (Shapiro & Jaffe, 1977). He hypothesized that the difference arises because respondents deliberately begin to underreport as they learn that a longer interview is the consequence of mentioning an additional use of medical services.

  d. On the null results side, studies conducted by Southern Illinois University for the National Center for Health Services Research suggested that varying lengths of interviews up to 45 minutes or an hour produced no trends in quality of data (e.g., score reliability, response bias) related to length of interview or position of particular items within the interview. Ware (1975), who reported those results, concluded that respondent burden has little or no adverse effect on quality of data collected in interviews of an hour or less.

  e. In summarizing the discussions, Bradburn suggested that longer interviews might reduce the cognitive burden of otherwise more difficult interviews. Task simplification resulting from reducing questions to smaller bits makes the interview longer and possibly more boring but less work for respondents.

3. Respondent burden can also affect the retention of interviewers. Warnecke reported that in a study of cancer in Black populations, high interviewer turnover was attributed to the stress of interviewing about upsetting subjects.

MEASURING RESPONDENT BURDEN

Scott described studies designed to obtain information about respondents' reactions to being interviewed. After each new study, the Survey Research Center (University of Michigan) sends a questionnaire to a sub-sample of respondents to learn their reactions. A summary of 18 such post-survey studies revealed a positive correlation between respondents' rating of length and actual length, but a negative correlation between interest in the survey and length—i.e., the shorter the interview the less the interest. Scott suggested that researchers recognized burdensome topics and compensated by designing shorter interviews

about them. Hensler thought the results could be interpreted differently. She said that, having granted a long interview, people explain their behavior to researchers and themselves as having been motivated by the importance or interesting qualities of the interview. (This is a classic example of Daryl Bem's "self-perception theory".)

Although agreeing in part with this suggestion, Scott described the subjects of the two longest and most interesting surveys as perennial favorites. One was mothers' attitudes toward their children and the other was about family growth and development.

By using respondents' appraisals, SRC found higher than average levels of interest in a long term panel survey in which respondents were interviewed between 80 and 90 minutes before and after elections in 1972, 74 and 76. But in a crossection sample, which was interviewed pre- and post-76 election, a slightly higher proportion expressed interest in the survey, and this suggests that repeated interviewing had lowered panel respondents' interest somewhat.

Bradburn, Bridge and Rothwell all agreed that "respondent burden" is not yet clearly defined, and other participants tacitly concurred so little time was spent trying to define the term.

Reversing the order of presentation, as we have done, clarifies the criticisms made by Bridge, as the session discussant, in his written and oral critique. Briefly summarized, he expressed disappointment in the narrow focus of concerns. Since respondent burden effects completion rates, most studies take completion rate as the sole dependent variable. Outcomes of interviewing are, however, multidimensional, and completion rates are but one kind of outcome. These other outcomes, include (a) data accuracy, (b) the respondent's willingness to be interviewed again, (c) changes in the respondent's self-esteem (which are obviously correlated with and mediate willingness to be reinterviewed), (d) the respondents' attitudes; that is, how they are changed or frozen by the interviewing experience, and (e) the respondent's tendency to communicate information to others. These outcomes of interviewing are particularly important in panel survey designs.

On the independent variable side, one and two factor experiments designed to investigate the causes of respondent burden are inadequate. The outcomes of interviewing are complex, and the causes of these outcomes are also complex. Small factorial experiments won't permit tests to the complexity of this reality. Research on mail surveys provides the horrible example. There are something like 200 methodological articles on self-administered questionnaires (Bridge, 1970), but few of these experiments varied more than one or two factors at a time, so potentially important higher-order interactions were obscured. As a result, there are incredible inconsistencies in findings, and we have no theory which could synthesize these disparate findings.

The discussion of "respondent burden" shows some signs of fragmenting and producing the seemingly contradictory findings so apparent in research about mail surveys. The recommendations made to avoid such a situation were: (1) devote more thought to developing an explicit theory of the interview process. This theory will permit selection of variables for research to maximize understanding of respondent burden; (2) avoid one and two factor experiments for the reasons cited.

Among the issues which should be addressed by a theory of response burden is one which was raised in Bradburn's paper, and was discussed briefly by Rothwell and Cannell, who took opposing positions. The issue is whether a trade-off between *data accuracy* and *sample bias due to low response* rates is possible and desirable. Bradburn suggested that it be explored, and Cannell thought such a trade-off was possible and might be desirable. Based on previously reported results of a study about hospitalizations, he believed that people who perceived participation in the survey as burdensome, did not reply except after repeated follow up, and then they tended to report inaccurately. Rothwell expressed the opinion that response bias is not so much the contribution of hostile or reluctant respondents as is it of uninformed and seemingly compliant people. Inability to understand questions and lack of information to answer them are among the major sources of response error, in her opinion. Rather than refuse, people reply the easiest way they can, the way they think is expected or desired, or the way they can best protect their self interest, and as a result they contribute to response bias. In this view, those who contribute most to response bias are least likely to contribute to bias from nonresponse, making a trade-off impossible.

Data collected only from acquiescent people is no more likely to be free of response bias than, for example, are data from mandatory surveys. Some results obtained in the 1960's from the Survey of Residential Alterations and Repairs provided relevant evidence which, however, cannot be considered conclusive. Interviewers identified 3 percent of their respondents as uncooperative. Average expenditures were the same for that small minority as they were for respondents not so identified.

Possible trade-offs and a better understanding of some seemingly conflicting research results, (for example, on the effect of interview length or the importance of question sequence), await development of theory. Issues raised and not discussed, like the relative burdensomeness of telephone versus face-to-face interviewing for long interviews, also might be addressed if there were more knowledge about what really constitutes burden. Finally, the phenomenon of burden must be considered in the context of specific surveys, and attention must be devoted to Fink's assertion that health survey interviewing may be deeply upsetting to people who are ill or who fear illness. Apropos of that, most participants felt that health was a subject people are interested in and enjoy talking about—making health interviews less burdensome than others.

STANDARD MEASURES OF STANDARD VARIABLES

Lu Ann Aday, University of Chicago
Ronald Andersen, University of Chicago

WHAT IS MEANT BY "STANDARDIZED MEASURES?"

Before launching a discussion of "standardized measures" it may be well to consider just what is implied by the idea of "standardized" indicators. The concept of "standard" itself is defined as "something set up as a rule for measuring or as a model to be followed" (Merriam-Webster, 1974). A report by the Social Science Research Council's Center for the Coordination of Research on Social Indicators (1975) further suggests that "standardized measures" represent a "model set" (p. 1) of questions and coding or scaling procedures that can be applied across a variety of data collection activities. One may well ask what "standards" or criteria of judgment were used to determine that any recommended items, however, are indeed the "best" ones and should themselves serve as models or points of reference for independent researchers interested in dealing with the same or comparable concepts. Perhaps the notion of "standardized" measures should also mean that there has been some methodological testing and verification of the validity and reliability of the questions and procedures on which they are based. It is not readily apparent that this principle has been exercised explicitly in the work reported thus far on the development of standardized items for use in social surveys, however (SSRC, 1975). Rather, judgments are usually made on the basis of face and consensual validity by knowledgeable persons in survey methodology drawing upon their own experience and knowledge of literature. While such an approach is most useful to survey methodologists a next step might be more explicit use of criteria validity in suggesting standards.

WHAT ARE THE ADVANTAGES OF DEVELOPING AND USING STANDARDIZED MEASURES?

What are the advantages of developing items that could be recommended for use in a variety of research situations? One advantage would be, of course, that it permits greater comparability of findings across population groups. The power to test a particular model or theory would be greatly enhanced if uniform methods of measuring the relevant concepts could be developed and data collected on them in a variety of settings. For example, models of health services utilization behavior have been tested in several cross-national comparative studies over the past ten years (Andersen, et al., 1970; Kohn and White, 1976). In these studies considerable attention was devoted to developing items which uniformly reflected the conceptual intent of the model across the diverse settings included in the study. Such an effort was crucial to provide the fairest test of the framework and to make possible generalizations about its utility from the findings in the several nations.

Secondly, administering similar questions to different populations further offers the possibility of doing methodological studies of the reliability and validity of the items themselves. Do certain ethnic groups, for example, tend to demonstrate more of a pattern of acquiescence to particular questions than do other groups? Findings of this kind may well challenge the credibility of using the same items in a variety of settings, however.

A third advantage of using standardized measures is that it permits greater comparability of findings over time. This is particularly applicable to time series analysis of selected social or economic indicators of well-being (Wilcox, et al., 1972). Also, secondary analyses of extant data which have been collected at different points in time are greatly enriched if comparable measures are included in the several data sets (Hyman, 1972).

A fourth advantage of using measures that have been employed previously is that greater economies in the construction of the data collection instrument and in the design of the data processing or coding specifications result. For example, if one were interested in doing a study of health care attitudes and behavior and would like to include some items on job satisfaction to see how they were correlated with the prime variables of interest, much time and effort could be saved by incorporating a scale which had been developed and tested by researchers whose major substantive interests were in the area of work satisfaction.

WHAT ARE THE DISADVANTAGES OF DEVELOPING AND USING STANDARDIZED MEASURES?

The disadvantages of using standardized measures should also be noted. One issue, of course, is that items developed for one situation may not be entirely applicable in others (Etzioni and Lehman, 1967). For example, questions used in national surveys may not be as focused or specific as those necessary for studies of particular communities or neighborhoods.

A second limitation is that items developed for one mode of data collection—such as by personal interview—may not be as applicable if other data collection strategies are employed, such as telephone interviews or self-administered mail questionnaires.

A third issue concerns whether standardized items are the most relevant, given the research question being asked. For example, standardized ways of asking and coding occupation questions may not be the best approach in a labor force study.

Another question is whether there might not be changes taking place over time in a concept that are not necessarily reflected in established ways of asking questions about that concept. In the health services research area, for example, medical services are increasingly being delivered via phone and by doctors' agents, such as paramedics or school or company nurses. Do standardized ways of asking people about how many doctor visits they had in a given period of time, for example, necessarily reflect these types of encounters?

A fifth problem to consider is whether a standard way of asking a question is the most appropriate, given the mode of analysis that is planned. Questions eliciting ordinal or categoric responses, for example, may prove to be troublesome if interval level measurements are required in the analytic plan developed for using the data.

And finally—and this relates to the issues raised initially concerning what precisely is implied in identifying "standardized" measures—certain questions may become reified without adequate testing and development. "Standardized items" may be recommended for use with no accompanying data on how valid or reliable they may actually prove to be. It is our impression that numerous items in current health survey questionnaires could probably be traced backward, sometimes through a circumlocutory route via several generations of questionnaires to early classics in the field. While we have previously enumerated advantages of this process, the simple fact that an item has been used before seems a slim reason for adopting it without some consideration of its reliability and validity.

WHAT IS THE CURRENT STATE OF THE ART REGARDING PROGRESS IN DEVELOPING STANDARDIZED ITEMS?

The types of items presented most often as ones which should be used uniformly in household surveys, for example, are the basic social and demographic background items—religion, occupation, education, income, etc. In 1965 the British Sociological Association organized a task force to study the problems and prospects of increasing the comparability of items used in various types of social research. The efforts of that task force culminated in 1969 in an edited volume which summarized approaches to standardizing measurement of the education, occupation, income and family composition variables (Stacey, 1969). Similarly in the United States in 1973 the Social Science Research Council organized a Working Group on Standardization of Survey Background Items, chaired by Philip Converse (Michigan), that prepared a set of recommendations concerning standardized variables that might be used to measure similar types of concepts: family structure and life cycle, ethnic origin and religion, socio-economic status, residence and political/ideological orientation (SSRC, 1975). In neither of the publications resulting from the deliberations of those groups, however, was there any significant discussion of the *methodological* rationale for recommending these particular questions or items instead of others.

Large-scale data collection organizations, which engage in long-term, repeated national studies, such as the Bureau of the Census, Survey Research Center (Ann Arbor), and the National Opinion Research Center (Chicago) also represent sources to which researchers often turn for traditional or recommended ways of asking questions.

In addition, there are sources of different types of sociometric or attitudinal scales and indexes which could be used by researchers interested in analyzing concepts of particular interest, such as social participation, alienation, self esteem, dogmatism, etc. (Miller, 1970; Robinson and Shaver, 1969). In some cases the methodological strengths of these particular items is much better documented than that of the general background items most often recommended for inclusion in social surveys.

In terms of the standardization of the items used most frequently in health care studies, there is no formal task force that has been designated to address the issues in the same way as was the British and Social Science Research Council groups. The National Center for Health Statistics (NCHS) serves as the informal standard bearer in that researchers designing surveys

of health care utilization and morbidity, for example, are quite apt to consult the NCHS Health Interview Survey as a model for the questions they might ask. Researchers may also informally share the wording for comparable items among themselves. This kind of exchange has, for example, been the case between the Center for Health Administration Studies, UCLA and the Rand Corporation in certain of the studies those groups have conducted. The most amount of "standardizing" activity in the health care field is probably occurring in the health attitudes and health status index development arenas. Some of this activity will be discussed in somewhat greater detail later. It would probably be a fair characterization of the "state-of-the-art" in health services research, however, that except for work on specific scales there has been little or no concerted effort to conduct comprehensive validity and reliability tests of items adapted for use in a variety of studies and that more often than not when researchers do include items that have been used before they are also apt to modify them somewhat to fit their needs better—as they perceive them.

WHAT ARE THE PARTICULAR ITEMS AND/OR ISSUES THAT SHOULD BE CONSIDERED IN DEVELOPING STANDARDIZED ITEMS FOR HEALTH CARE SURVEYS?

At this point, it may be well to introduce (1) some of the general issues that may need to be considered in any efforts to develop standardized measures in the health care field, and (2) describe specific areas in which the work might be focused. In an effort to explore the parameters of the problems associated with developing standardized measures for the health care field, the discussion that follows is intended to raise more questions than it purports to answer.

One generic issue concerns what standardized procedures will be developed for actually testing the validity and reliability of items? What models of total survey error, for example, should be applied to evaluate the recommended measures? Other questions overlap those which might be raised in the evaluation of any survey item: What problems might arise in asking the questions of proxies, for example? ..... or of certain population subgroups, e.g., non-English speaking minorities? ... or how effective is the measure in operationalizing all aspects of the concept under consideration?

Specific types of indicators toward which work on the standardization of measures might be directed are, for example, measures of utilization, morbidity, health attitudes and insurance coverage and whether or not

people have a regular source of medical care.

Current sources of data on utilization measures include the National Center for Health Statistics and the Center for Health Administration Studies. Examples of specific issues that would have to be addressed in efforts to develop standardized approaches to measuring utilization include what types of encounters constitute a visit to a physician (e.g., should telephone contacts and paramedic visits also count?); what recall period is most relevant (e.g., two weeks or a longer period of time); how might different dimensions of health care utilization be distinguished (e.g., preventive and illness-related use) or should they be considered separately?

Measuring the need for medical care is another important substantive concern in the health services research arena. Once again, the National Center for Health Statistics is probably the best source of particular items for tapping general health or morbidity levels of a survey population. The Clearinghouse on Health Status Indexes is a good resource for identifying the considerable activity in the whole area of health status index development. Some questions at issue here might be, even if uniform methods of reporting causes of death can be identified, how useful are such measures as indicators of health? Should the descriptions of symptoms in checklists provided respondents be more specific or general in nature? If disability days are used as a measure of need, how can one effectively capture any differences in the intensity or severity of the respective days of disability in terms of how the questions are phrased? How do patient perceptions of health correlate with physician evaluations? How might one adequately and uniformly code conditions reported by lay respondents in a household survey?

A third area of interest to health services researchers is the measurement of health care attitudes. John Ware (Rand) and Barbara Hulka (North Carolina) are currently doing considerable work on the development of standardized scales for measuring patients' attitudes toward medical care. The Institute for Survey Research at Michigan, of course, has a long history of work in the whole area of attitude measurement. The types of issues that may well arise here in any efforts to develop standardized items is if people are asked in general about their attitudes toward the health care system, we are not apt to know any more about the precise experiential referent for their attitudes, but if we are too specific (e.g., how satisfied were you with the time you had to wait in your doctor's office the last time you went?), then might there be some bias that results from focusing on a particular physician and situation?

Other important types of data for health care researchers are such things as the kinds of insurance

coverage and types of regular source of medical care. Once again, the National Center for Health Statistics, the Center for Health Administration Studies and the Rand Corporation are examples of groups that have devoted considerable resources to collecting information on these types of variables. Some questions that might be raised in efforts to develop standardized indicators of these concepts might be: What aids could be used to stimulate the respondent to provide accurate information (e.g., require old bills, cancelled checks, etc. at time of interview)? And what do previous experiences with comparisons of respondent reports with provider record checks suggest about how the questions might be phrased?

WHAT STEPS MIGHT BE TAKEN TO FORMALIZE THE PROCESS OF STANDARDIZING ITEMS FOR INCLUSION IN HEALTH CARE STUDIES?

Though as mentioned at the outset of the previous discussion, many of the problems and issues raised in the development of standardized items for health care surveys plague any efforts to develop indicators of those concepts, there are several questions which we should perhaps ask ourselves before we proceed too fast apace with recommending solutions or approaches to standardizing variables in the health services research field:

1. Can we, for example, decide upon those variables which are of greatest importance to the health care field and for which it is *most* important that standardized measures be developed?
2. Can we develop formal methodological approaches to testing the reliability, validity and feasibility of using certain items? What might some of these approaches be?
3. What organizations and/or individuals might be charged with the responsibility for carrying out such an activity?
4. What methods might be used, via the funding process, for example, to enforce the use of "well-tested" items or should such enforcement mechanisms be considered?

DISCUSSION OF STANDARDIZED MEASURES

Seymour Sudman, University of Illinois at
   Urbana-Champaign, Chair
Thomas Bice, United Hospital Fund of New York,
   Recorder

The discussion was led off by Andersen's glance backward to the history of health services research from which much of current practice derives. Noting that measures of use of health services, health status, attitudes and beliefs and other variables have accumulated from a series of studies beginning in the 1930's, he observed that the current choice of measures often is dictated by tradition. While this practice is desirable in that it has led to a degree of *de facto* standardization, he and several others were concerned that we lack conceptually and methodologically defensible rationales for preferring one measure to another. Most conferees agreed that, in the absence of a considered reason for developing new measures, researchers would do well to employ items that have been used in previous research. It was also generally agreed, however, that more systematic information regarding a variety of methodological characteristics of measures is required to inform and justify choices.

Attention turned to the types of information that would be needed to evaluate particular measures. Several participants stressed the importance of viewing the item or question in the larger context of the survey design in which it is imbedded. Comparability should not be defined in the narrow sense of identical question wording. The properties of an item should be seen as the resultant of all the characteristics of the study. Therefore, researchers should be encouraged not only to compute and report measures of an item's reliability and validity, but also to describe in detail the nature of the population surveyed, the sample design and other features of the context within which it was used.

These considerations led Horvitz to propose that an information matrix is needed for health survey research. Taking the concept as a unit, the matrix would include various question wordings, information about the context within which they were used and data describing their formal statistical properties. Such a matrix would serve several purposes. At a minimum, it would provide an inventory of items to which researchers could turn for guidance in formulating questionnaires. More importantly, however, an information matrix so constructed would be the basis for a science of survey research, in which the many and

varied factors that influence responses to items are codified and analyzed. With such information, researchers could choose measures for particular purposes and situations with some assurance that they will perform in a predictable manner.

Ideas and recommendations regarding the possible dimensions of an information matrix were contributed by many participants. They identified several factors that should be included, owing to their known effects on responses and their usefulness in evaluating survey items. Among these were the mode of administration (face-to-face-, telephone, mail), the nature of the sample (national cross-section, special populations, such as the aged, the poor, various ethnic groups, etc.), the general purpose of the questionnaire and the location of the particular items in the questionnaire. Suggestions were also offered pertaining to types of formal characteristics that would describe items' reliability, validity and temporal stability. Having endorsed the concept of an information matrix, conferees pointed out prototypical efforts already completed or underway including the standards for measurement promulgated by the American Psychological Association (1974), collection and summary of methodological research by the Research Center for Measurement Methods of the Bureau of the Census (e.g., U.S. Bureau of the Census, 1974; Survey Methodology Information System, undated a, b, c, d, e, f), the methodological series on the HIS published by NCHS (1972), the inventory of health-related measures assembled by Reeder and his associates (1976), and the cataloging of standard items by the CCPDS of the National Cancer Institute.

Discussion moved to implementation and recommendations about where to begin. A consensus emerged on the necessity for a feasibility study centering on the likely costs and benefits of the information matrix. Specifically, the following recommendations were made:

1. *The need for agreement on concepts used in health research.* The standardization of concepts must be agreed on primarily by the users of data, although the data collectors must be involved to insure that the concepts are operational. While it is necessary to agree on the concepts used in

health research, it was pointed out that reasonably good agreement has already been reached and the remaining conceptual fuzziness is small. Generally, it is probably better to use broader measures that can be narrowed in later analyses than to use narrower initial measures. Thus, it may be possible to agree to include all medical contacts, both personal and telephone, to physicians and other medical providers in measuring utilization.

2. *Careful consideration and agreement on the dimensions of quality and their relative importance.* This must be done by data collectors. There are several models and substantial experience that will facilitate this task. An OMB committee headed by Monroe Sirken is preparing a report on this topic to be presented at the 1977 Annual Meeting of the American Statistical Association.

3. *The need for agreement on procedures for measuring the components of error in surveys including reliability and validity of instruments, sample biases, interviewer effects and simple response variance.* This is another task for the data collectors. There is no need to invent these procedures. Methods already exist, although not all procedures are of equal quality. Andrews suggested that the merger of structural equation methods and multitrait-multimethod approaches to isolating error components provides one especially promising direction to this effort.

4. *There is a great need for a central source of information on the components of survey error—reliability and validity—of individual items and scales along with the other characteristics of a survey. The major health data collection agencies, NCHS, NCHSR and the Bureau of the Census, are the logical locations for such a central source.* The central source could be located either in one of the suggested agencies or an inter-agency group could be established.

5. *The need for additional research to measure the components of error in surveys.* Two principal methods for conducting such research are special methodological studies and research added on to existing large and/or continuing projects. It is unreasonable to expect that such measures will be gathered routinely unless additional budget is provided.

6. *Determination of the feasibility of a computer retrieval system containing the cumulated measures of the components of survey errors.* A feasibility study would attempt to determine a) the market potential for such a services, i.e., how many users would there be and how intensive would the use be; b) the operational issues in gathering such data and establishing the data bank; and c) costs and time involved in establishing and maintaining such a system. This feasibility study need not wait until points 1-4 have been completed, but could be conducted in parallel with these other projects. It would be highly desirable for NCHS and NCHSR to set up a joint committee along with outside members to facilitate such a feasibility study.

7. Ultimately, the development of standardized measures based on the most valid procedures will be desirable. This will come naturally, however, and forced adoption of standardized variables should be avoided, especially before data on the validity of such measures are available.

HEALTH SURVEY RESEARCH AND THE OFFICE OF MANAGEMENT AND BUDGET

Joseph Duncan, Office of Federal Statistical
  Policy and Standards

When I was invited to speak at the conference, the chairman and I agreed that this would be a very informal exchange, designed mostly so that I could hear some of your interests and concerns. He also suggested that my introductory remarks should focus on the broad set of activities of the Statistical Policy Division, not simply forms clearance which affects all of you in your survey work.

The title of this session is very general. It refers to the relationship between researchers and the Office of Management and Budget. I want to speak about four OMB concerns this evening. I am not going to speak about zero base budgeting, although that, too, of course, has some effect on researchers.

Tonight I will discuss first, the President's program for reducing the burden of paperwork on the public; second, the development of "A Framework for Planning U.S. Federal Statistics, 1978-1989."; third, the current activities related to the reorganization of Federal statistical activities; and, finally, I would like to make a few comments about the reorganization of the Executive Office of the President, of which my office is a part.

On paperwork reduction, I think most of you were confronted with President Ford's program to cut the number of Federal forms by 10 percent. As you know, the Government bureaucracy responded and that goal was exceeded. There was a 12-1/2 percent reduction in the number of different forms that are used by the Federal Government.

But, as some of you know, it had the impact of virtually freezing the clearance process in some agencies. That was not our intent. If you look only at survey research, the paperwork reduction guidelines provided a mechanism so that there would really be very little impact on survey research since as one set of forms expired, an agency could develop another set of forms to conduct another research study or survey.

In the current paperwork reduction effort, President Carter is emphasizing a slightly different twist on the paperwork reduction program—that is the focus on reducing the total man-hours of reporting burden. The impact on statistical activities or survey research should not be heavy since most reporting burden is in administrative and management areas and not in survey and research activities.

Statistics and survey activities account for between 10 and 15 percent of the total reporting burden. To have a major impact on reporting burden, you have to deal with administrative forms, and each of the departments now has its own program to achieve some burden reductions by the end of this fiscal year.

I would comment that many of the new people coming into Government that I have talked with seem to feel, based on their experiences, that there is a large volume of useless data that is collected by the Government. At the same time, they also feel that when they need data for making an individual decision, they never have the data or statistics that they need. I think that the process internally in individual departments for justifying each data collection, for defining the purposes for each data collection very clearly, will be quite valuable and important.

There are a number of other interesting activities related to paperwork reduction which we could talk about, and I would just mention one. In the education area, a contractor has recently completed a handbook describing how to get through the clearance process. It is a very good handbook, and I suggest you examine it if you are interested, remembering that the education area has its own unique situation *vis-a-vis* clearance. Consequently, some of the guidelines are specific to the education arena, but it does discuss the Office of Management and Budget process very well.

Now let me talk about "A Framework for Planning U.S. Federal Statistics, 1978-1989." I think most of you probably read the *Statistical Reporter,* and therefore know that for the last three years our office has been developing this program. Some people ask why we chose the particular dates of coverage. The reason is that the first budget in which that material will really have an impact is the 1978 budget process, and we look largely at the next seven or eight years.

But when we examine efforts, such as the mid-decade census and the total decennial census program, we begin to set forth ideas that will have an impact throughout the 1980's. That is not meant to suggest that we have predicted the state of policy for health

statistics through 1989. We fall short of that, to be sure. But for the overall structure, we have an agenda for improvement that will take most of the decade to achieve.

I would like to review very briefly the process we followed to develop the Framework materials. In 1975, we convened an ad hoc committee from all the major statistical agencies to help design the planning process with the hope that this would effectively interrelate with agency planning activities. We decided that the most efficient way to get the materials drafted was to have our staff draft individual chapters and to have the relevant agencies review them rather than have each agency write its own set of materials in a heterogeneous and uncoordinated fashion.

I think this approach has worked reasonably well. We have some obvious limitations in our staff capabilities and hence we are not fully up to the total task of topics, so some chapters will be better than others. But overall, I think the chapters represent very good working drafts from which to begin.

We began circulating materials to the agencies in October 1976, and we are still circulating some first-draft materials. Many of them now, however, are in the process of public review. We plan for public review and comment to continue through the year.

There are several keys to the Framework that I would mention very briefly. First, in the area of economic statistics, we have recently obtained the results of a project known as the Gross National Product Data Improvement Project. It is a project that was started in 1973. The final report, which will be published this summer, is very detailed, with a line-item-by-line-item analysis of the national accounts; a description of how the numbers are estimated; a description of the data base for the estimates; an analysis of what is wrong with the estimates (that is, the difference between the preliminary and the final numbers); and a discussion of what can be done to improve the estimates. This report will be a major factor in integrating the chapters on economic statistics.

The second major item, which we have for statistical integration in our Framework, is the mid-decade census which was authorized last year by Congress. We have some proposals in the Framework materials describing the use of the mid-decade census to integrate many of the special-purpose social surveys that are undertaken for narrow purposes, and we have some rather controversial ideas for "nested surveys" in the mid-decade census that I am sure will generate much debate in the years ahead.

We also have some of the elementary ideas that were first proposed back in the mid-1930's by the Committee on Government Statistics and Information Systems which issued its report in 1938. We are still trying to implement some things which they recommended—things such as standard concepts and classifications to be used in a variety of areas. For example, we need to have more than the standard industrial classification system which we presently

have to meet the needs of a consistent statistical system. Finally, we have some recommendations concerning the process for setting priorities and for overall planning and coordination controls in various functional areas.

There are a great many topics in the Framework materials. Altogether there are about 1,200 pages of draft materials. Very few people are able, or could be expected, to read all of them, but most find something of interest.

For example, we have a chapter on confidentiality; we have a chapter on health statistics, which I am sure will be of interest to all of you in this meeting; we have a chapter on Federal-State cooperative programs; we have a chapter on statistical methodology; and so forth—altogether about 24 chapters. We can talk about these in the discussion period, if you like, but first let me just touch on two other topics very briefly.

Next, we have the chapter on reorganization of statistical activities. As you know, the President feels that he was given a mandate in the election to reorganize the Federal Government. The Congress has given him reorganization authority, and a reorganization program is now well underway.

The form of that program is, I think, very interesting. It is going to be managed out of the Office of Management and Budget, under the direction of a committee chaired by the President himself.

The OMB team will be a relatively small team. In fact, they are adding 32 positions to OMB, which is a rather small number. But through detailees, borrowed people in OMB and other places, there are about 150 people working on reorganization right now.

Several topics have been selected for initial attention. I am sure you know that the first reorganization report will focus on the Executive Office of the President itself. The President will probably make some recommendations by June 15, 1977, if he holds to the present schedule.

Some other areas that are being attacked initially are equal employment opportunity activities scattered throughout the various agencies and statistical activities. My office has been given responsibility for the statistical reorganization background work within the Office of Management and Budget, and we have a three-phase activity.

Currently we are examining all of the small statistical units defined as follows: units which have fewer than 10 forms labeled statistical, cleared through our office. It turns out that there are between 50 and 60 such small units. As you know, the Commission on Federal Statistics in 1971 was very critical of the small units that collected statistics. Their view was that there was not adequate staff capability, and there was a very poor statistical quality in those activities.

Now that is not universally true, obviously, because some of those small activities collect their statistics by doing things like paying the Census Bureau to do it for them. The quality of that type of activity is different

from somebody sitting down and designing his own survey form within a small unit.

We will have a report to the reorganization task force very shortly, and that will feed into the ongoing activities within each of these departments. As you know, each department has its own internal reorganization activity.

Then as the departments move more deeply into reorganization, we have in the Framework some proposals for consolidation of statistical activities. For example, we propose a National Center for Criminal Statistics, a National Center for Transportation Statistics, and several other consolidations.

Those ideas will feed into the departmental review. Finally, after the overall team comes up with the exchange of activities among various departments (that is, the Government-wide reorganization recommendations), we will be involved in reallocating roles and missions among the various major statistical activities.

I am sure that topic of a central statistical office will be addressed throughout that process. Now, in the meantime, the Executive Office of the President is subject to reorganization. Since my office is in the Executive Office of the President, we are subject to reorganization very shortly.

Some of you know that there have been a number of proposals for how forms clearance can be improved and how statistical policy can be improved. The Commission on Federal Paperwork has some ideas for improving forms clearance. While they have not issued a report yet at this point in time, there are some indicators.

For example, the Co-Chairman of the Commission on Federal Paperwork, Senator McIntyre, has stated that the thing that is wrong with forms clearance is that the statisticians are in charge. The forms clearance process would be better managed, he believes, if it were in management information specialists' hands.

I am not certain that that is going to be the recommendation of the Commission, but he is certainly an influential member of the Commission. The other Co-Chairman of the Commission is Representative Horton. He was the person that created the Office of Federal Procurement Policy, which is in the Office of Management and Budget, but which also reports to the Congress directly. So it is possible that a dual reporting responsibility for the forms Clearance Office will be recommended by the Commission with the Clearance Office reporting both to Congress directly and to the Director of the Office of Management and Budget.

There are many other possibilities, but those are two. You may have seen the report of the Joint Ad Hoc Committee on Government Statistics that appeared in the *Statistical Reporter*. As you know, that report suggested that the Office of Management and Budget has not carried out its responsibilities well. As a followup to the Committee's report, there is now in draft form a report which proposes that a central statistical coordinating agency be created, perhaps composed of the Statistical Policy Division, the Bureau of the Census, and the Bureau of Economic Analysis and all assigned to the Executive Office of the President.

I personally feel that that last proposal is out of tune with the President's goal to cut the Presidential staff by one-third because if you add in the Bureau of the Census, you have moved in the opposite direction by a substantial margin.

There are some other proposals. The National Advisory Committee on Urban Growth Processes suggested creating a new unit of about 250 people to set statistical standards, to coordinate data collection, and to develop long-range growth models of the economy as an input to the Council of Economic Advisers and the White House.

So you see, there is no dearth of proposals, but at this point I really cannot predict for you which way it is going to come out.

With that, let me just open the floor to comments and discussion.

DISCUSSION OF HEALTH SURVEY RESEARCH AND THE OFFICE OF MANAGEMENT AND BUDGET

Bernard Greenberg, University of North
    Carolina, Chair
John Ware, The Rand Corporation, Recorder

In response to a question by Sudman, Duncan discussed the problem of length of review of clearance procedures in OMB, a concern to all survey researchers engaged in federally-sponsored data collection activities. It was noted by him that approximately 67 percent of the instruments/forms submitted for OMB review and clearance during the prior six months were processed within 18 days. The handbook regarding the clearance process (referred to earlier in the talk) estimates one to five weeks for clearance. This range covers the vast majority of submissions. An important consideration, which investigators must keep in mind, is the length of time set aside for review within the Department of Health, Education, and Welfare and other agencies before reaching OMB. Several ways of expediting the OMB review process itself were suggested, as follows: (1) Meetings between OMB and a given agency prior to the publication of an RFP for discussion of the scope, sample size, and methods of collecting the data; many RFP's are defective resulting in rejection by OMB of data-gathering instruments. (2) Discussions between OMB and the parties involved early in the project and prior to construction of instruments and designing the survey. This process often identifies other agencies that may be collecting data for the same purpose or in the same area. Gross problems or violations of survey design can also be identified at this stage thereby preventing further development and submission of a faulty plan. (3) Another device for expediting the review process suggested by Duncan consists of a parallel review within the department.

In a question from Horvitz, several issues were raised; staff costs caused by delays in fielding instruments while review is in process; demands for the study arising from the timeliness of the data; and need to meet a target date (e.g., need to make a report to Congress). This is an issue of whether the costs to the survey outweigh the benefits of the review by OMB, or vice versa. In reply, Duncan commented on the poor planning of survey researchers such as the hiring of field work staff prior even to initiation of the OMB review process. Better planning, for example, would consist of hiring the field staff at the time the instrument package nears the end of the review process. The

speaker emphasized that in his judgment the total benefits of OMB review exceed the costs, and that OMB review could be extremely expedient. An example was cited in which OMB clearance was forthcoming five minutes after submission for a legitimate emergency situation. Much of this was possible as a result of prior planning and discussion with OMB early in the process. With 3,000 instruments to review each year, it is not possible for OMB to provide such rapid turnaround of all requests.

Following this, Woolsey requested examples showing that the clearance process actually improved the quality of the data that were collected in the health field. Duncan complimented the submissions by investigators working in the health field, specifically citing the National Center for Health Statistics as an agency that needs very little review by OMB. Eisinger, a member of the OMB staff concentrating on the health field, was also asked to comment on this question. He indicated that participants at this conference were usually not the persons submitting forms that caused problems during the review process. He cited a recent experience with the Health Interview Survey in which good prior planning resulted in approval within a day or so. Without naming other DHEW agencies, Eisinger cited a few of the problems of surveys fielded by agencies, other than NCHS, which involve project officers who have little or no training in survey methodology. Moreover, he pointed out, unless such projects involve contracts with well-known survey institutions, problems arise.

Many submissions to OMB require modification prior to fielding. Duncan noted that only about ten percent of all submissions are refused clearance. Data were not available regarding a disaggregation of the percentage as to how many were survey form requests and how many were for routine report forms.

The Chairman, Greenberg, posed the issue of whether or not collaboration between a health agency, such as NCHS, and agencies that do not have experience in fielding surveys would be desirable for the purpose of improving the product. The example of the Census Bureau working with other government agencies was cited as an illustration. Duncan noted that it is the policy of OMB to make the various centers within DHEW (e.g., NCHS, NCES) the focal points

within their areas. This can be promoted by encouraging agencies to seek the advice of agencies with focal responsibility, e.g., the NCHS in the health field. There are certain situations within government where this may be a problem when it involves crossing departmental lines.

Attention was focused by Eisinger on the initial screening process within the Public Health Service itself (e.g., that provided by the Public Health Service Office of Data Policy) as a factor in the high quality of submissions to OMB. Thus, many problems are corrected prior to OMB review. Eisinger also clarified the role of OMB in the review process; their intention is not to redesign a survey. Hopefully, OMB will be able to address statistical policy questions within the government and not get involved with the day-to-day statistical operations of the agencies. This can be accomplished only through rigorous technical reviews at agency and departmental levels.

Reeder asked for comments regarding the general view of the OMB with respect to the quality of federally-sponsored health survey research in comparison to survey research in other fields such as agriculture, economics, etc. Duncan noted that comparisons are difficult to make because systematic comparisons are not performed routinely by OMB. He noted that the problems involved in designing and fielding surveys are different in the various survey areas and that the results are also very dissimilar. An example is monitoring in the field of environmental research where relative chaos exists. This problem stems from the fact that there is little agreement regarding what should be measured and how it should be measured. In comparison to the environmental health field, health survey research is far advanced. However, in comparison to survey research in the economic area, Duncan pointed out that health survey research has many more problems, especially in deciding what to measure. Specifically, the health field lacks the unified analytic framework that exists in economics which links the various pieces together so that the data can be collected in a systematic fashion. *The health field is characterized by some indicators that are relatively well understood but the capability of preparing a report which systematically draws data from all of the health surveys for the purpose of saying how the nation is doing in the health area does not presently exist.* Duncan also noted some exceptions to this generalization in the health area such as the NCHS report on shifting patterns of death rates over the years. The quality of information and the ability to synthesize information in that area of health research are relatively well developed.

In attempting to compare health survey research to agricultural survey activities, Duncan called attention to the importance of response rates and OMB policy regarding minimum requirements for surveys, namely, an anticipated 75 percent response rate. Thus, some requests for clearance in the agriculture area have projected response rates of approximately 25 percent and will have to be rejected for that reason. Nevertheless, a few of these will have to be approved because they involve important policy issues and no one has designed a better sampling frame for the agricultural sector. In some instances where those responding to a survey represent only a 25 percent response rate, the respondents may account for a very large proportion of the production (i.e., 90 percent) and the situation may not be as bad as may appear initially. However, this relationship may not be consistently predictable and we may not know which respondents will account for the most production from year to year.

Bradburn posed the question of whether, without imposing double clearances, OMB could become involved in the process sooner (e.g., at the RFP stage) in order to improve the quality of resultant surveys and reduce clearance problems. In response, Duncan indicated that OMB is happy to provide pre-clearance consultation, such as a review at the time of RFP writing, to any agency that requests such assistance. However, if this assistance was requested universally, the resources in OMB would not be sufficient to meet the demand. Also, despite prior agreement as to the strategy and survey at the RFP stage, the final OMB submission may not be approvable because of departures from the original strategy and survey design that were previously agreed upon.

Several questions regarding the necessity of multi-level reviews were raised by Bryant. Responding, Duncan pointed out that, at least in theory, each review stage has its own separate purpose. At the level of NCHS, for example, the technical aspects of the survey instruments and total survey design should be addressed. At the level of Assistant Secretary for Health, the review should focus on the relationship of the survey to other data collection efforts throughout the health sector. The departmental review is largely focused on policy and overall goals regarding the burden placed on the public. Policy questions regarding total burden must be addressed at this level and departments must make choices. The OMB review should focus primarily on government-wide relationships. Most of the problems noted by Duncan resulted when reviews of a technical nature were done poorly at a lower level. Reviews at subsequent levels are based on the assumption that the survey design is technically sound. However, some technical problems still remain when proposals reach OMB.

According to Duncan, the problem is not that there are multiple levels of review but rather the problem is the amount of time that elapses with each review. Submissions often stack up on desks at various stages in the review process not because of concern with the submission but because of personnel overload. This problem can be easily solved with a small increase in the size of the staff, sometimes even on a temporary

basis. A second problem with multiple levels of review occurs when the same questions are raised repeatedly, even when they should have been resolved at a previous level. This problem can be resolved by getting everyone involved at each level to agree on the purpose of that level's review. An example from the education field was cited where clarification had established the purpose of each review.

It was noted that the goals of review at different levels in the health field are pretty well worked out. OMB is eager to discuss issues pertaining to the goals of review at each level whenever questions do arise. However, even when reviews are done well at the department level, other problems may still be identified when submissions reach OMB. For example, three separate governmental agencies have submitted competent proposals to conduct education surveys of the American Indian and it is possible the same problem could arise in the health field. It is the responsibility of OMB to recognize any overlap in such efforts and to correct the conflicts of duplication.

Woolsey called attention to what he considers to be very substantial imbalance in the federal statistical system. The relatively large amount of resources devoted to economic data in which there is a plethora of economic statistics. He lamented the fact that there is a relative paucity of data regarding social statistics and yet Congress consistently mandates more data regarding social well-being. For instance, he noted that we do not even know how many abortions are performed in America each year. Woolsey asked for comment or clarification of how OMB plans to deal with such imbalance over the next decade.

Duncan agreed that this is an important question and noted that he hopes that he will be able to correct this imbalance. Even before his assuming the duties at OMB, however, the budget during the past decade for social statistics has grown relatively faster than that for economic statistics although the focal centers within the department (NCHS, NCES) have not grown as rapidly. He cited the criminal justice area as an example of an entirely new point of focus in gathering social statistics.

With respect to correcting the imbalance regarding social statistics, Duncan pointed out two barriers that need to be removed. The first is an almost insatiable thirst for social statistics and it is difficult to establish proper priorities. Part of this insatiable thirst involves the desire to acquire data for small geographic areas. To satisfy individual community needs, the costs would have to increase as much as a thousand-fold. A second problem is that we do not have the conceptual model and analytic framework in social statistics as we do in the economic field. He indicated that preliminary estimates of profits in the economic field, for example, can be compared with the IRS results in order to improve the data collection process. Unfortunately, the same does not hold true for social statistics. He is,

nevertheless, looking forward to some improvement resulting from the mid-decade census so as to permit social scientists to create some order and better estimation procedures in the various fields of social statistics areas. He hopes that new theory developed by social scientists will make social statistics more focused and better managed.

De la Puente cited problems with the Medicaid program as an example of policy decision-making in the absence of a minimum data set. He asked Duncan to comment on what plans OMB has with respect to determining the content of minimum data sets and insuring the consideration of such sets in policy formation.

Duncan agreed with the desirability of minimum data sets in support of decision-making in policy areas and the responsibility of OMB in this regard. He cited two major problems that make progress toward these goals difficult. First, it is difficult and sometimes impossible, to convince Congress when things are impractical from a statistical point of view. For example, when legislation was proposed for the use of local labor statistics to be used under the Comprehensive Employment Training Act, OMB pointed out that such statistics were not available. The advice was ignored and we were forced to come up with a make-shift system for doing so. Secondly, the concept of a minimum data set has been employed in only a few areas (e.g., education, and health). The common core of data in education, which has been under development for seven years, is a specific example of the difficulty in establishing a minimum data set. It has been difficult to achieve full agreement regarding definitions and terms needed for management at the state and local level. The PSROs are another example where attempts have been made by OMB to get the personnel to use existing data collection efforts. The problem is that the PSROs want data immediately and they want to be in control of their own program.

A byproduct of the current statistical reorganization activity, according to Duncan, may be "more teeth" in some of the functional central activities. The problem may never be solved completely because it is a complicated problem from both the political and bureaucratic points of view.

Warnecke raised additional issues related to statistical reorganization. Duncan was asked to comment upon the relationship between the Privacy Act and the need in health research for record linkages (e.g., a national death registry). Duncan characterized the Privacy Act as a substantial compromise between the House and the Senate resulting in a somewhat internally-inconsistent document.

Duncan indicated that he is on record for protected enclaves of data that absolutely cannot be released on an individual basis (except to other enclaves). Obviously, this will not solve all problems. For example, he raised the question of the conditions under which it is ethical to link health data with income data

when the individual did not consent specifically for such linkage. Also, a control mechanism for making such decisions is needed, e.g., review by some public body. It has not been determined whether these issues will be dealt with during statistical reorganization. In Duncan's opinion, one central statistical agency (which has been proposed) will not be a result of statistical reorganization. There are too many sets of interests regarding statistics in the federal government.

These differences are related to legitimate differences in definitions and data requirements. Duncan's proposal would be, in general terms, a series of large-scale data collection units that are very small in number and a larger set of analytic teams that are more related to specific policy makers. He pointed out that at present there are 108 agencies in the federal government that collect statistical data and that number is far too many and should be reduced to less than twenty.

DEVELOPING ISSUES IN THE ETHICS
OF SOCIAL RESEARCH ON HEALTH

Bradford Gray, National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research

On February 8, 1966, in response to growing public concern, U.S. Surgeon-General William H. Stewart issued a policy statement on the topic of "Clinical Investigations Using Human Subjects." He announced that the Public Health Service would not support research involving human beings unless the grantee had "provided prior review of the judgement of the principal investigator by a committee of his institutional associates." This was the origin of the current system of human subjects review committees.[1] (Incidently, in December, 1966 the Surgeon-General explicitly stated that it was intended that this policy apply to investigations in the "behavioral and social sciences.") In time the Surgeon-General's policy became DHEW policy, then DHEW regulations, and, finally, in the National Research Act of 1974, institutional human subjects review committees became a legislative requirement at institutions receiving PHS support for research involving human subjects.[2] Since then, an obscure amendment to DHEW appropriation bills has made informed consent a requirement in any DHEW funded "program, project, or course which is of an experimental nature."[3] Recent years have also seen legislation (the Privacy Act[4] and the Buckley Amendment[5]) which has affected researchers' access to certain materials in files or records of individuals, and the creation of the Privacy Protection Study Commission.[6] In addition, the National Research Act created the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research.[7] Obviously there has been a lot of activity in recent years concerning the involvement of human beings in research, and most of this activity has had importance to those who would do social research on health.

Some of the concerns underlying this activity pertain to the use of information or data for purposes other than for which they were obtained. The use of hospital or school records, for example, raises serious issues about who should legitimately have access, how to prevent abuses in which disclosed materials are used against data subjects, and the appropriateness of informed consent requirements. These are among issues currently under study by the Privacy Protection Study Commission. Their draft recommendations[8] on the topic of research and statistics reflect an appreciation for the importance of the research use of materials gathered for other purposes, as well as a sensitive concern for the rights and welfare of data subjects. Their basic approach has been to distinguish between the research and administrative use of records, to specify carefully the circumstances under which records can flow from administrative to research use, and to prohibit the reverse flow. The Privacy Protection Study Commission is still a few months away from the completion of its work, and its recommendations are not yet final.

A second area of concern is the protection of the confidentiality of research data, particularly protection against legal process. With a few narrow exceptions, e.g., research on alcohol and drug abuse, there is little statutory protection available to researchers to prevent legal process for disclosure of research records. In recent years, according to a study by James Carroll and Charles Knerr at the University of Syracuse, there have been at least 17 subpoenas issued for research records and 26 other instances of substantial governmental demands for such information.[9] Most of these cases involved either criminal activity or materials relevant to a civil action. The threat of subpoena has produced researcher responses ranging from destruction of identifiers, to shipping data out of the country, to pressing for legislative relief.[10] Regarding the latter, there is considerable activity at the present time, including a possible recommendation from the Privacy Protection Study Commission and the activities of DHEW's Task Force on Health and Medical Records which has been working on a draft confidentiality statute. A complicating factor in all of this is that as the link between research and policy grows, the public's interest in the validity and integrity of research findings increases. The General Accounting Office has a legitimate auditing function which may be brought to bear on federally-funded research, such as the income-maintenance experiments, and there is some indication that public interest lawyers, aware that there have been past allegations of faked data (particularly in pharmaceutical testing), are beginning to express interest in having access to materials which would show whether research of policy importance was con-

ducted properly. They may wish to contact research subjects. (In this regard, in the research in which I have been involved for the Commission for the Protection of Human Subjects, the Survey Research Center at Michigan had to deal with a threat of suit to *prevent* destruction of identifiers in a data set.) In the past, the purpose of subpoenas was generally to obtain information which might be used against the individual data subject. Although these more recent efforts to breech the confidentiality of data do not represent the same kind of a threat to subjects, they may nevertheless involve inconvenience and even an invasion of privacy. No matter what the source, however, all threats to confidentiality share important ethical and methodological implications.

People who have written about the rights of human research subjects generally contend that the consent process should include disclosure to subjects of the limits of the investigator's ability to protect the confidentiality of data, at least in studies where it could be reasonably anticipated that subjects would want knowledge of such limits. However, many survey researchers believe that such disclosures, even if the possibility of subpoena is described as remote, may affect response rates and introduce biases into data. This is an area in which we need more methodological work as well as more creative solutions to ethical problems.[11]

A third, more general area of concern is the protection of human subjects, including such matters as the DHEW regulations,[12] human subjects review committees, and informed consent. It is these matters which comprise the core of activities of the National Commission for the Protection of Human Subjects.

The Commission was established by the National Research Act of 1974 with a mandate to recommend policy for the government on a number of specific concerns such as research on the human fetus, children, prisoners and the mentally infirm, as well as more general policy regarding the ethical standards which should underlie research on human subjects, mechanisms to translate those standards into practice, and guidance on a series of thorny topics such as the boundary between research and practice, the assessment of risks and benefits of research, and the nature and definition of informed consent. (The Commission is not responsible for the content or administration of present DHEW regulations for the protection of human subjects.) The eleven members of the Commission were appointed by the Secretary, DHEW, and include no social researchers. However, social researchers have made their concerns known to the Commission in a variety of ways,[13] and the Commission is becoming familiar with some of the features of social research which makes it vulnerable to unintended effects of protection measures designed with biomedical research in mind. For example, the distinction between informed consent and the documentation of consent (i.e., consent forms) has arisen several times in Commission discussions; though the Commission has not yet made the

relevant recommendations, recognition of that distinction is a prerequisite for introducing some flexibility into requirements for documentation of consent. The Commission is scheduled to complete its work in the spring of 1978.

One interesting aspect of the Commission's mandate is the way it was asked to reach its general recommendations for protection of human subjects. Congress did *not* ask the Commission to study the extent to which the rights and welfare of subjects are jeopardized and to recommend solutions on the basis of such study. Rather, it directed the Commission to specify the basic ethical principles which should underlie the conduct of research involving human subjects and to recommend ways of assuring that research is conducted in accordance with those principles. This may explain why the Commission may not accept some of the most commonly made assumptions about the ethics of research on human subjects. To take a single example, it is generally stated that research involving human subjects requires informed consent, with the theory underlying this requirement, to the extent to which it is considered at all, usually thought to be to protect human subjects from harm. Thus, for example, the DHEW regulations link the need for consent to the presence of risk. The Commission, however, is moving in the direction of saying that the consent requirement derives from a basic stance of respect for human beings as moral agents, or, in the codewords that have developed, "respect for persons." It is not entirely clear yet what this may imply, but it may suggest both that the need for consent is not limited to the presence of risk and that there may be alternatives to consent in certain circumstances. In any event, it is certainly a more sophisticated stance than saying that the need for consent derives simply from the fact that you are doing research.

While one cannot predict with certainty, it appears that the Commission's presently evolving stance regarding the ethics of research involving human subjects may be compatible with some approaches to adding clarity and flexibility into the DHEW regulations to resolve some of the survey research and field work problems which have arisen at some institutions. This might include clarification of the definition of a human subject—now it is not clear, for example, whether a record is a subject; recognition that the principle of respect for persons may not imply the need for informed consent in certain situations in which there is no intervention in subjects' lives; and putting more flexibility into the use of written documentation of consent. Such steps will seemingly not require the Commission to set up separate standards for biomedical and social research.

Thus, survey research may not, in principle, present serious issues for the Commission's developing approach to issues involving protection of human subjects. The most difficult question in survey research would arise when it is felt that disclosure of the purpose of research might affect both the quality of the

data and subjects' willingness to participate. The Commission's approach would not allow for the withholding of information from subjects on the grounds that disclosure might cause them not to participate. That argument has been totally discredited in biomedical research, largely on the basis of the public scandal of 15 years ago after live cancer cells were implanted beneath the skin of geriatric patients at the Brooklyn Jewish Chronic Disease Hospital. These patients were not told of the nature of the implantation, because the researchers believed they would not agree to participate if they were told.[14]

Serious issues regarding the rights of human subjects may be more inherent in evaluation research and social experimentation than in survey research. A recent lawsuit, *Crane v. Mathews,* is instructive in illuminating the problems of applying current ethical concepts and procedural requirements to certain activities.

In 1975 the Georgia Department of Human Resources submitted to DHEW a demonstration and research project called "Recipient Cost-Participation in Medicaid Reform."[15] The proposal requested a waiver of certain regulations to allow the imposition of a co-payment requirement for physician and hospital services under Medicaid. The Medicaid law allows no enrollment fees or premiums and no deduction or cost-sharing charges; however, Section 115 of the Social Security Act allows this restriction to be waived for a period of time by the Secretary, DHEW in the case of an experimental, pilot, or demonstration project which the Secretary judges likely to assist in promoting the objectives of Medicaid. Secretary Mathews used this provision to allow the co-payment feature in the Georgia project for the purpose of "proving, under properly controlled research conditions, that co-payment is in fact a viable option for Medicaid reform."

The project began in January 1976, without any review by a human subjects review committee. DHEW approved the project by using an interpretation that human subjects were not involved in that the co-payment procedure was an administrative change of the sort which occurs in the administration of any program.

Suit was brought in the U.S. District Court for the Northern District of Georgia by two Medicaid recipients, Fannie Crane and Evelyn Jackson, who contended that this was an experiment which should be reviewed and conducted in accordance with DHEW regulations for the protection of human subjects. The court ordered review of the project, including review of the co-payment itself, by a human subjects review committee. The court rejected the contention that the co-payment feature was essentially a policy or administrative change and ruled that it was broadly experimental in nature. During the court hearings, Secretary Mathews attempted to prevent the court from reaching this conclusion by publishing in the *Federal Register* an "interpretation" of the term, "subject at risk," in the DHEW regulations.[16] Secretary Mathews stated that the regulations were designed to

protect subjects at risk in various kinds of standard biomedical research and were "not intended to protect individuals against the effects of research and development activities directed at social and economic changes, even though these changes might have an impact on the individual." Essentially, he attempted to exclude all experimental or demonstration alterations in governmental service programs from coverage by the DHEW regulations.

The court explicitly rejected Secretary Mathews interpretation of the regulations, stating that it "defies logic" to contend that "the actual imposition of co-payment is not within the scope of the regulations." On the contrary, the court held that "requiring a co-payment exposes these individuals to a method which is not a standard or accepted method in meeting their needs." Thus, the court held the project fell under the regulations and required review. In that review, the Department of Human Resources Review Board voted 10 to 2 that the co-payment procedure exposed subjects to risk, because it might prevent them from seeking needed medical care. Under the regulations, the board could have approved the project anyway, provided that they found the benefits of the research to outweigh the risks. Had they done this, informed consent procedures would have been required by the regulations; it is not clear what this would have meant, but it seems unlikely that many people would have consented to the imposition of the co-payment charges. However, the board ruled that because of a series of research design flaws ranging from a lack of controls to poor survey instruments, the risks of the research were not outweighed by the benefits. Whereupon, the project apparently died.

This one case, which was not appealed, hardly resolves the issues raised by social experiments. However, it clearly shows that the ideas and procedures developed to protect human subjects may not be compatible with certain social experiments, particularly those imposed in existing service delivery programs.

This may mean that existing models of the ethics of research are inadequate. If so, an adequate model awaits development. Its elements might include clarification of the proper limits of discretion in the administration of a service delivery program and the development of a better understanding of the consequences of changing programs without pilot or demonstration work. The paper prepared by Campbell and Cecil[17] at the Commission's request is helpful in this regard, as is the Brookings Institution's book on *Ethical and Legal Issues of Social Experimentation.*[18]

However, the incompatibility of human subjects protection procedures and certain social experiments may not mean that those procedures are wrong. It may mean that certain experiments, in fact, do violate important ethical principles and should not be done. In ordinary research involving human subjects, some procedural solutions have been developed to determine the limits of what can be done with human subjects— this is a matter for a review committee to judge by

applying certain criteria and guidelines, and it is a matter for informed consent of subjects. This is how we distinguish operationally between the permissible and the impermissible. One of the most important challenges in the field of social experiments, demonstration projects, and so forth, is the development of satisfactory ways of distinguishing the permissible from the impermissible under the discretionary authority of administrative officials, and, by so doing, protecting the rights and welfare of the participants. This is an issue that the Commission is not likely to resolve in its general deliverations about the protection of human subjects in research.

In concluding, I would offer the opinion that the activities which I have described have not been born out of hostility to research. Both the Commission for the Protection of Human Subjects and the Privacy Protection Study Commission have explicitly worked from the premise that ways of doing research, while protecting the rights and welfare of subjects, can be found. Similarly, among the policy recommendations in Alan Westin's recent study for the National Bureau of Standards— *Computers, Health Records, and Citizen Rights*—is the following: "The importance of health-care evaluation and medical research calls for developing special procedures so that these activities can be carried on without jeopardizing citizen rights."[19] The research community is being asked to come forward and help develop ways to accomplish such goals.

NOTE: Subsequent to the conference, the Privacy Protection Study Commission issued its Report to the Congress on July 12, 1977.[20]

FOOTNOTES

1. For information on the background and development of DHEW policy, see Mark S. Frankel, *The Public Health Service Guidelines Governing Research Involving Human Subjects: An Analysis of the Policy-Making Process,* Washington, D.C.: George Washington University Program of Policy Studies in Science and Technology, 1972; Bradford H. Gray, *Human Subjects in Medical Experimentation: A Sociological Study of the Conduct and Regulation of Clinical Research,* New York: Wiley-Interscience, 1975; Robert M. Veatch, "Human Experimentation Committees: Professional or Representative?" *Hastings Center Reports,* Vol. 5 (October, 1975), pp. 31-40. Some emerging problems with human subjects review committees are briefly described in Bradford H. Gray, "The Functions of Human Experimentation Review Committees," *American Journal for Psychiatry,* forthcoming.
2. P.L. 93-348, Title II.
3. P.L. 94-296, Section 411.
4. P.L. 93-579.
5. The Family Educational Rights and Privacy Act of 1974.
6. Established under Section 5 of the Privacy Act of 1974 (P.L. 93-579).
7. P.L. 93-348, Title II.
8. *Federal Register,* Vol. 41, No. 243, December 16, 1976, p. 55005.
9. Testimony presented by Carroll and Knerr to Privacy Protection Study Commission, January 5, 1977.
10. See Paul Nejelski and Howard Peyser, "A Researcher's Shield Statute: Guarding Against the Compulsory Disclosure of Research Data," in the National Academy of Sciences report, *Protecting Individual Privacy in Evaluation Research,* 1975.
11. Particularly important has been the work summarized in the forthcoming book, *Methods for Assuring Confidentiality of Social Research Data* by Robert F. Boruch and Joseph S. Cecil, to be published by Cambridge University Press in the Rose Monograph Series of the American Sociological Association.
12. Code of Federal Regulations, Title 45, Part 46.
13. These include having a sociologist on the staff, soliciting position papers from social researchers (including Donald Campbell, Albert Reiss, Bernard Barber, and David Mechanic), receiving testimony from social researchers at public hearings, and through informal contacts at Commissioner's home institutions and elsewhere.
14. Chapter One of Jay Katz's book, *Experimentation with Human Beings* (New York: Russell Sage, 1972) was devoted to this case.
15. Most of my information on the details of this case come from an unpublished paper entitled "Clarification Statement Regarding the Georgia Department of Human Resources Human Research Review Board's Review of the Proposal Entitled 'Recipient Cost-Participation in Medicaid Reform,'" by Russell J. Bent, the chairperson of that board.
16. *Federal Register.* Vol. 41, No. 125, June 28, 1976, pp. 26572-26573.
17. Donald T. Campbell and Joseph S. Cecil, "Protection of the Rights and Interests of Human Subjects in the Areas of Program Evaluation, Social Experimentation, Social Indicators, Survey Research, Secondary Analysis of Research Data, and Statistical Analysis of Data from Administrative Records," unpublished paper prepared for the National Commission for the Protection of Human Subjects, Second Draft, January, 1977.
18. Alice M. Rivlin and P. Michael Timpane (eds), *Ethical and Legal Issues of Social Experimentation,*

Washington, D.C.: Brookings Institution, 1975. See also, Henry W. Riecken and Robert F. Boruch (eds.) *Social Experimentation: A Method for Planning and Evaluating Social Intervention,* NY: Academic Press, 1974.

19. Alan F. Westin, *Computers, Health Records, and Citizen Rights,* National Bureau of Standards Monograph 157, Washington, D.C.: U.S. Government Printing Office, 1976.

20. "Personal Privacy in an Information Society." Report of Privacy Protection Study Commission, July 12, 1977, Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. Stock No. 052-003-0095-3.

81

DISCUSSION OF DEVELOPING ISSUES IN THE ETHICS OF SOCIAL RESEARCH ON HEALTH

Leo Reeder, University of California at Los
  Angeles, Chair
Jack Elinson, Columbia University, Discussant

One thing that I can agree with Brad about is that we need more data. It is hard for social researchers not to agree on that kind of issue.

Brad touched only briefly on the issue of privacy and confidentiality. I would like to pick that up and expand on that one issue.

Privacy, of course, is a value. We do need data about who regards what kind of privacy situation as important relative to what. I do not know whether the study by Eleanor Singer is going into this sort of thing, but we need that kind of information. Privacy is not an absolute. We note that the Commission is declaring ethical principles, as if they were to be handed down from on high. If they are going to tell us what privacy is, it seems to me that is hardly a sociological or social science way of going about it.

To illustrate privacy as a value, I will tell a story about Donnie Rothwell. She and I were associated in some work shortly after World War II. We were looking at what soldiers thought about the quality of housing they were getting. She had used an open-ended question for this purpose. When soldiers complained that the barracks were cold, they did not say anything about whether the barracks were dirty or crowded and lacking in privacy. When soldiers did not complain about the cold, but complained about the dirtiness, they still did not complain about the crowdedness and lack of privacy. If the barracks were warm and clean, only then did they complain about lack of privacy. We have here a hierarchy of values in a special situation.

In the history of privacy in the U.S., Alan Westin in his book reminds us of the paper by Brandeis and Warren called "the rights of privacy," or "the right of privacy," which had the following origin. Brandeis was on his way up at Harvard. In the Boston area one of his wealthy friends held a party for a daughter getting married. The next day in the newspaper there appeared a story about the goings-on at this party. This was referred to as the "yellow press" in those days (today it is called "investigative reporting"). Brandeis and Warren set about establishing the civilized right of privacy.

According to Westin, who is quite a spokesman for privacy and confidentiality, the Warren and Brandeis essay was essentially a protest by spokesmen for patrician values against the rise of the political and cultural values of mass society.

The major observation I would like to direct our attention to is the unequal benefits of privacy. The protection of privacy does not benefit all sectors of society equally. I suspect it is valued most by those whom it benefits most and those who would sustain some sort of loss if their privacy were invaded. While this can be applied to all of us to some extent, I suspect that there are some groups that have more to hide than others, for example, persons with untaxed wealth, and persons who have committed serious and undetected crimes who stand to lose quite a lot if their veils of privacy are penetrated.

Practically no survey researcher that I know of, government sponsored or not, is foolish enough to inquire about untaxed wealth. It leaves our economists frustrated with rather wild conjectures and inferences about this subject. Not only wealth, but also income, as we all know, is frequently concealed. We survey researchers have learned that in surveys and polls the question about income is one of those most frequently not answered.

I suspect that poor people and sick people do not complain very much about the lack of privacy, nor does anybody else in very dire need. On the contrary, I suspect that poor people would like the rest of society to know something about their plight so that maybe they could eat better and have better living conditions. We look away from the people in the streets of Calcutta. We wish they would do their living, such as it is, in more private ways. Their lack of privacy offends us.

Similarly, sick people, who place themselves in the hands of physicians and the total apparatus of modern medicine, announce in effect they are ready to forego all semblance of privacy and to permit the grossest invasions of their body, and mind, in the hope that by doing so somehow their distress will be alleviated. In other words, where illness and other forms of distress are concerned, we are ready to abandon our rights, needs, and pretensions of personal privacy. The ills of mankind—malnutrition, starvation, child abuse, slum living—are quarantined with a cloak of privacy.

Tore Dalenius is the editor of a volume called *Personal Integrity and the Need for Data in the Social Sciences,* based on a symposium held in Sweden recently. One of the interesting comments I picked out of his book illustrates the difference between the way Swedes look at this subject and the way Americans do. I do not know whether it is valid or not, we will ask Tore to comment on it, but the comment was that in Scandinavian countries they hardly regard the state as something threatening. On the contrary, the state in Sweden is thought of as a protector of the poor and underprivileged. By contrast, remarks by the then Director of the U.S. Bureau of the Census, noted the suspicion that Americans have of government in general; and what use will be made of government data.

May I cite for you a summary of polls that were made by Yankelovich and others over the past 15 years which shows an increasing suspicion of government, of institutions of all kinds, a very sharp decline in the confidence that Americans have of institutions, coupled with an increase in self-concern and, in a sense, of narcissism.

We all recognize that Commissions are political bodies. Brad has pointed out there are no social researchers sitting on either of these two Commissions. While Brad says they are somehow represented and their voices are heard somehow, the interests of social researchers are not heard in the same way they would be heard if some social researchers were there. Now, it does say that social researchers constitute a very weak constituency, if you will; and have not been able to get themselves on Commissions of this type. The interests of those who are serving on the Commission are more likely to be served than those not serving on the Commission.

As for the so-called basic ethical principles which Brad referred to—what harm social research might do somebody and, on a different level, respect for persons—I suggest that these, too, like the Bill of Rights, are negotiable. These are not absolute principles. They have to be hammered out just as the Bill of Rights is hammered out, case by case, court by court, decree by decree, situation by situation. We shall just have to be in there along with everyone else representing whatever points of view that we have.

I would close saying that we ought—I agree with Brad—we ought to know more about privacy as a value, who wants it and why, and under what circumstances people will trade some privacy for some other benefits. I have to note in passing that a whole privacy industry has grown up. That is, there are those who break privacy like social researchers and doctors, and detectives and journalists. Some of them go to jail for it. There are those who protect privacy, the locksmiths and the private secretaries and the guards and so on; and those whose living is intrinsically related to the maintenance of a privacy relationship with their clients, doctors, lawyers, psychoanalysts, priests, councelors of all kinds, and the like. There is a whole privacy industry. By way of conclusion, I might quote from the Warren and Brandeis article. It said that "the common law has always recognized a man's house as his castle." That is okay for those who regard their houses as castles. The trouble is there are many of us that do not.

REFERENCES

Alvey, W. and C. Cobleigh
  1975 "Exploration of differences between Linked Current Population Survey and Internal Revenue Service Income Data for 1972." Pp. 121-128 in American Statistical Association Proceedings, Social Statistics Section.

Alvey, W. and B. Kilss
  1976 "Further exploration of CPS-IRS-SSA wage reporting differences for 1972." In American Statistical Association Proceedings, Social Statistics Section.

American Psychological Association
  1974 Standards for Educational and Psychological Tests. Washington, DC: American Psychological Association.

American Statistical Association
  1974 "Report on the ASA Conference on Surveys of Human Populations." The American Statistician 28:30-34.

Andersen, R. and O.W. Anderson
  1967 A Decade of Health Services. Chicago: University of Chicago Press.

Andersen, R., et. al.
  1970 Medical Care Use in Sweden and the United States: A Comparative Analysis of Systems and Behavior. Research Series No. 27. Chicago: Center for Health Adminstration Studies.
  1976 Two Decades of Health Services. Cambridge, MA: Ballinger Publishing Co.

Anderson, O.W. and J.J. Feldman
  1956 Family Medical Costs and Voluntary Health Insurance. New York: McGraw Hill.

Andrews, F.M. and R. Crandall
  1976 "The validity of measures of self-reported well-being." Social Indicators Research 3:1-19.

Bailar, B.A.
  1976 "Some sources of error and their effect on census statistics." Demography 13(May):273-286.

Bailar, B. and C.M. Lanphier
  1977 Development of Survey Methods to Assess Survey Practices. Report to the National Science Foundation, April 1977.

Bradburn, N.M., et. al.
  1977 Question Threat and Response Bias. Unpublished manuscript.

Bridge, G.
  1970 Mail Questionnaires: An Annotated Bibliography. Los Angeles: UCLA Survey Research Center, No. 7001.
  1974 Non-response Bias in Mail Surveys. Santa Monica, CA: Rand Corporation.

Bryant, E.E., et. al.
  1975 "A study of the effect of remuneration upon response in the Health and Nutrition Examination Survey - United States." Vital and Health Statistics Series 2,67(October).

Cannell, C.F. and F.J. Fowler
  1963 "Comparison of a self-enumerative procedure and a personal interview: a validity study." The Public Opinion Quarterly 27 (Summer): 250-264.

Champion, D.J. and A.M. Sear
  1969 "Questionnaire response rate: a methodological analysis." Social Forces 47(March):335-339.

Chromy, J.R. and D.G. Horvitz
  1974 "The use of monetary incentives in National Assessment Household Surveys." Proceedings, Social Statistics Section, American Statistical Association, 171-179.

Colombotos, J., et al.
  1975 "Physicians view national health insurance: a national Study." Medical Care 13(May): 369-396.

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

Dohrenwend, B.S.
  1971 "An experimental study of payments to respondents." Public Opinion Quarterly 34:621-624.

Etzioni, A. and E.W. Lehman
  1967 "Some dangers in 'valid' social measurement." The Annals of the American Academy of Political and Social Science 373(September):1-15.

Ferber, R. and S. Sudman
  1974 "Effects of compensation in consumer expenditure studies." Annals of Economic and Social Measurement, National Bureau of Economic Research 3(April):319-332.

Folsom, R.
  1974 "Working Paper No. 2—National Assessment No-Show Analysis: a study of nonresponse bias." National Assessment of Educational Progress (January).

Frankel, M.
  1971 "An empirical investigation of some properties of multivariate statistical estimates from complex samples." Institute for Social Research Monograph, Ann Arbor, Michigan.

Freedman, J. and S.C. Fraser
  1966 "Compliance without pressure: the foot in the door technique." Journal of Personality and Social Psychology 4:196-202.

Greene, S.
  1976 "Maintaining High Response Rates in Studies Involving Repeated Personal Interviews." Presented at a seminar sponsored by the American Marketing Association, October 7, 1976.

Hansen, M.H., et. al.
1961 "Measurement errors in censuses and surveys." Bulletin of the International Statistical Institute 38:Part II:359-374.

Hawkins, D.F.
1977 "Non-response in Detroit Area Study Surveys: A ten year analysis." Working Papers in Methodology, No. 8, Chapel Hill, Institute for Research in Social Science, forthcoming.

Herriot, R.A. and E.F. Spiers
1975 "Measuring the impact on income statistics of reporting differences between the Current Population Survey and Administrative Sources." Pp. 147-158 in American Statistical Association Proceedings, Social Statistics Section.

Horvitz, D.G.
1952 "Sampling and field procedures of the Pittsburg Morbidity Survey." Public Health Reports 67:1003-1012.

Hyman, H.H.
1972 Secondary Analysis of Sample Surveys: Principles, Procedures, and Potentialities. New York: John Wiley and Sons, Inc.

Inderfurth, G.P.
1972 "Investigation of Census Bureau interviewer characteristics, performance and attitudes: a summary." Working Paper 34, Bureau of the Census, U.S. Department of Commerce, June.

Janis, I.
1972 "The effects of contact with supportive persons on adherence to stressful decisions." In Symposium on Adaption to Stress, presented at the meetings of the American Psychological Association, Honolulu, September 3, 1972.

Johnson, W.T. and J.D. Delamater
1976 "Response effects in sex surveys." Public Opinion Quarterly 40:165-181.

Juster, F.T.
1976 "Methodology, Data Quality, and Interpretation: Some Notes on Consumer Surveys." Paper prepared for the Conference on the Economic and Social Outlook, The University of Michigan, November.

Kalsbeek, W., et. al.
1974 "The National Assessment No-Show Study: An examination of nonresponse bias." Pp. 180-189 in Proceedings of the American Statistical Association, Social Statistics Section.

Kanuk, L. and C. Berenson
1975 "Mail surveys and response rates: a literature review." Journal of Marketing Research 12:440-453.

Kilss, B. and M.T. Millea
1975 "Exploration of differences between Linked Social Security and Internal Revenue Service Wage Data for 1972." Pp. 138-146 in American Statistical Association Proceedings, Social Statistics Section.

Kish, L.
1965 Survey Sampling. New York: John Wiley and Sons, Inc.
1974 "Optimal and proximal multipurpose allocation." Pp. 111-118 in Proceedings of the Social Statistics Section, American Statistical Association.

Koch, G.G.
1973 "An alternative approach to multivariate response error models for sample survey data with applications to estimators involving subclass means." Journal of the American Statistical Association 68(December):906-913.

Kohn, R. and K. White
1976 Health Care: An International Study. London: Oxford University Press.

Koons, D.A.
1973 "Quality control and measurement of nonsampling error in the Health Interview Survey." Vital and Health Statistics, Series 2 (March).

Kovar, M.G. and R.A. Wright
1973 "An experiment with alternate respondent rules in the National Health Survey." Pp. 311-316 in the Proceedings of the Social Statistics Section, American Statistical Association.

Lessler, J.T.
1974 A Double Sampling Scheme Model for Eliminating Measurement Errors in Surveys. Institute of Statistics Mimeo Series No. 949, University of North Carolina, August.
1976 "Survey designs which employ double sampling schemes for eliminating measurement process bias." Pp. 520-525 in Proceedings of the Social Statistics Section, American Statistical Association.

Lipset, S.M.
1976 "The wavering polls." The Public Interest 43:70-89.

Locander, W., et. al.
1976 "An investigation of interview method, threat and response distortion." Journal of the American Statistical Association 71(354):269-275.

Love, L.T. and A.G. Turner
1975 "The Census Bureau's experience: respondent availability and response rates." Paper prepared for presentation at the 135th annual meeting of the American Statistical Association, Atlanta, Georgia.

M.R.S. (Market Research Society)
1976 "Response rates in sample surveys: report of a working party of the Market Research Society's Research and Development

Committee." Journal of the Market Research Society 18:113-142.

Madow, W.G.
1965 "On some aspects of response error measurement." Proceedings of Social Statistical Section, American Statistical Association.

Merriam-Webster
1974 The Merriam-Webster Dictionary. New York: G and C Merriam & Co.

Miller, D.C.
1974 Handbook of Research Design and Social Measurement. New York: David McKay Co., Inc.

NCHSR and NCHS (National Center for Health Services Research and the National Center for Health Statistics)
1975 "Advances in Health Survey Research Methods: Proceedings of a National Invitational Conference." Airlie, Virginia, DHEW Publication No. (HRA) 77-3154.

National Center for Health Statistics
1972 "Reliability of estimates with alternate cluster sizes in the Health Interview Survey." Vital and Health Statistics Series 2, Number 52, April.
1975 "A study of the effect of remuneration upon response in the Health and Nutrition Examination Survey." Vital and Health Statistics, Series 2 (October).

Noelle-Neumann, E.
1976 Die Empfindlichkeit demoskopischer Messinstrumente. In Allensbacher Jahrbuch der Demoskipie, 1976. Wien: Verlag Fritz Molden.

Orne, M.
1969 "Demand characteristics and the concept of quasi-controls." Pp. 143-179 in R. Rosenthal and R.L. Rosnow (eds), Artifact in Behavioral Research. New York: Academic Press.

Palmer, S.
"On the character and influence of nonresponse in the Current Population Survey." Unpublished paper, Bureau of the Census, U.S. Department of Commerce, Washington, DC.

Platek, R., et. al.
1977 "Adjustment for nonresponse in surveys." Paper presented at the Symposium on Survey Sampling and Measurement, April, 1977.

Politz, A.N. and W.R. Simmons
1949 "An attempt to get the 'not-at-homes' into the sample without callbacks." Journal of the American Statistical Association 44:9-31.
1950 "Note on an attempt to get the not-at-homes into the sample without callbacks." Journal of the American Statistical Association 45:136-137.

Quenouille, M.H.
1949 "Approximate tests of correlation in time series." Journal of the Royal Statistical Association 11, Series B:68-84.

Reeder, L.G., et. al.
1976 Handbook of Scales and Indices of Health Behavior. Pacific Palisades, CA: Goodyear Publishing Company.

Roberts, R.E., et al.
forthcoming "Dental care for handicapped children reexamined: I - Dental training and treatment of the handicapped." Submitted to Journal of Public Health Dentistry.

Robinson, J.P. and P.R. Shaver
1969 Measures of Social Psychological Attitudes. Ann Arbor: Institute for Survey Research.

Sailer, P. and L. Vogel
1975 "Exploration of differences between Linked Current Population Survey and Internal Revenue Service Income Data for 1972." Pp. 129-137 in American Statistical Association Proceedings, Social Statistics Section.

Scott, J.C.
1971 "Response rate trends 1955-1970, and response by season, size of place of residence and length of interview." Technical Report 5a, Field Office, Survey Research Center, Institute for Social Research. The University of Michigan, Ann Arbor.
1976 "Response to P.468683." Memorandum to the Omnibus Steering Committee, Survey Research Center, The University of Michigan, July 13.

Sechrest, L.
1976 "Another look at unobtrusive measures: an alternative to what?" Pp. 94-107 in Sinaiko and Broedling (eds.), Perspectives on Attitude Assessment: Surveys and Their Alternatives. Champaign, IL: Pendleton Publications.

Shapiro, S., et. al.
1976 "Medical Economic Survey: Methods Study: Design, Data Collection, and Analytic Plan." Medical Care 14 (November):893-912.

Shapiro, S. and Yaffe, R.
Medical Economics Survey Contract HRA 106-74-150. National Center for Health Statistics Final Report 1977. John Hopkins Health Services Research and Development.

Sheth, J. and A.M. Roscoe
1975 "Impact of questionnaire length, follow-up methods, and geographical location on response rate to a mail survey." Journal of Applied Psychology 60(2):252-254.

89

Sinaiko, H.W. and J.A. Broedling (eds.)
1976 Perspectives on Attitude Assessment: Surveys and Their Alternatives. Champaign, IL: Pendleton Publications.

Social Science Research Council
1975 Basic Background Items for U.S. Household Surveys. New York: Social Science Research Council.

Sprately, E.
1973 "An analysis of the efficiency with which Health Interview Survey Data were collected during the first three quarters of 1972." Unpublished Trainee Report, National Center for Health Statistics, July 13.

Springer, P.
1976 "Characteristics of student OASDI Beneficiaries in 1973: an overview." Social Security Bulletin 39(November):3-32. (See especially Technical Note, pp. 23-32.)

Stacey, M.
1969 Comparability in Social Research. London: Heinemann Educational Books, Ltd.

Sudman, S.
1967 Reducing the Cost of Surveys. Chicago: Aldine Publishing Co.

Sudman, S. and N.M. Bradburn
1974 Response Effects in Surveys: A Review and Synthesis. Chicago: Aldine Publishing Co.

Sudman, S. and R. Ferber
1974 "A comparison of alternate procedures for collecting Consumer Expenditure Data for frequently purchased products." Journal of Marketing Research 11:128-135.

Survey Methodology Information System
undated (a) "Use of advance or preliminary letters and inserts." Statistical Research Division, Bureau of the Census, U.S. Department of Commerce, Washington, DC.
undated (b) "Use of incentives." Statistical Research Division, Bureau of the Census, U.S. Department of Commerce, Washington, DC.
undated (c) "Question sequence: placement of screening questions and effect on recall." Statistical Research Division, Bureau of the Census, U.S. Department of Commerce, Washington, DC.
undated (d) "Telephone interviewing." Statistical Research Division, Bureau of the Census, U.S. Department of Commerce, Washington, DC.
undated (e) "Effects of types of postage in mail surveys." Statistical Research Division, Bureau of the Census, U.S. Department of Commerce, Washington, DC.
undated (f) "Collecting and validating data on income and assets." Statistical Research Division, Bureau of the Census, U.S. Department of Commerce, Washington, DC.

Tukey, J.W.
1958 "Bias and confidence in not-quite large samples." (Abstract), Annals of Mathematical Statistics 29:614.

U.S. Bureau of the Census
1974 Index to Survey Methodology Literature. Technical Paper No. 34, Washington, DC.: U.S. Government Printing Office.
1975 Statistical Abstract of the United States: 1975, 96th edition. Washington, DC.

Vaughan, D. and R. Yuskavage
1976 "Investigating discrepancies between Social Security Administration and Current Population Survey Benefit Data for 1972." In American Statistical Association Proceedings, Social Statistics Section.

Walker Research, Inc.
1975 "Favorable findings revealed by Walker Industry Image Study." The Marketing Researcher, April, (800 Kune Road, Indianapolis, IN 46250).
1976 "Updated survey indicates little change in industry image study." The Marketing Researcher, October.

Walsh, T.C.
1976 "Selected results from the 1972-73 Diary Survey." Paper presented at a seminar sponsored by the American Marketing Association, October 7.

Ware, J.E., Jr., and Snyder, M.K.
1975 "Dimensions of Patient Attitudes." Medical Care, 13 (August): 669-682.

Weber, D.
1973 "Who's home when." Working Paper No. 37, Bureau of the Census, U.S. Department of Commerce, Washington, DC, January.

Westin, A.F.
1976 Computers, Health Records, and Citizen Rights. National Bureau of Standards, Monograph No. 157, U.S. Department of Commerce, Washington, DC, December.

Wilcox, L.D., et. al.
1972 Social Indicators and Societal Monitoring. San Francisco: Jossey-Bass, Inc.

Wilson and Wilson
1976 "Legal and ethical aspects of nonreactive social psychological research." American Sociologist (November).

Wolter, K.M.
1975 "A class of regression estimators for Lessler's Double Sampling Scheme." Unpublished mimeographed paper, Research Center for Measurement Methods, U.S. Bureau of the Census, February, 1975.

Yaffe, R., et. al.
forthcoming "Medical Economic Survey: cost effectiveness of alternative survey strategies." Submitted to Medical Care, April, 1977.

90

CONFERENCE PARTICIPANTS

Murray Aborn
National Science Foundation
1800 G. Street, NW
Washington, DC 20006

Lu Ann Aday
Center for Health Administration Studies
University of Chicago
Chicago, IL 60637

Ronald Andersen
Center for Health Administration Studies
University of Chicago
Chicago, IL 60637

Frank Andrews
Institute for Social Research
University of Michigan
Ann Arbor, MI 48104

Morris Axelrod
Department of Sociology
Arizona State University
Tempe, AZ 85281

Barbara Bailar
Research Center for Measurement Methods
U.S. Department of Commerce
Bureau of the Census
Washington, DC 20233

Thomas Bice
Institute for Health Care Studies
United Hospital Fund of New York
3 East 54th Street
New York, NY 10022

Norman Bradburn
Department of Behavioral Science
University of Chicago
Chicago, IL 60637

Gary Bridge
Department of Psychology
Teachers College
Columbia University
New York, NY 10027

E. Earl Bryant
Division of Data Systems
National Center for Health Statistics
Center Building
3700 East-West Highway
Hyattsville, MD 20782

James Byrne
Reports and Statistics Services
Veterans Administration
Washington, DC 20420

Charles Cannell
Institute for Social Research
University of Michigan
Ann Arbor, MI 48106

John Colombotos
School of Public Health
Columbia University
New York, NY 10032

Jean Converse
Institute for Social Research
University of Michigan
Ann Arbor, MI 48104

Archer Copley
Division of Health Services Research and Analysis
National Center for Health Services Research
Center Building
3700 East-West Highway
Hyattsville, MD 20782

Kathleen Crittenden
Department of Sociology
University of Illinois
Chicago, IL 60680

Ronald Czaja
Survey Research Laboratory
University of Illinois
Chicago, IL 60680

Tore Dalenius
Brown University
Providence, RI 02912

Joseph de la Puente
Division of Health Services Research and Analysis
National Center for Health Services Research
Center Building
3700 East-West Highway
Hyattsville, MD 20782

Donald Dillman
Department of Rural Sociology
Washington State University
Pullman, WA 99163

Joseph Duncan
Office of Statistical Policy
Office of Management and Budget
Washington, DC 20503

William Eckerman
Research Triangle Institute
P.O. Box 12194
Research Triangle Park, NC 27709

Richard Eisinger
Office of Management and Budget
New Executive Office Building
Washington, DC 20503

Jack Elinson
National Center for Health Statistics
Center Building
3700 East-West Highway
Hyattsville, MD 20782

Jacob Feldman
National Center for Health Statistics
Center Building
3700 East-West Highway
Hyattsville, MD 20782

Raymond Fink
Research and Statistics Division
Health Insurance Plan of Greater New York
625 Madison Avenue
New York, NY 10022

Robert Fuchsberg
Division of Health Interview Statistics
National Center for Health Statistics
Center Building
3700 East-West Highway
Hyattsville, MD 20782

Cedric Garagliano
School of Public Health
The Johns Hopkins University
Baltimore, MD 21205

Earle Gerson
Demographic Surveys Division
Bureau of the Census
Washington, DC 20233

Bradford Gray
National Academy of Sciences
2101 Constitution Avenue
Washington, DC 20418

Lois Gray
Intramural Research Unit
National Center for Health Services Research
Center Building
3700 East-West Highway
Hyattsville, MD 20782

Robert Gray
Department of Sociology
University of Utah
Salt Lake City, UT 84112

Bernard Greenberg
School of Public Health
University of North Carolina
Chapel Hill, NC 27514

Paul Haberman
School of Public Health
Columbia University
New York, NY 10032

Deborah Hensler
RAND Corporation
1700 Main Street
Santa Monica, CA 90406

Norman Hiller
Reports and Statistics Services
Veterans Administration
Washington, DC 20420

Daniel Horvitz
Research Triangle Institute
P.O. Box 12194
Research Triangle Park, NC 27709

Thomas Jabine
Office of Research and Statistics
Social Security Administration
1875 Connecticut Avenue, NW
Washington, DC 20009

Charles Jones
Statistical Methods Division
Bureau of the Census
Washington, DC 20233

William Kalsbeek
Research Triangle Institute
P.O. Box 12194
Research Triangle Park, NC 27709

Timothy Keller
Program Evaluation
Veterans Administration Hospital, Brentwood
Los Angeles, CA 90073

William Kitching
Division of Health Services Research and Analysis
National Center for Health Services Research
Center Building
3700 East-West Highway
Hyattsville, MD 20782

David Kruegel
Health Services Research and Development
Veterans Administration
Washington, DC 20420

Judith Lessler
Research Triangle Institute
P.O. Box 12194
Research Triangle Park, NC 27709

David Maglott
Intramural Research Unit
National Center for Health Services Research
Center Building
3700 East-West Highway
Hyattsville, MD 20782

Alfred Marcus
School of Public Health
University of California
Los Angeles, CA 90024

Kent Marquis
RAND Corporation
1700 Main Street
Santa Monica, CA 90406

Cheri Marshall
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08540

Robert Mason
Research Triangle Institute
P.O. Box 12194
Research Triangle Park, NC 27709

James Massey
Division of Data Systems
National Center for Health Statistics
Center Building
3700 East-West Highway
Hyattsville, MD 20782

Linda McCleary
Division of Health Services Research and Analysis
National Center for Health Services Research
Center Building
3700 East-West Highway
Hyattsville, MD 20782

Richard Meile
Department of Sociology
University of Nebraska
Lincoln, NE 68508

Peter Miller
Institute for Communication Research
University of Illinois
Champaign, IL 61820

Lois Monteiro
Department of Sociology
Brown University
Providence, RI 02912

Sam Myers
Intramural Research Unit
National Center for Health Services Research
Center Building
3700 East-West Highway
Hyattsville, MD 20782

Lois Oksenberg
Institute for Social Research
University of Michigan
Ann Arbor, MI 48104

Charles Palit
Survey Research Laboratory
University of Wisconsin
Madison, WI 53706

Clyde Pope
Health Services Research Center
Kaiser Foundation Hospital
11526 SW 33rd Avenue
Portland, OR 97219

Leo Reeder
School of Public Health
University of California
Los Angeles, CA 90024

Sharon Reeder
School of Nursing
University of California
Los Angeles, CA 90024

Robert Roberts
Departmentof Psychiatry
University of Texas Medical School
Houston, TX 77025

Gerald Rosenthal
National Center for Health Services Research
Center Building
3700 East-West Highway
Hyattsville, MD 20782

Naomi Rothwell
U.S. Bureau of the Census
Washington, DC 20233

Jaime Salazar
Health Research Services and Analysis, Inc.
1100 Glendon Avenue, Suite 1151
Los Angeles, CA 90024

Jeffrey Salloway
Department of Sociology
University of Illinois
Champaign, IL 61801

Jack Scharff
Social Security Administration
Meadows East
6401 Security Blvd.
Baltimore, MD 21235

George Schnack
Office of Statistical Research
National Center for Health Statistics
Center Building
3700 East-West Highway
Hyattsville, MD 20782

Nancy Schroeder
Health Services Research and Development
Veterans Administration
Washington, DC 20420

John Scott
Institute for Social Research
University of Michigan
Ann Arbor, MI 48104

Gary Shannon
Intramural Research Unit
National Center for Health Services Research
Center Building
3700 East-West Highway
Hyattsville, MD 20782

Sam Shapiro
Health Services Research and Development Center
The Johns Hopkins University
Baltimore, MD 21205

Seymour Sudman
Survey Research Laboratory
University of Illinois
Urbana, IL 61801

Joseph Waksberg
Westat, Incorporated
11600 Nebel Street
Rockville, MD 20857

Daniel Walden
Intramural Research Unit
National Center for Health Services Research
Center Building
3700 East-West Highway
Hyattsville, MD 20782

John Ware
RAND Corporation
1700 Main Street
Santa Monica, CA 90406

Richard Warnecke
Department of Sociology
University of Illinois
Chicago, IL 60680

Dee West
Department of Family and Community Medicine
University of Utah Medical Center
Salt Lake City, UT 84112

Elijah White
National Center for Health Statistics
Center Building
3700 East-West Highway
Hyattsville, MD 20782

Gail Wilenski
Intramural Research Unit
National Center for Health Services Research
Center Building
3700 East-West Highway
Hyattsville, MD 20782

Sherman Williams
Health Services Research and Development Center
The Johns Hopkins University
Baltimore, MD 21205

Kirk M. Wolter
Research Center for Measurement Methods
U.S. Bureau of the Census
Washington, DC 20233

Herbert Wooley
National Institute of Health
Bethesda, MD 20014

Theodore Woolsey
Health Statistics Consultant
8121 Rayburn Rd.
Bethesda, MD 20034

Gooloo Wunderlich
Division of Policy Development and Planning
Department of Health, Education and Welfare
Parklawn Building
5600 Fishers Lane
Rockville, MD 20857

Richard Yaffe
Health Services Research and Development Center
The Johns Hopkins University
Baltimore, MD 21205

94

Current NCHSR Publications

National Center for Health Services Research publications of interest to the health community are available on request to NCHSR, Office of Scientific and Technical Information, 3700 East-West Highway, Room 7-44, Hyattsville, MD. 20782 (telephone: 301/436-8970). Mail requests will be facilitated by enclosure of a self-addressed, self-adhesive mailing label. These publications also are available for sale through the National Technical Information Service (NTIS), Springfield, VA. 22161 (telephone: 703/557-4650). PB and HRP numbers in parentheses are NTIS order numbers. Publications which are out of stock in NCHSR are indicated as available only from NTIS. Prices may be obtained from the NTIS order desk on request.

Research Digests

The *Research Digest Series* provides overviews of significant research supported by NCHSR. The series describes either ongoing or completed projects directed toward high priority health services problems. Issues are prepared by the principal investigators performing the research, in collaboration with NCHSR staff. Digests are intended for an interdisciplinary audience of health services planners, administrators, legislators, and others who make decisions on research applications.

(HRA) 76-3144   Evaluation of a Medical Information System in a Community Hospital (PB 264 353)

(HRA) 76-3145   Computer-Stored Ambulatory Record (COSTAR) (PB 268 342)

(HRA) 77-3160   Program Analysis of Physician Extender Algorithm Projects (PB 264 610, available NTIS only)

(HRA) 77-3161   Changes in the Costs of Treatment of Selected Illnesses, 1951-1964-1971 (HRP 0014598)

(HRA) 77-3163   Impact of State Certificate-of-Need Laws on Health Care Costs and Utilization (PB 264 352)

(HRA) 77-3164   An Evaluation of Physician Assistants in Diagnostic Radiology (PB 266 507, available NTIS only)

(HRA) 77-3166   Foreign Medical Graduates: A Comparative Study of State Licensure Policies (PB 265 233)

(HRA) 77-3171   Analysis of Physician Price and Output Decisions (PB 273 312)

(HRA) 77-3173   Nurse Practitioner and Physician Assistant, Training and Deployment (PB 271 000, available NTIS only)

(HRA) 77-3177   Automation of the Problem-Oriented Medical Record (PB 266 881)

Research Summaries

The *Research Summary Series* provides rapid access to significant results of NCHSR-supported research projects. The series presents executive summaries prepared by the investigators at the completion of the project. Specific findings are highlighted in a more concise form than in the final report. The *Research Summary Series* is intended for health services administrators, planners, and other research users who require recent findings relevant to immediate programs in health services.

(HRA) 77-3162   Recent Studies in Health Services Research, Vol. 1 (July 1974 through December 1976) (PB 266 460)

(HRA) 78-3183   Recent Studies in Health Services Research, Vol. II (CY 1976)

(HRA) 77-3176   Quality of Medical Care Assessment Using Outcome Measures (PB 272 455)

(PHS) 78-3193   Optimal Electrocardiography

Policy Research

The *Policy Research Series* describes findings from the research program that have major significance for policy issues of the moment. These papers are prepared by members of the staff of NCHSR or by independent investigators. The series is intended specifically to inform those in the public and private sectors who must consider, design, and implement policies affecting the delivery of health services.

(HRA) 77-3182   Controlling the Cost of Health Care (PB 266 885)

Research Reports

The *Research Report Series* provides significant reports in their entirety upon the completion of the project. Research Reports are developed by the principal investigators who conducted the research,

and are directed to selected users of health services research as part of a continuing NCHSR effort to expedite the dissemination of new knowledge resulting from its project support.

(HRA) 76-3143 Computer-Based Patient Monitoring Systems (PB 266 508)

(HRA) 77-3152 How Lawyers Handle Medical Malpractice Cases (HRP 0014313)

(HRA) 77-3159 An Analysis of the Southern California Arbitration Project, January 1966 through June 1975 (HRP 0012466)

(HRA) 77-3165 Statutory Provisions for Binding Arbitration of Medical Malpractice Cases (PB 264 409, available NTIS only)

(HRA) 77-3184 1960 and 1970 Spanish Heritage Population of the Southwest by County

(HRA) 77-3188 Demonstration and Evaluation of a Total Hospital Information System (PB 271 079)

(HRA) 77-3189 Drug Coverage under National Health Insurance: The Policy Options (PB 272 074)

(PHS) 78-3204 Experiments in Interviewing Techniques: Field Experiments in Health Reporting (PB 276 080)

Research Management

The *Research Management Series* describes programmatic rather than technical aspects of the NCHSR goals, research objectives, and priorities; in addition, this series contains lists of grants and contracts, and administrative information on funding. Publications in this series are intended to bring basic information on NCHSR and its programs to research planners, administrators and others who are involved with the allocation of research resources.

(HRA) 76-3136 The Program in Health Services Research (Revised 9/76)

(HRA) 77-3158 Summary of Grants and Contracts, Active June 30, 1976

(HRA) 77-3167 Emergency Medical Services Systems Research Projects (Active as of June 30, 1976) (PB 264 407, available NTIS only)

(HRA) 77-3179 Research on the Priority Issues of the National Center for Health Services Research, Grants and Contracts Active on June 30, 1976

(HRA) 77-3194 Emergency Medical Services Systems Research Projects, 1977 (PB 273 893)

(HRA) 78-3202 NCHSR Research Bibliography (July 1, 1976 to June 30, 1977) (PB 273 997)

Research Proceedings

The *Research Proceedings Series* extends the availability of new research announced at key conferences, symposia and seminars sponsored or supported by NCHSR. In addition to papers presented, publications in this series include discussions and responses whenever possible. The series is intended to help meet the information needs of health services providers and others who require direct access to concepts and ideas evolving from the exchange of research results.

(HRA) 76-3138 Women and Their Health: Research Implications for a New Era (PB 264 359, available NTIS only)

(HRA) 76-3150 Intermountain Medical Malpractice (PB 268 344, available NTIS only)

(HRA) 77-3154 Advances in Health Survey Research Methods (PB 262 230)

(HRA) 77-3181 NCHSR Research Conference Report on Consumer Self-Care in Health (PB 273 811)

(HRA) 77-3186 International Conference on Drug and Pharmaceutical Services Reimbursement (PB 271 386)

(PHS) 78-3195 Emergency Medical Services: Research Methodology

Program Solicitations

(HRA) 77-3196 Conference Grant Information

(HRA) 77-3200 Grants for Dissertation Research Support

(PHS) 78-3206 Grants for Cost Containment Research for Health Planning