

NCHSR

RESEARCH PROCEEDINGS
SERIES

**Advances in
Health Survey
Research Methods:**

**Proceedings of
a National Invitational
Conference**

Airlie House, Airlie, Virginia
May 1 - 2, 1975

*This conference was jointly sponsored by the
National Center for Health Services Research
and the National Center for Health Statistics.*

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
Health Resources Administration
National Center for Health Services Research
DHEW Publication No. (HRA) 77-3154

**ORGANIZING
COMMITTEE**

Sherman R. Williams, NCHSR, Chairman
Robert R. Fuchsberg, NCHS
Charles F. Cannell, Ph.D., University of Michigan
Bernard G. Greenberg, Ph.D.
 University of North Carolina
Daniel G. Horvitz, Ph.D.,
 Research Triangle Institute
Leo G. Reeder, Ph.D., University of
 California at Los Angeles
Seymour Sudman, Ph.D., University of Illinois
William M. Kitching, NCHSR
B. William Lohr, NCHSR
Joseph de la Puente, NCHSR

**CHAIRMEN
AND
RAPPORTEURS**

Leo G. Reeder, Ph.D.—
 Overall Conference Chairman

Session Chairmen

Seymour Sudman, Ph.D.
Charles F. Cannell, Ph.D.
Bernard G. Greenberg, Ph.D.
Daniel G. Horvitz, Ph.D.

Rapporteurs

Ronald M. Andersen, Ph.D.
Floyd J. Fowler, Ph.D.
Monroe G. Sirken, Ph.D.
Kirk M. Wolter, Ph.D.

FOREWORD

The National Center for Health Services Research and the National Center for Health Statistics support research in survey methods in order to increase the validity and reliability of measures of health, the availability of health services, and the use of health services. While much technical progress has been made in the refinement of health survey methods and measures, the dissemination of the state of the art to the general health services research community remains problematic, and there is a need to identify needs and priorities for continued research activities.

Recognizing these needs, the two Centers jointly sponsored this invitational conference to bring together leading researchers in health survey methodology. Participants were charged to review the current state of the art of health survey research and to identify areas and issues for continuing research. It is hoped that this digest of conference proceedings will: (a) acquaint health services researchers whose primary skills are not in survey methods with the limitations and difficulties inherent in health surveys, and (b) apprise researchers whose interests and skills are in the area of health survey methodology of the

research needs and priorities identified by conference participants. i

This report could not have made its timely appearance without the dedicated efforts of our internal staff and consultant planning group. It is our pleasure to acknowledge the generous assistance of Sherman Williams, Joseph de la Puente, William Lohr, William Kitching, Rita Delmont, Linda McCleary, Juanita Locke and Annabelle Ridenour, National Center for Health Services Research; Robert Fuchsberg, Monroe Sirken, Elijah White and Nancy Pearce, National Center for Health Statistics; Kirk Wolter, Bureau of the Census; and Leo Reeder, Seymour Sudman, Ronald Andersen, Charles Cannell, Floyd Fowler, Bernard Greenberg, and Daniel Horvitz, of the non-Federal planning group. The major credit for the success of this effort lies with the participants of the conference, the many skilled and dedicated individuals who have committed themselves to the improvement of health services in the United States. The listing of conference participants in Appendix B is token recognition of their outstanding contribution.

Dorothy P. Rice
Director, NCHS

Gerald Rosenthal, Ph.D.
Director, NCHSR

INTRODUCTION

ii Brief Historical Overview

Survey research has a long and honorable tradition. Its roots, especially in health, go back to the population surveys in France in the late 18th Century and to the *Medical Polizie* in Germany (Rosen, 1972). Later, in the late 19th Century, this method of systematically collecting data from populations or samples of populations through the use of personal interviews was elaborated by the British social surveys of Booth and others. Perhaps in no other country, however, has the survey method reached such a broad range of application as in the United States.

Although sample surveys have spread throughout the world, the American experience is particularly broad and versatile. This is due in no small measure to the fact that opinion and attitude surveys, professional and amateur, are an integral part of the administrative structure of power in both political and business life.

The systematic study of errors, bias, and other problems associated with the application of the survey method began in the early 1930's with the work of the U.S. Department of Agriculture and the U.S. Bureau of the Census. These agencies provided impetus to problems of statistical sampling and other measurement problems. The development of techniques of scaling of attitudes developed by Cantril, Likert, Stouffer, and Lazarsfeld as well as Mosteller's work provided great impetus to the adoption of surveys and public opinion polling in the thirties and early forties.

Perhaps the largest and, certainly, one of the most impressive programs of survey research was conducted during World War II by the Research Branch of the Information and Education Division of the War Department. The behavioral scientists associated with this organization carried out over 300 separate studies on army personnel covering a broad array of topics. In what has become almost a classic case of the adoption of social policy based upon survey research, the Research Branch studies of demobilization priorities of troops led to the adoption of the so-called "point system" of demobilization. Other social surveys carried out during this period for

other Government agencies also are illustrative of the application of this technique for social policy purposes.

A principal benefit of these surveys was their strong contribution to the methodology of survey research. These studies addressed themselves to problems of sampling, questionnaire construction, and interviewing. Several volumes and a large number of research papers were subsequently published that have had a significant impact upon the use of sample survey by not only Government but also business, industry, the mass media, and a wide variety of agencies in the health and welfare fields.

The Use of Surveys in Health Research

In the field of health, the survey method has become a major tool for the systematic collection of health-related data. It is used by epidemiologists, statisticians, medical care and health services researchers, medical sociologists, health economists, psychologists, and of course, various Government agencies. In the mid-thirties, the United States Public Health Service undertook what was perhaps the first Government-sponsored survey of the Nation's health. But it was not until the establishment of the National Center for Health Statistics that systematic, periodic health interview surveys were undertaken. This agency immediately addressed itself to the methodological problems of reliability of interview responses, validity, and other nonsampling measurement issues involved in the Health Interview Study. To this day, NCHS is vitally concerned with the methodological improvement of its surveys in order to improve the quality of the data obtained from respondents.

In addition, the National Center for Health Services Research has supported a considerable number of extramural research projects aimed at improving the quality of health surveys. Much of this activity occurs in the course of substantively oriented research projects such as the work of the Human Population Laboratory at the California State Department of Public Health; the Washington Heights Master Sample Survey of Columbia University; the Los Angeles Metropolitan Area Survey at UCLA; the research of

the Center for Health Administration Studies at the University of Chicago; the Survey Research Center of the University of Michigan, and others. Finally, the National Institutes of Health, the Social Security Administration, National Science Foundation, and others have contributed in recent years to the body of knowledge concerning survey methodology especially as related to health. Indeed, it is difficult to separate health and non-health survey methodology; survey methods developed by statisticians or sociologists have direct relevance for health surveys and likewise, epidemiological survey methods are often of critical interest to social scientists.

Despite the considerable advance in survey methodology, it should be noted that systematic studies of the methods of this research tool are of recent origin. There is much to be done and the professional researchers of the many disciplines that use this method are most sensitive to the limitations of the method. Much has been written on the methods and procedures of survey research; there is also a considerable body of literature on the problems of survey research in the field of health and health services. The material discussed in the present volume is not intended as introductory material; rather it is hoped to add to the existing basic knowledge in the field.

Planning for This Conference

The present conference, sponsored by the National Center for Health Services Research and the National Center for Health Statistics, developed out of a series of symposia and workshops that these two units held during the past two to three years. Among other things, these small meetings addressed themselves to specific topics such as: the use of diaries as a memory aid in retrieving data from respondents; scales to measure the dimensions of patient satisfaction; sample designs and data-collection strategies, and so on. In discussions between several participants of these meetings and staff of the National Center for Health Services Research, the need for a national invitational conference was proposed as a way to synthesize the state of the art with respect to certain key methodological concerns and to identify needs and priorities for additional research.

Subsequently, a planning committee was appointed, and this committee determined that the most useful format for a national conference would be a

relatively small number of invited participants utilizing a semi-structured program. Thus, no papers were to be prepared for this meeting.

Rather, the planning committee prepared an agenda of four major topics that included a number of salient issues under each topic. A planning committee member served as the chairman for one topic on the agenda; each chairman invited a specific individual to serve as the rapporteur or recorder for his session. Each of the major topics was given approximately one-half day for open or free discussion. The objectives of this conference were:

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 scientists.
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.

iii

We hope that this report of the conference will be found useful by those who read it. No attempt has been made to present a verbatim transcript; rather, the chairman and recorder of each session have prepared a report that presents the various issues that were discussed, the comments made about them by various participants, and a summary and recommendations.

This report is tentatively planned as Volume 1 of a series of such conference proceedings on advances in health survey research methods. It is hoped that conferences and reports such as this will occur on a biennial or triennial basis.

References

- Rosen, George, "The Evolution of Social Medicine," *Handbook of Medical Sociology*, 1972, pp. 30-60.

CONTENTS

iv	i	Foreword
	ii	Introduction
	1	Summary and Conclusions
	4	Policy Issues and Communication of Results
	7	Health Survey Research Instruments
	13	Interviewers and Interviewing Techniques
	23	Validity Problems
	36	Total Survey Design
	42	Index of Subjects
	44	Glossary
	47	Consolidated References
	55	Conference Program
	56	Attendees

SUMMARY AND CONCLUSIONS

Leo G. Reeder, Ph.D.

Introduction

Attempting to summarize a technical working conference is always a difficult task; it is doubly so in the present instance because the content of the conference was so rich in substantive contributions. Conference participants are generally not noted for their post-conference enthusiasm; the present meeting was a rare exception in that several participants took the time to write letters not only to the Conference Planning Committee but also to high-level government officials favorably commenting on the format and outcome of the conference. In the final analysis, however, the usefulness of any conference is determined by its products, only a portion of which is visible in the form of a proceedings such as this volume.

The conference was organized to discuss the state of the art of knowledge in certain critical methodologic areas of health survey research and to identify problems for high priority research on non-sampling measurement errors. Four main topics were selected for review: (1) research instruments; (2) interviewing; (3) problems of validity; (4) total survey design. Critical methodological issues in each of these areas were isolated and discussed with reference to current state of knowledge and the needs for further research.

In many respects, health surveys have advantages over surveys in other areas. As the conference participants noted, health surveys are usually more complex and are longer than other surveys. Despite the complexity and length, respondents participate in substantially high proportions, and frequently enjoy the opportunity to respond to questions about their health. The importance of this commitment and willingness to participate in health surveys should not be understated because such a commitment provides an unusual environment and opportunity for health investigations and simultaneously places a serious obligation on the shoulders of investigators provided with the time and trust of the respondent.

In order to fulfill our responsibilities to our scientific colleagues and to participating respondents, the conference gave special emphasis to the concept of total survey design which attempts to provide a

framework for the assessment of survey errors and their cost components.

Total survey design is a concept that facilitates the planned allocation of resources towards the optimal reduction of the total error of estimate. The urgent need for survey results is dramatically portrayed by the number of dollars being invested to conduct the surveys, and the policy decisions being made as a consequence of the availability or lack of availability of results. Similar investments in the development of more efficient surveys are also dramatic in their paucity. The fact is that each component of error: sampling, response, interviewers, and their interactions, are manageable at a cost and do affect the results of every survey. While studies addressing single component of error are important, consideration of all components of error, as an integral part of such studies, is essential.

Throughout the conference, there was consensus that while the results of survey research document the methods for controlling and monitoring various specific sources of errors, the continued development of an information matrix depicting not only methods but costs should be actively sponsored.

Knowledge and application of survey methodology is an indispensable ingredient in the orderly growth of health services research. Most of the useful products of research depend on the applicability, validity, and reproducibility of survey results. Indeed, policy decisions are being made today on the basis of evidence obtained by conducting national, area, and local surveys. Paradoxically, those responsible for designing, conducting, analyzing, and providing the results of surveys, seldom have all the specialized expertise vital to the successful design, conduct and reporting of the surveys being conducted.

The questions of validity and reliability often focus upon sampling errors as the major measurement issue, whereas non-sampling measurement errors are equally important. While many journal articles report sampling errors, very little attention is given to the reporting of non-sampling errors. This conference focused particularly on the importance of addressing non-sampling measurement errors.

Surveys are conducted through a variety of media: mail, telephone, and face-to-face interviews as well as through several variants and combinations of these. The conference did not attempt to address all of these forms of surveys in a systematic fashion, however, since face-to-face and telephone interviews are the most frequent types of survey used in health studies. Mail surveys are used more often for limited initial or follow-up to a personal interview.

This report is being developed to serve two groups of health services research specialists: those engaged in methodological research in health or other areas and those investigators who are the users of survey research methodology in their substantive work.

2

Obtaining valid and reliable data from respondents, i.e., various kinds of publics of interest to health researchers, has always been a matter of concern to research investigators. While considerable progress has been made in the development of better procedures to elicit valid and reliable responses from our various study populations, certain problems remain to be solved. This conference afforded a rare opportunity for an exchange of knowledge concerning a variety of issues including: use of proxy respondents in obtaining data; procedures to aid recall (such as use of diaries and memory aids); questionnaire length; use of the telephone and so on. Basically, the conferees were concerned with improvements in the quality of the data obtained from survey respondents through the use of such data-collection instruments and several recommendations were made of problem areas where methodological research should be given priority.

There is another fundamental aspect to data-collection in survey research, namely, the quality of the *interviewing process* itself and the relationship between interviewers and respondents. It became clear that telephone interviewing was a feasible alternative to face-to-face interviewing but questions remain concerning sample representativeness, interviewer characteristics, and on the quality of the data obtained. Clearly, this procedure requires additional research and merits high priority in methodological studies to improve health surveys. In addition to reaching consensus on the usefulness of the telephone in interviewing, the conference participants also agreed that racial matching of interviewers and respondents was unimportant when non-racial issues were the subject of the investigation. But the problem of relative status differentials between interviewers and respondents merits further study. One of the least understood facts in survey research is that the *quality* of interviewing can affect the data as much or more than response rates, sample design, and so on. The conference indicated that too little is known about the role of interviewer behavior on interview results. There is virtually no systematic body of research data on the evaluation of differential interviewer training strate-

gies so that appropriate guidelines for better methodology can be established.

Governmental agencies supporting studies using the survey method might give serious consideration to the development of a set of guidelines to be used by research grant applicants or contractors. Such guidelines can assist applicants to adhere to established "good practices" in survey methods and procedures.

A recurrent problem in survey research, particularly studies concerned with the collection of sensitive or confidential data, is the issue of validity. How can bias caused by deliberate or unintentional untruthful reporting be reduced or perhaps eliminated? Several procedures were discussed at this conference that suggest reasonable avenues for accomplishing such a goal. Such procedures as randomized response, coding systems, multiple respondent or network surveys, and so on, were considered and evaluated. Clearly, these procedures have much utility in survey research and merit wider applicability. Nevertheless, considerable additional research appears to be required. For example, we know very little about the acceptability of these alternative procedures to the respondent; moreover, our knowledge is scanty with respect to the utility of these newer procedures in other than face-to-face interviews such as mail questionnaires and telephone interviews. The conference participants were also concerned about the effects of recent legislation aimed at protecting the privacy of individuals on the legitimate research and validation procedures which have been used by statisticians and others for decades. Certain pitfalls were discussed in studies involved with record linkage as a means of checking validity.

A major issue in survey research concerns the matter of costs. Although professional survey researchers are aware of this issue, it tends to be considered independent of other variables in the design and conduct of methodological studies. The conference highlighted a concept, Total Survey Design (TSD), that can be operationalized and used effectively by investigators to measure the cost components of given measurement designs. An information matrix was suggested for determining survey error and cost components. This session, in particular, generated considerable discussion concerning the relative lack of adequate funds committed toward sophisticated research on various components of health survey research methods.

A major concern elicited in the TSD discussion was the widespread variability in terminology and definitions of major methodological concepts. As a step toward clarification of this issue and to promote standardization, the Conference Planning Group formulated definitions for concepts used in each session; thus, the glossary has been provided in this volume. Obtaining consensus on the definition of such concepts is central to advancing the TSD strategy and provide the data needed to improve survey work gen-

erally. It might be noted, parenthetically, that this effort follows the lead of the Social Science Research Council's Center for the Coordination of Social Indicators' recent report on the standardization of certain common personal background data typically collected in surveys.

Clearly, this conference did not solve or even highlight all of the problems of survey research. It did provide a forum in which survey methodologists and professional users of the method in health research could exchange ideas, and agree on certain problematic issues, and suggest new lines of inquiry. Because the publication lag is sufficiently long to inform the community of researchers of new developments and methodological findings, conferences such as this one are an important means of scientific communications.

Finally, much of the content of this conference has sometimes been perceived by mission agencies in the Federal Government as "basic" or "fundamental" and thus of low priority for them to encourage research or award grants for methodological work. But, as the conference forcefully demonstrates, such work has a special interest for health services. Although other agencies, such as the National Science Foundation, have a mission to sponsor investigations that are contributory to the body of knowledge concerning the substantive and statistical bases of the surveys, mission-oriented Federal agencies have an obligation also to undertake such research programs. If health policies and programs are to rest upon a sound data base, the mission-oriented agencies have an obligation to support methodological survey research. (See *Support of Basic Research by Mission Agencies* National Science Foundation, National Science Board, NSB-74-225, October 23, 1974.)

The following issues merit special consideration:

- A. There are several trade-offs to be considered in determining the length of the recall period to be used in a survey. Some of the considerations are often epidemiological in nature. If the attribute to be recorded happens to be an unusual event, one would need a rather long period of recall to obtain a robust numerator. On the other hand, as one increases the length of the recall period, telescoping and other sources of error may plague the survey.
- B. There was general agreement that certain interviews can easily last from one to two hours without serious effects to the respondent. The major

problem in this area appears to be that of *interviewer fatigue*.

- C. Advantages and disadvantages of telephone interviews were discussed. The advantages presented had a decided edge over disadvantages. The major disadvantages encompassed sample problems and frame development. Various strategies designed to reduce errors due to these sampling problems were presented. It was the consensus that, with the proper strategy, telephone interviews can be efficient. There are many advantages of being able to use telephone interviews in health surveys. Among the advantages discussed were the following:

1. One can assess interviewer performance through monitoring.
2. One can interview in areas where interviewers would hesitate to go, particularly during the evening when one may want to reach working members of the household.
3. Health professionals are more likely to participate in telephone surveys.
4. Interviewer restrictions (mobility, transportation) are substantially lessened.
5. Field costs are reduced.

- D. Trade-offs between additional training of interviewers and compensation of respondents were discussed. The participants agreed that there are greater gains in the quality of data obtained by additional training of interviewers than by compensating the respondents.

- E. The position was taken that unless the survey addresses racial issues, there is very little evidence that the race of the interviewer has any effect on the response.

- F. It was agreed that there is no simple or direct method for measuring response bias and that this is a very critical issue. One method of addressing this problem is that of assessing the internal consistency of records. To this effect, the complexity of record linkage and methods for deciding which is the valid record were discussed. The principal use of the information to be obtained should determine the priorities existing towards reducing the range of errors that have to be considered in observing the principle of Total Survey Design.

**POLICY
ISSUES
AND
COMMUNICATION
OF
RESULTS**

4

Dr. Leo G. Reeder, Chairman of the Conference, introduced Dr. Gerald Rosenthal, Dr. Bernard Greenberg, and Dr. Daniel Horvitz as a panel to lead the discussion of policy issues and communication of research results. Dr. Reeder emphasized the importance of providing the results of methodological research in health surveys to the larger body of researchers in health services research and the need for an explicit recognition of the requirements of support of methodological research.

Dr. Rosenthal, Director of the National Center for Health Services Research led the initial discussion. He indicated that the importance of methodological research in health surveys is fundamental to the viability of the research for several reasons: (1) A significant proportion of analytic work in health services research is based on survey data; (2) The quality of initial requests for research support is diminished by inadequacies in design and inappropriate specification of data pertinent to the research issues; (3) The analyses of data developed by surveys is often deficient because we cannot correct for errors in measurement; (4) The evaluation of the demonstration efforts of the National Center require baseline and follow-up surveys (to obtain the data for evaluation purposes) and the responsible persons may not be sufficiently acquainted with surveys to conduct a proper evaluation; (5) There is a need for improved health surveys in terms of the time frame of the research in order to avoid delays in the completion of the studies. Investment in the overall design of the survey could result in significant savings and improved quality of the data being obtained.

It is estimated that 55 to 65 percent of the research supported by the National Center is based upon obtaining data through survey research. Yet, in terms of the methodological research to improve the quality of the data being obtained, we are probably not making a sufficient investment. We frequently use very refined statistical techniques with all their own assumptions and limitations on survey data without worry to correcting for measurement errors. It is a pleasure to learn of the work of investigators in the methodological aspects of survey work and the advances being made in improving the quality of data.

The communication and use of research findings is, in essence, the major reason for the existence of the National Center. It is a function that is recognized as critical. The special consideration in the dissemination of methodological research should make the communication problem a little more amenable to solution than research findings destined for the world of health policy and decision-makers. The target audience is much more identified, and it is a group with which we relate actively on a repetitive basis. As a result, the strategies for communication give more promise of success. The National Center is actively interested in exploiting the value of methodological research to improve the overall quality of health services research. It is envisioned that the brokering of methodological research results will be pursued in a variety of ways including: direct technical consultation, publications, and specific targeted training. The National Center for Health Services Research is the obvious place for the research community to look for help in survey research in health services, and we ought to be available and visible. The questions raised by the conference are important ones and the results of the deliberations should themselves be widely disseminated.

Further discussion of the issues by the conference participants resulted in elaboration of both policy issues and important concerns regarding the dissemination of research results. It was noted that methodological research conducted to date has produced a body of knowledge that should be more widely applied in health services research. There are also many significant questions and issues to be addressed by methodological research.

Another way of improving the quality of data is to promote secondary analysis of the data by investigators other than those who originally collected the data. This may be especially important in large-scale national surveys and in evaluations of important social programs initiated through demonstrations or experiments. It was pointed out that we frequently seem unwilling to invest a few extra thousand dollars in secondary analyses when we have invested many more dollars and several years to acquire the information.

In the remainder of this chapter, we briefly outline the policy issues that emerged in the Conference. It is generally recognized that a substantial investment of federal funds is devoted to the collection of data through the use of the survey method. The uses to which the data are put may or may not often result in policy decisions but it is fair to state that ideally, this would be the case. As noted by Dr. Rosenthal, there is an insufficient investment in methodological research to improve the quality of the data obtained through surveys. The Conference itself will no doubt have some impact upon this situation by calling attention to the need for increased attention to methodological issues. Thus, we can summarize the policy issues as follows:

1. There is a need for increased investment in the methodological aspects of health surveys, particularly non-sampling problems.
2. Communication of research findings to the broad community of users of health data is essential. Government agencies such as the National Center for Health Services Research may act as brokers in disseminating research results of methodological studies through a variety of mechanisms.
3. There is a need for greater collaboration between federal agencies such as the NCHSR, NCHS, and NIH to advance the quality of data collected through surveys. Both the mission agencies of the federal government and the agencies concerned with so-called "basic" research make substantial investments in substantive investigations using survey research methods of data collection. Investigators use this method to make contributions to the body of knowledge regarding scores of substantive issues of direct concern to such agencies as the National Institutes of Mental Health; Child Health and Human Development; Cancer; Heart and Lung; as well as others in NIH. Similarly, the mission agencies such as NCHSR and NCHS also utilize the survey method for substantive purposes. Ultimately, the generation of this knowledge is intended to improve the nation's health and hence, the quality of the data base is of critical importance. Clearly, with such common goals, these health agencies have an obligation to support methodological studies aimed at the improvement of the state of the art.
4. Government agencies supporting studies using the survey method might give serious consideration to the development of a set of guidelines to be used by research grant applicants or contractors. Such guidelines can assist applicants to adhere to established "good practices" in survey methods and procedures. Thus, new technique, as well as standard approaches in instruments construction as they apply to health surveys might be distributed in some systematic fashion to those applying for grants and contracts involving health

survey work. Moreover, such materials might also be made available to project officers in the granting agencies (NCHSR, NCHS, NIH, NSF, and so on). Perhaps, a clearinghouse could be established similar to the NCHS Clearinghouse on Health Status Indices. Obviously, the more "experienced" investigator would benefit from such a project, as would the novice.

5. Although the development of a guide of standards would be useful to the less-experienced survey researcher, it must be recognized that survey research is a complex research methodology which is deceptively simple to the uninitiated. In terms of cost effectiveness, as a protection of respondents, and the quality of data, surveys should be undertaken only by those who are skilled in the methods or who have expert consultants readily available.
6. In order to assess research results and to advance methodological procedures, survey investigations should contain a statement of procedures used including: interviewer training and supervisory procedures, other quality control methods, and response rates. The development of standardized guidelines for monitoring and reporting the quality of the data collection would be particularly useful.
7. Explicit attention to the issue of validity should be given in all projects involving data collection. For example, validation checks should be strongly encouraged where feasible, as well as comparison of the results with data from other studies, and so on. Such checks would provide useful data for evaluating the validity of the findings as well as increase our knowledge about validity and the factors that affect it. Nowhere is research more urgently needed than in the area of confidential data and sensitive issues, because the problems of refusal to answer queries and untruthful reporting seriously affect the validity of the data. There is a special need to encourage more research on novel techniques and methods of improving validity in the area of confidential data and sensitive issues.
8. The use of computers and the linkage of records has caused problems regarding violation of confidentiality and the invasion of privacy. This has given rise to the publication of "Records, Computers and the Rights of Citizens" by the DHEW Secretary's Advisory Committee on Automated Personal Data Systems in July, 1973. Many of the recommendations in this report are sound and fundamentally just. There are, however, other recommendations which would prohibit even the matching of death certificates with birth certificates and, when implemented, essentially prohibit valuable health research that has been conducted for decades and which did not violate confidentiality. It is essential that professional organiza-

tions interested in health research and survey research express their position regarding some of the provisions of this Act to the Committee on National Statistics of the National Academy of Sciences. The American Statistical Association and the American Sociological Association have recently established, with three other professional societies, an Ad Hoc Committee on Government Statistics; the American Statistical Association also has an Ad Hoc Committee on Privacy and Confidentiality. This body should also be apprised of the possible implications of the Privacy Act of 1974.

- 6
9. The value, implications, and application of the Total Survey Design (TSD) concept should be made more readily available to the survey research community. Conferences, research and training programs, and publication of special monographs can be very instrumental not only in this instance but also with respect to other issues discussed here.
 10. Presupposing standardization of terms and definitions, an information system containing per-

minent data on the various error components and cost components associated with specific measurement designs used in sample surveys should be established for use by the survey research community. Perhaps the establishment of a national clearinghouse, as noted in Policy Recommendation No. 4 above, would be the vehicle developing, maintaining, and disseminating such information.

11. The National Center for Health Statistics, and the National Center for Health Services Research should sponsor an annual Summer Session addressing survey research in health including the review of the state of the art and analytical techniques and processes pertinent to recent approaches to survey research in health.

The discussion ended with agreement that the suggestions made were worthy of serious attention. The conference has produced numerous questions meriting research attention and the strategies for dissemination seem to hold promise. Operational plans for proceeding in both of these areas deserve immediate attention and adequate resources.

HEALTH SURVEY RESEARCH INSTRUMENTS

Seymour Sudman, Ph.D., Chairman
Ronald M. Andersen, Ph.D., Rapporteur

Introduction

Researchers collecting health data by survey methods are fortunate on the one hand, because most respondents find discussing their health and medical care interesting and important. On the other hand, medical studies increasingly require complex information, thereby increasing the burden on the respondent and making data problems more likely. In this session, the major emphasis was on how questionnaires and other survey instruments might be designed to reduce respondent burden and increase the quality of data.

In the initial planning for the session, four topics were included: 1) respondent burden; 2) length of time to complete; 3) standardized modules or standardized measures; and 4) effects of instrument complexity on different groups. In the actual discussion that took place, the first two and last two topics seemed to group themselves naturally. Thus, they are combined in this summary. Length of the interview is really one example of a burden on respondents. Also the use of standardized questions and modules assumes that these modules are understood and reacted to similarly by all different groups of interest.

This summary of the proceedings tries to capture the flavor of the remarks, but the interpretations are the responsibility of the chairman and recorder. Because we are also deeply interested in these issues, our experiences and biases may creep in. Wherever possible, however, we have tried to identify the people with their ideas so that interested readers may contact the conference participants direct for additional information or clarification.

The general framework is to present a problem in collecting health data and give some indication of the nature of that problem. Then, the current procedures used to handle these problems are described. Reference is made to research on these problems, some of which is either still in process or has not yet been widely adopted. Finally, we end with a list of unsolved problems and a future research "want list."

Respondent Burden

Respondent burden concerns the level of demand placed on the respondent necessary to answer the survey instrument questions. As the burden increases, a number of problems become more serious for the research community: 1) the population may become more alienated toward survey research, and level of cooperation for future studies may decline; 2) the reliability of the data collected may be lessened; and 3) response rate in subsequent interviews in panel studies may decline. The following discussion of respondent burden refers mainly to general population surveys, although many of the problems and attempts to solve them are also germane to surveys of physicians, other health services providers, and administrators in the health delivery field.

The major types of burden discussed in the session on health survey instruments related to issues of 1) recall period; 2) salience of information requested; 3) frequency of events the respondent is asked to report on; 4) use of proxy respondents; 5) complexity of the instrument; and 6) length of the instrument. Each of these topics will be treated below and an attempt will be made to summarize the problems associated with each and the procedures that were suggested to ameliorate the problem.

Recall Period

The major problem with asking respondents about events that happened six months to a year or more prior to the interview is that respondents are likely to forget details of the events or even the occurrence of the event itself. There are trade-offs in reducing the length of the recall period, however (Horvitz). The costs of data collection increase. Telescoping or reporting of events that actually occurred outside of the accepted recall period may also increase.

The National Center for Health Statistics (NCHS), in weighing the memory loss with a long recall period against the days of experience lost with a short recall period, arrived at a two-week reference

period for physician and dental visits. Studies conducted by the Survey Research Center at the University of Michigan have shown that reporting of hospitalization using a twelve-month recall period is less complete for hospitalizations that occurred early in the recall period (Cannell and Fowler: 1965). Consequently, NCHS is currently calculating hospitalization rates using only hospitalizations reported in the preceding six months. However, they continue to collect data on hospitalizations for an entire year. A shorter recall period, such as six months, appears to result in some telescoping (Fuchsberg). The Current Medicare Survey uses periods of one month to study hospital re-admissions. Some reporting omissions are found for short, overnight stays for diagnostic tests (Scharff).

8

The issue was raised as to whether there is fall-off in reporting even with a recall period as short as two weeks for physician visits. The answer appears to be yes (Fuchsberg). Nonetheless, NCHS judged the costs would be too great if the recall period were reduced to, say, a week for physician visits.

Another problem with a long recall period is how to account for the experiences of people who died during the recall period (Horvitz). NCHS has sponsored methodological studies to determine the bias introduced by deaths in the population during the recall period (Sirken). Using a six-month recall period, Horvitz showed that some of the losses in reporting were due to death with the rest due to memory problems. The problem is more serious if the event being studied is correlated with deaths, as is the case with hospitalizations, than if the event is not correlated, as is the case with dental visits (Sirken). In the former case, the loss is an increasing monotonic function of the recall period.

The bounded interview is one means sometimes used to decrease the effects of telescoping that result from a shorter recall period (Neter and Waksberg: 1964) (Jabine). This process involves a baseline interview and follow-up interviews which solicit the reporting of events that happened subsequent to the first interview. Sudman conducted a recent study involving an initial interview with a three-month recall period and three subsequent monthly interviews concerning physician visits and disability days. This procedure compared the bounded-interview approach and diary method for the three monthly interviews (Sudman, Wilson, and Ferber: 1974). The results show telescoping can be eliminated by reminding people of what they said in earlier periods. Omissions in recall, however, still occur.

The diary was suggested as a possible solution to the omission problem (Sudman). While major events such as hospitalizations are less likely to be forgotten, the diary approach is most helpful in aiding recall of physician and dental visits and disability days. The work of Mooney was cited as an early but still useful

comparison of bounded interview and diary approaches (Mooney: 1962) (Woolsey).

Variations in the use of the diary were discussed. In some cases, the diary is a relatively complete form which is filled out by the respondent and is used as the primary document for subsequent data processing (Survey Research Laboratory, University of Illinois). In other instances, the diary or calendar form is less formalized. It is used primarily as a memory aid for the respondent, and an interview schedule is subsequently filled out by an interviewer and used as the processing document (Johns Hopkins Medical Economics Study and the Rand Health Insurance Study). A major advantage of the more complete diary technique is that it reduces the time and cost necessary to collect the data. It might also stimulate more complete recording of events. The use of the diary as a memory aid, seemingly, would also allow the collection of more complex forms of information.

Assumptions about ability to retrieve information lost through memory decay influence the approaches used to combat such loss (Marquis). If the assumption is made that information lost through memory decay cannot be recalled, the general approach is simply not to ask the respondent for that information, but to attempt to collect it through some other source. An alternative view is that little that has been experienced cannot be recalled with the proper memory aids. In that latter case, much more effort is likely to be devoted to recall techniques.

A calendar year is sometimes used for the recall period. One assumed advantage of the calendar-year recall period has been that it corresponds to certain types of records, such as income tax, and to traditionally defined intervals, such as yearly salary. Also, advantages accrue if events are remembered as occurring before or after the start of the New Year. The Current Population Survey, however, has compared weekly and monthly reporting of income with yearly income reporting, but found little difference (U.S. Bureau of the Census, 1963) (Gerson).

Given the differences in reporting according to length of recall period, the question of why the mean estimates on utilization of health services produced by the Center for Health Administration Studies and the National Opinion Research Center (CHAS-NORC) using a one-year recall period have been similar to those produced by the National Health Survey using much shorter recall periods. Procedures in the CHAS-NORC studies that might make estimates similar to those from NCHS include: 1) incorporating into the estimates verifying information provided by hospital physicians and insurance companies; 2) encouraging respondents to consult records of income tax, doctor bills, and insurance policies; and 3) using aided recall methods, such as having people report utilization separately for each episode of illness experienced during the year and the number of times a particular doctor was seen for each episode (Andersen, et al, 1976).

Saliency of Event

In general, it has been assumed that the more salient the event the less burden placed on the respondent in reporting it. Respondents who pay fees directly for the physician visits may see these events as more salient than those who get care at no direct cost. There is some evidence that fee-for-service visits are more likely to be reported (Fuchsberg). Similarly in the Rand Health Insurance Study, completeness of reporting of doctor visits was compared to the level of reimbursement people received for their medical costs (Marquis). Those who were to receive reimbursement for the visits they reported would probably view these medical events as more salient than those who were not. Indeed, when the information was collected monthly, the findings were that persons reimbursed for their visits were more likely to report their visits than those who were not. However, there was no difference using a weekly diary. This finding suggests the need to use a short recall period for events considered less salient by the respondent, whereas a longer one can be tolerated for the more salient events.

Generalizations about saliency still must be made with caution, however. For example, some events that are salient but extremely painful may be repressed. An example given was the underreporting of infant death in social surveys when compared to vital statistics records (Horvitz).

Frequency of Event

The more frequently or common the event the more difficult it may be to recall, particularly if the respondent is required to recall specific events and details. For example, nutrition surveys asking people what they have eaten show considerable memory loss with recall periods of as short as a week (Greenberg). Diaries may prove particularly useful for this type of event.

Proxy Respondent

A proxy respondent is often very useful to summarize information and provide details for a family member not available for an interview. A proxy respondent is essential in certain instances, such as when soliciting information on young children, on seriously ill or senile persons, and on deceased family members.

For many types of questions, however, proxy reporting appears to be less valid, because events are not as salient for the proxy or, in some cases, because the proxy simply does not have the necessary information. For example, personal expenditure data tends to be less well reported by proxies than by self respondents (Sudman and Ferber: 1970).

For certain other types of information, such as reasons for hospital admission, a reasonably informed proxy serves about as well as the self respondent. For diseases that are socially undesirable or that involve

considerable threat to the individual, such as alcoholism, diabetes and cancer, proxy reporting may actually be superior (Sirken).

It is important to define the kind of proxy respondent. Having a spouse respond for the subject is very different from having any other related person in the family serve as proxy. Having the head of the family report on family expenditures and income is likely to be more helpful than using just any available adult. The results depend on the proxy's relationship to the subject, the kinds of questions asked, and the length of the recall period. All of these factors must be considered in deciding whether or not a proxy will be an acceptable respondent. It appears that, except when using "sensitive" questions or collecting information on children, senile persons, and others who would have difficulty remembering and communicating, respondent burden is probably reduced by questioning the subject himself rather than a proxy.

Complexity

The impression of many experienced researchers is that health questionnaires, on average, are complex. Respondents are often requested to provide details on a variety of subjects such as expenditures for medical care; utilization of hospitals, physicians, dental care and drugs; sequence of events in an episode of illness; perceptions of symptoms and illness conditions, comprehensive accounts of their perceptions; and evaluations of their medical care. When these types of information, particularly in combination, are solicited from the respondent, the burden may be excessive.

One strategy to reduce this burden and, at the same time, to hold costs down is to use simple, less complex questionnaires in the initial study phases on a particular topic (Dalenius). The complex details would be collected in later phases and, then, only on the issues which seem warranted by preliminary analysis. For example, if frequency of events in the preliminary phase is low, the reliability and/or validity appear unacceptable, or if adequate data is supplied by the simple form to answer the research question, then using more complex questions might not be indicated.

It was suggested that this "incremental approach" is somewhat analogous to calibration techniques used by physicists (Dalenius). Ideally, a point might be reached for those measures that require complex instruments for acceptable validity where the complex form would need to be used only on a subset of the data. The results from this subset might then be used to adjust the major data set collected by the simple method. This approach would have the double benefits of reducing collection costs and the burden on the respondents.

One cannot assume that the results achieved from calibrating one measure, such as disability, can be

applied to other measures such as doctor visits or dentist visits (Sirken). The reliability and validity of the different measures may vary considerably, and each important measure should be calibrated separately. One study of people discharged from hospitals shows little correlation in reporting error for different items including diagnosis, length of stay, and date of admission and discharge. (Cannell and Fowler: 1965).

It should also be noted that the validity of the measure may differ according to the order of the questions asked and the other kinds of information that are collected in a given instrument. For example, some differential reporting of disability days might be expected according to whether or not other questions about illness experience preceded those about disability days.

Length of Instrument

There was general agreement that personal interviews that lasted an hour or less caused no serious problems in health surveys whereas interviews lasting over two hours caused major problems with both respondents and interviewers due to fatigue (Fuchsberg, Hensler, Kulley, Losciuto, White, and Woolsey). There were mixed opinions about problems with interviews lasting from one to two hours, but consensus that problems increased when the interview length increased from one to two hours. Sirken noted that it was not sufficient to observe the average length of interview, inasmuch as interviews that averaged 45 minutes might require three hours for some respondents.

Length of the questionnaire was perceived as an even more critical problem in self-administered forms (Barnes, Boisen, Bradburn, and Dillman). Here it is the number of pages rather than the length of time to complete that influences cooperation rate. Dillman reported substantial reduction in both quality and cooperation with mail samples of general populations when the questionnaire was longer than 12 pages (Dillman, et al: 1974).

The chief method for dealing with length is by conscious efforts to cut out questions that are interesting, but not critical. Another suggested method is the use of matrix sampling in which not all respondents are asked all questions. That is, using a rotating order, each respondent is asked only a subset of all the questions of interest (Horvitz and Waksberg).

Meyers stressed the importance of being honest with respondents and telling them in advance how long the interview might last. Cannell suggested that it was the subjective, not the objective, length of the interview that was more important. If the respondent enjoyed the interview and the topics were interesting, the interview would seem shorter. Even with an interesting interview, however, extremely long questionnaires caused major difficulties (Woolsey).

Effects of Instrument Complexity on Different Groups

In the abstract, the use of standardized modules or measures seems to be highly desirable. Not only is it wasteful to re-invent the wheel, but doing so makes it difficult or impossible to compare the results of different studies if different *ad hoc* forms are used. Thus, the questionnaires used in the National Health Interview and the model Neighborhood Health Centers and the CHAS-NORC National Health Expenditures and Utilization questionnaire provide models to be followed.

Most of the discussion, however, focused on the limitations of standardized questionnaires and scales for different populations. There was general agreement that questionnaires that work well with middle-class respondents may have serious problems when used with lower education respondents. Other respondent characteristics such as age, ethnicity, social status, and acquiescence were also discussed. It appears that little is being done, however, in most surveys to handle some of the problems raised and there seems to be only limited research in progress or contemplated, although research in this area is both vitally needed and feasible.

Starting with education, it is obviously impossible to get accurate response if the respondent does not understand the words used or the meaning of the question. It is often impossible to detect this on examination of a finished questionnaire, but this language barrier is more readily evident if the interview is observed or if one listens to a tape recording. The problem becomes worse as one moves from behavioral to attitudinal questions and from the more specific to the global. Several examples were cited of mail surveys in which the quality of the data declined with successive mailings. The major reason was that less educated respondents were more likely to respond on later waves and also to have more difficulty with a self-administered questionnaire (Dillman and Sirken).

The current procedures of pretesting questionnaires prevent some of the more serious problems, but often standardized questions or modules are assumed to be satisfactory with limited or no testing. In some cases, particularly dealing with attitudinal and personality scales, these forms were standardized on college students and it is dangerous to assume that they are valid for other less educated populations.

The current procedures standardize on wording, but the content may be perceived differently by respondents with different levels of education. The converse procedure of attempting to standardize by content is extremely difficult and no reports of its use were presented at the conference, although this method is in wider use in cross-cultural studies.

Age differences are also found in standardized questions. Since both memory and energy decline for persons above middle age, questions that can be answered accurately by younger persons are much

more difficult for the aged. Although these problems are recognized in studies that focus entirely on the aged, they are generally ignored in studies of the total population. Solutions could include reducing the length of the recall period, so as to lessen the burden on the respondent's memory and energy, or obtaining data on this population from the records of medical providers.

Obtaining data from medical providers presents special problems. The chief methods for securing a high degree of cooperation and accurate response from providers are to have the study sponsored by an organization the provider respects and to make clear to the provider group the positive applications for the findings.

Ethnicity differences also affect standardized questions. Different ethnic groups use different forms of para-medical help such as curanderos by Spanish and felshers by Slavic respondents. Certain types of symptoms and folk medicine practices are also unique to some ethnic groups. Particularly, if native language questionnaires were used instead of considering the non-English-speaking respondent as a refusal, these special topics might be included in the questionnaire.

The lengthiest discussion related to respondent social and psychological characteristics and, particularly, acquiescence. As an example, the work of Carr was cited as indicating the effect of acquiescence on the Srole anomie scale (Carr: 1971). Since acquiescence is highly related to the respondent's social class, much of the data that show increased alienation in lower class black respondents may be caused or contaminated by acquiescence.

Woolsey gave examples of memory errors that were generally consistent across respondents over several events. Most respondents who were asked to recall the date of important news events tended to telescope the dates forward. A subgroup, however, consistently did the reverse. Exactly what the characteristics are of respondents who err in different directions is not understood.

The need for social approval is another variable that varies by respondent and influences response. One unexpected consequence of the need for social approval is the overreporting of some items by members of subgroups that most respondents would underreport. The smoking of marijuana, sexual experiences and participating in deviant activities may be overreported by teenagers whose peers admire such activities (Boruch).

Overreporting of socially desirable activities is reduced by using less personal methods such as self-administered forms and telephone, instead of face-to-face, interviews. Acquiescence can be reduced by making the questions more specific and less global. No method, however, completely eliminates underreporting of highly threatening events, although this problem can be alleviated by methods suggested in later sessions.

References

1. Andersen, Ronald, Joanna Lion and Odin W. Anderson
1976 *Two Decades of Health Service: Social Survey Trends in Use and Expenditures*, Cambridge, Massachusetts: Ballinger Publishing Co.
2. Cannell, Charles F. and Floyd J. Fowler
1965 "Comparison of Hospitalization Reporting in Three Survey Procedures," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 8, Washington, D. C.: U.S. Government Printing Office.
3. Carr, Leslie G.
1971 "The Srole Items and Acquiescence," *American Sociological Review*, Vol. 36, No. 2 (April), pp. 287-293.
4. Dillman, Don, Edwin Carpenter, James Christenson and Ralph Brooks
1974 "Increasing Mail Questionnaire Response, A Four State Comparison," *American Sociological Review*, Vol. 39, No. 5 (October), pp. 744-756.
5. Elinson, Jack, E. Padilla and M. E. Perkins
1967 *Public Image of Mental Health Services*, New York City: Mental Health Materials Center, Inc., for N.Y.C. Community Mental Health Board.
6. Haberman, Paul W.
1976 "Psychiatric Symptoms Among Puerto Ricans in Puerto Rico and New York City," *Ethnicity*, Vol. 3, in press.
7. Mooney, H. William
1962 *Methodology in Two California Health Surveys*, Public Health Monograph No. 70, Washington, D.C.: U.S. Government Printing Office.
8. Neter, John and Joseph Waksberg
1964 "A Study of Response Errors in Expenditures Data from Household Interviews," *Journal of the American Statistical Association*, Vol. 59, No. 305 (March), pp. 18-55.
9. Sudman, Seymour and Robert Ferber
1970 *Experiments in Obtaining Consumer Expenditures in Durable Goods by Recall Procedures*, Urbana, Illinois: Survey Research Laboratory, University of Illinois.
10. Sudman, Seymour, Wallace Wilson and Robert Ferber
1974 *The Cost-Effectiveness of Using the Diary as an Instrument for Collecting Health Data in Household Surveys*, Urbana, Illinois: Survey Research Laboratory, University of Illinois.

11. U. S. Bureau of the Census
1963 *The Current Population Survey—A Report of Methodology*, Technical Paper No. 7, Washington, D. C.: U.S. Government Printing Office.

Summary and Conclusions

The topics discussed in this section are how health survey research instruments relate to respondent burden and differential responses of subgroups in the population.

Respondent burden is the level of demand placed on the respondent in order to answer questions in the survey instrument. Burden is generally considered to be reduced by a short recall period, questions about salient and infrequently occurring events, questioning the subject personally rather than a proxy respondent, and limiting the length and complexity of the survey instrument. Various techniques were considered that might allow the researcher to collect the needed information while not unduly taxing the respondent.

There was general agreement that questionnaires that worked well for middle-class respondents may cause problems for certain minority groups. Thus, while standardized questionnaire modules are highly desirable for purposes of standardization and comparison, caution must be exercised in applying these modules to subgroups. Respondent characteristics that must be taken into account include education, age, ethnicity, social status, and tendency toward acquiescence. Ways to reduce bias caused by these factors were discussed.

Needed Research

The following needs for additional research were mentioned during the session. The listing is random, not by priority.

1. Judgments on length of recall period are already supported by substantial research, but additional work is needed on the effects of varying length of the recall period on proxy respondents vs. self-reports and on old vs. young respondents.
2. Although diaries are being used and tested for the collecting of health data, additional work is needed to see how they work with differing educational levels. Also necessary are special studies of techniques to improve respondents' ability to deal with more complex diary forms.

3. More systematic measures of the effectiveness of aided recall procedures and the use of records are needed, again especially with differing education levels.
4. Questionnaire length can be reduced by subsampling along the items of interest. The impact of this method on the accuracy and completeness of the remaining data needs to be measured. Also, procedures should be further developed for estimating total events on the basis of a sample.
5. A similar area of research is the measurement of improvement in quality due to subsampling members in large households. Again, procedures are required for making estimates of total household experiences and expenditures on the basis of a subsample.
6. Systematic measures of the effects of the time required to complete health interviews could easily be obtained from existing data by measuring quality vs. the actual length of the interview. Experimentally, one could measure the effects of time, length, and fatigue by varying the order of certain sections of a standard questionnaire.
7. There is a need for measuring the effects of various kinds and levels of incentives on cooperation and quality of data for especially complex medical studies and for special subpopulations for whom the interview is more difficult. These incentives can be fiscal or other kinds of goods and services.
8. Determining equity in health services involves comparison among subgroups of the population. Special studies for dealing with data collection from various ethnic and minority groups need to be supported.
9. Given the complexity of some health surveys, approaches to reduce respondent burden should receive special attention. Examples include bounded interviews, matrix sampling, and respondent compensation.
10. With the proliferation of health surveys and attempts to use them for health planning and prediction of utilization, reliability and validity studies of questionnaire items are becoming increasingly important. Funding should be provided for both secondary analysis of existing data and new study designs to measure reliability and validity.

INTERVIEWERS AND INTERVIEWING TECHNIQUES

Charles F. Cannell, Ph. D., Chairman
Floyd J. Fowler, Ph.D., Rapporteur

Introduction

This conference is a formal recognition of the growing interest in the quality of data in health surveys and of the need to focus research attention on improving survey methodology. Historically emphasis has been on interviewer bias, suggesting that the researcher was blameless in the quality of reporting and the interviewer was at fault.

Over time has come the realization that the data collection interview is a complex interactive process between two people. This interaction can be positive and lead to accurate and complete reporting, or it can have profoundly negative effects that bias or inhibit accurate reporting. Recently, emphasis has been focused on the components of the interaction, and research has been undertaken to use the interactive force to improve reporting.

Attention has also been given to the growing awareness that the task given the respondent is frequently too difficult for him to perform. Sometimes he does not possess the information sought or he is unwilling to report it, or he is insufficiently motivated to put forth the necessary effort to produce accurate reports. Consent to be interviewed is not a commitment candidly to answer all conceivable questions or to expend an unlimited amount of time or effort on the interviews. Respondents are volunteers who control and limit this participation. The limits of data collection are those pieces of information that almost all people can report with modest effort or can be universally induced to exert the effort required to report.

Session I focused most on these tasks. This session is primarily about the way the task is presented: the interviewer and his or her procedures.

Telephone Interviewing

One of the most important recent changes in the interaction process of the interview is the increased use of the telephone. Although in the past the use of telephone interviewing for Government sponsored health-related research has been minimal, it has become widely used in recent years.

A distinction needs to be made between three different applications of telephone interviewing:

1. For re-interviews of persons initially interviewed in person.
2. For a one-time survey, but using personal follow-up procedures to include those who either have no telephones or who cannot be reached by telephone.
3. For a one-time survey conducted solely by telephone interviews.

The Federal Government has used the telephone primarily for panel survey re-interviews. Waksberg cited the use of telephones for re-interviews in the Current Population Surveys by the Bureau of the Census. Jabine noted the use of telephones with a panel survey of Medicare recipients. The Health Interview Survey is experimenting with a panel to obtain health expenditure data by telephone.

When a sample initially interviewed in person is re-interviewed by telephone, there are no special sampling issues. Gerson reported that those without telephones or for whom the interview is particularly complicated are visited by a personal interviewer in the CPS.

The cost savings in using the telephone for interviews are substantial in some studies; other studies have shown only minimal cost savings. Moreover, the use of the telephone has been carefully evaluated for both the CPS and Medicare panels. Comparisons of response rates, the reliability of answers across waves, and the distribution of responses reveals no statistically significant difference between the success of telephone re-interviews and personal re-interviews. LoSciuto validated reports of data of birth via telephone and personal interview against birth certificates and found no differences. The reliability of reporting of items previously reported was the same whether the re-interview was in person or by telephone (Institute for Survey Research: 1975).

Telephone interviews supplemented by personal interviews have also been used for one-time interviews. For household-based samples, the basic problem is to obtain a significant number of telephone numbers. Hochstim had field listers attempt to make contact

with households, obtaining telephone numbers whenever possible without callbacks (Hochstim: 1967). In Rhode Island, City Directories (such as those published by R. L. Polk) were used to obtain telephone numbers (Thornberry and Scott: 1973). A combination of these procedures was also used with a Vermont sample (Fowler: 1973). In all three cases, a personal interview was carried out with those addressees for whom an interview could not be completed by telephone.

The percentage of all interviews completed by telephone varied from 60 to 80 percent. In all cases, significant cost savings resulted from use of the telephone. The overall response rates obtained using the combined procedures have been at least equal to those obtained by personal interviewers. Tully reported that his evaluation of some 20 surveys conducted in experimental health areas that commonly relied on this combined strategy generally had response rates in the 85 to 90 percent range.

Comparisons were carried out (Hochstim: 1967; Thornberry and Scott: 1973) between subsamples interviewed entirely in person and comparable subsamples interviewed with the combined telephone-personal strategy. Neither study found any significant differences for standard health items. There were some differences in the Hochstim study for other types of items included in the survey.

It should be noted that household-based samples permit sending an advance letter to respondents. All the studies cited above in which response rates were in the 85 to 95 percent range used advance letters. Dillman said he had evidence that response rates via telephone were significantly better if an advance letter could be sent (Dillman and Freg: 1974). Colombotos and Boisen mentioned similar experiences.

Two questions of note were raised about the above procedures. First, if the percentage of households for which telephone numbers can be obtained is sufficiently low, the cost savings associated with telephone use may be mitigated. Second, if the addressees for which personal interviews are necessary because of the absence of telephones are widely scattered, the costs of the personal interviews may be disproportionately high, again reducing the cost benefits of telephone usage.

The problem of the coverage of a sample was discussed as a greater problem for studies that rely solely on the telephone, without personal follow-up procedures. The alternatives for sampling in this case include lists; such as telephone directories, and random digit dialing. Of course, good lists, such as those of organizational members, are no problem. However, telephone directories are a very weak source from which to draw a population sample. Mobility—the problem of people listed in the directory moving and new households moving into an area that are not listed—is one major concern. People who request that

their telephone numbers be unlisted or unpublished are another omission. Although less than 10 percent of the households in the country do not have telephones, there are major differences by race and income, with the rate of non-telephone-ownership being as high as 25 to 30 percent for blacks. The latter rate appears to be primarily a function of income and probably applies to poor whites as well. In any case, there are biases nationally and in some areas if non-telephone owners are omitted from a sample. Waksberg cited the availability of not fully analyzed 1970 census data that would be helpful in describing telephone ownership.

Sudman pointed out that there are areas in which there is high stability, and almost universal telephone ownership, where these problems are minimal. The mobility problem is reduced in some places where telephone directories are issued every three months and are available from the telephone company or one of the directory publishers (such as R. L. Polk). Schuman cited a study in Cincinnati (Klecka and Tuchfarber: 1974) that showed little difference in estimates from a household-based personal interview survey and a comparable survey using random digit dialing. Random digit dialing eliminates the problems of incomplete telephone number lists, but, of course, does not provide a way to include those without telephones.

It was clear that there are circumstances in which each of the procedures suggested as a way to obtain telephone samples can be useful. However, it was also clearly agreed that the researcher must be aware of who will be omitted by the sampling procedure he uses. When in doubt about omission biases, it is probably advisable to use a procedure that is supplemented by personal interviews to include those without telephones and/or who are omitted from lists.

Next discussed were the strengths and limitations of telephone procedures for the kinds of questions that can be asked. It is clear that questions requiring visual materials have to be modified for telephone use. Income, where a card is typically presented to respondents with a large number of detailed categories, was cited as an example. Sudman reported success in achieving comparable results on the telephone by presenting a series of intervals. He felt that most of the objectives achieved by visual materials or cards could be achieved on the telephone with imaginative questionnaire design. Others (notably Dillman) felt that there were objectives that are difficult to achieve on the telephone—for example, rank ordering of lists.

Colombotos cited an example of a question that initially yielded biased information from doctors (Colombotos: 1969). When asked how many journals they read, doctors gave a lower number on the telephone than in person. But, when the question was changed to ask them to list the journals they read regularly, the difference between telephone and personal interview procedures disappeared.

Data collection that benefits from interviewer observation clearly suffers. Haberman felt that alcoholics were more accurately identified in personal interviews because the questionnaire data were supplemented by interviewer observations. When respondents are asked to check such things as medical records, checkbooks, or labels on prescription bottles (White) to verify their recall, the task is much more easily accomplished via personal interview than by telephone.

The question was raised as to whether or not there were certain subjects that should be avoided on the telephone. The consensus seemed to be that there were not. Properly done, it appeared that the range of topics likely to be covered in health-related studies have been "successfully" carried out on the telephone.

Coombs and Freedman (1964) have been using telephone procedures for follow-up to personal interviews, asking questions on pregnancies, family planning, and related topics. They report that the telephone appears to be a method as satisfactory as the personal interview for collecting such data. They stress, however, that these interviews followed a personal interview. Mooney, Pollack and Corsa (1964) report similar success on sensitive topics such as menstruation.

Sudman validated reporting of "threatening" information such as drunken driving and going into bankruptcy. Although there was considerable error in reporting both in a telephone and a personal interview, there was no marked difference in data accuracy obtained in the two procedures.

Cannell reported some data from a study in Kansas City using standard mental health scales that suggested the following pattern: for those times that were extremely threatening or only mildly threatening, there were no differences between telephone and personal interview responses. For items that appeared to be moderately threatening, however, respondents seemed more likely to describe themselves in a positive way on the telephone than in a personal interview (Henson, Roth and Cannell: 1974). Bradburn cited work using his "happiness" item that showed no difference; but Cannell thought those items were probably at the less threatening end of the continuum of the Kansas City questionnaire. He concluded that there was need for some caution and further research as we apply telephone interviewing to other substantive areas.

Overall, with proper introduction by advance letter or with an initial personal contact, there is no obvious restriction on the subjects about which interviews can be conducted on the telephone—either through concern about accuracy or response rates. However, the data cited by Cannell and the lack of comparative research for one-time surveys on topics other than fairly basic health measures suggest a need for some caution and some further research with respect to the application of telephone to the more sensitive topics.

In a similar vein, some people expressed concern about how long a telephone interview could be. Some organizations try to limit telephone interview length to 20 minutes or half an hour. However, other researchers, such as Sudman, reported no difficulty in having interviews that last an hour or more. There seemed to be no basis for saying that the restrictions on length of telephone interviews were any greater than those on personal interviews; although it is clear that most users tend to keep telephone interviews shorter than personal interviews; moreover, there was little experience with telephone interviews that lasted over an hour except with special populations.

The suggestion that telephone interview schedules, and perhaps the training of interviewers, should be different from those for personal interview schedules was discussed in some detail. There was a dearth of hard evidence, but there was a considerable amount of feeling that some compensation was needed for the absence of visual cues in the interaction between interviewer and respondent. Dillman reported that he has interviewers frequently summarize answers and verify them with respondents. Bradburn discussed the theoretical and laboratory work of Ingve that there are "back channel" (feedback) noises people make that keep conversations going that should be standardized in telephone interviews. This seems to be an area in which some comparative research is needed. At the moment, there do not seem to be any clear guidelines for different procedures to account for the special type of interaction on the telephone, although the need for such procedures seemed likely to the participants.

Finally, several advantages of using the telephone in the administration of surveys were cited. Following respondents who relocate through panel studies and coverage of respondents widely dispersed geographically is greatly facilitated by telephone. Persons difficult to find at home can sometimes be more readily reached by telephone. Persons residing in places where interviewers are reluctant to go such as urban high-crime neighborhoods or high-rise apartments with extensive security systems, may be reached more successfully in a telephone survey. Busy professionals or elite respondents may be contacted more successfully and interviews completed at their convenience via telephone, as Colombotos has shown.

It was also suggested that telephone interviewing might reduce the between-interviewer differences. Visual cues that might produce bias are not factors. However, Colombotos finds as much between interviewer variance on the telephone as in person in his surveys of physicians. Apparently, that potential has not yet been realized. Nevertheless, the potential for close supervision of telephone interviewers—monitoring their actual interaction with respondents (making certain not to violate Federal laws regarding tapping) and the potential to select and use interviewers without restrictions on age, appearance, car ownership and mobility—should permit higher standards for inter-

viewer performance to be achieved, and increase standardization of techniques. There is also the potential for rapid entry of computer processing; data telephone interviewers with a terminal can enter answers during the interview.

In summary, there appears to be consensus that for many purposes for which personal interviews have been used, telephone techniques can produce data of equal or higher quality, often at lower cost. The possibility of excluding significant segments of a study population by exclusive reliance on the telephone needs careful attention in any given study design. There seem, however, to be few if any bases for saying *a priori* that the telephone is a less satisfactory data collection modality than the personal interview. The telephone interview has the potential for solving some problems that have plagued personal interview procedures. There is need for further research as its applications are extended, but there are few obvious limits on its utility at the moment.

Compensation

The question of paying respondents or compensating them in some other way for their participation in a survey has been debated for years. In this conference, the focus was only on the rate of cooperation or the quality of response. Issues such as the appropriateness of paying low-income respondents for reasons of justice may be very important in certain contexts but are not considered here.

One could argue that compensation could increase respondent commitment to a task and relieve feelings of exploitation; or that it could detract from reporting accuracy by making people feel that they are being bribed. Sudman has some data indicating that diary keeping may be somewhat more complete when respondents are compensated; but, in general, there is no conclusive evidence on this point.

There is, in contrast, a good deal of data on the value of compensation to increase cooperation in data collection efforts. It appears that when respondents are being asked to accept a moderate task, within the range of the standard one-time interview of about an hour, compensation does not have a significant effect on response rate. However, when the positive forces on respondents to cooperate are fairly low—as in a mail survey—or when a great deal is being asked of respondents, compensation appears to be helpful. Panel studies using diary techniques benefit from compensation, particularly in the third and fourth waves (Sudman and Ferber). Success with payment to induce a sample of young adults to take a series of tests that took several hours was also reported (Chromy and Horvitz: 1974).

Compensation need not always be monetary. Feedback on panel results has proven helpful to maintain panel cooperation (Sudman), and Greenberg pointed out that providing diagnostic information results from

health examination surveys is an incentive for cooperation.

Hagerman reports that his interviewers liked being able to compensate respondents in a study of alcoholism, although it was not clear it affected the response rate. The role of the interviewer was also cited by LoSciuto: when interviewers had small gifts from which respondents could choose, slightly higher response rates were obtained.

Knowing at what level to compensate respondents was discussed briefly. Horvitz cited the need for empirical testing to decide the amount of compensation to offer for a given task. This issue was reinforced by findings in psychological laboratory experiments, in which paying too much actually reduced task performance while moderate compensation increased the performance.

In general, there was little enthusiasm for compensating respondents unless unusual demands were made of them, such as repeated interviews, lengthy interviews, difficult tasks, etc. Perhaps the sense of the conference was best reflected in a Bureau of the Census experiment, reported by Gerson. When trying out a diary technique, Census paid matched groups different amounts of money and compared the response rates with an unpaid control group. The response rates were unsatisfactory for all groups. They then proceeded to work harder on training their interviewers, who without compensating respondents subsequently obtained much higher response rates than any of the preceding groups. The moral may be that there are many ways to enlist respondent cooperation. Although it may seem only fair to compensate those of whom a great deal is to be asked, for more modest tasks there are other better understood, more reliable, and probably more effective ways to enlist cooperation; and most researchers would probably do best using those.

Response Sets

The group spent some time discussing “response set.” The ideas that emerge reflected a breadth of ideas and concepts ranging from a description of symptoms to more causal hypotheses. In contrast to some of the other topics discussed at the session, it is clear that the issues are complicated, and they interact in complex ways. The discussion was illuminating even though many issues were raised and no firm conclusions were reached. Clearly, this an area on which much more research attention needs to be focused.

Marquis’ statement of the problem is that it is useful to think of a response set such as acquiescence or conformity as a response that is generated by some stimulus other than the question content itself.

Such extraneous stimuli include a wide variety of factors: the form or wording of the question; the difficulty of producing an adequate response; the perception of, or the expectations of, the interviewer

(especially if the interviewer is seen as being of higher status); the level of effort the respondent is willing to exert; and the social desirability of alternatives offered to respondents.

These can be subsumed under two general headings. In one case, it appears that there are forces that dominate a question, leading many respondents to respond in a way that does not reflect their true response. In the other case, the respondent either has no ready response or the task of generating a valid response requires greater effort than he is willing to put forth. In the latter situation, his response may be based on some extraneous cue from the question, from the interviewer, or from some other aspect of the situation.

The discussion differentiated the source of the cue from the form of the response that is generated. In some cases the source and the response mode are closely linked, and in others they are quite independent. For example, if the interviewer is perceived as having more education than the respondent, the respondent may be reluctant to report that he seldom reads books. If an abstract question is not understood, but only a yes or no response is required, the respondent may merely pick one answer rather than admit his lack of understanding.

The following is a list of the various factors discussed at the conference that have been found to influence the kinds of responses that are obtained:

Relative status of interviewer and respondent

Indicators of differences in status may influence the respondent and the interviewer. The status difference may be mediated by such factors as ingratiation, resistance or conformity. Indicators of status may include education, social class, or income.

Some research (Fowler: 1965; Cannell, Fowler and Marquis: 1968) has found clear differences in interviewer behavior toward respondents with different levels of education, particularly in their interpersonal interaction and feelings. Respondent perception of the interview situation and behavior in the interview also differed by education. For example, when the respondent was of lower education level than the interviewer, respondent behavior was more ingratiating and submissive. In experimental work, when the interviewer had higher education than the respondent, feedback resulted in better respondent performance. However, when the respondent was of higher education, performance did not improve, and in some cases, worsened (Marquis, Cannell, Laurent: 1972). One hypothesis to explain these results was that the relative status made feedback appropriate and welcome by less educated respondents and inappropriate and resented by higher educated respondents.

There were various comments on the effects of these status differences on the quality of data. Weiss

found middle-class interviewers obtained better reporting from people on welfare than indigenous interviewers (Weiss: 1968). Bradburn cited Hyman results that showed that similarity of interviewer and respondent led to unwarranted assumptions by the interviewer that he understood the view expressed by respondents (Hyman 1954). An anecdote was told about past ingratiating behavior of low-income Blacks in the South with respect to white interviewers—the status difference produced apparent cooperation but little actual cooperation in obtaining good data.

Other demographic differences

Interviewer age and sex have been thought to be important to the quality of responses. Age or sex matching of respondent and interviewer has been advocated, especially for topics in which these may be expected to influence the response. Bradburn reported a consistent finding that young interviewers (most often college students) were especially poor interviewers. He attributes this, however, not to age but inadequate training and experience. Other comments also suggest that the major variable involved is the adequacy of the training the interviewer receives rather than the age characteristic. LoSciuto, Colombotos, Meyers and others report that when training was adequate, sex or age of interviewer showed no effect, even in studies in which differences might be expected.

One of the most extensive discussions dealt with interviewers' race. The results of that discussion are summarized in a subsequent section.

The form of the question

There was some discussion centering around the idea that some question forms, such as agree-disagree, yes-no or unbalanced format may be distinctively likely to lead to particular response sets. These apply particularly to attitude scales commonly used in sociological research. It was suggested that acquiescence scales can be built into questionnaires to identify people most likely to respond with certain sets. It was also noted that these procedures are difficult to apply; moreover, the issue has somewhat limited relevance to standard health surveys.

Carr reported on his study of acquiescence response with the Srole Anomie Scale (Carr: 1971). There was considerable discussion of acquiescence and other response sets, their characteristics and causes. Ware suggested that what appears to be acquiescence is often a reflection of the form in which the question is asked, especially those in which the alternatives given the respondent are in fact not alternatives. He cited other forms of scale items that may give rise to acquiescent-appearing responses but which in fact reflect other factors. It was generally agreed that more needs to be learned about response sets—what they are and what factors underlie them.

The subject of the question

There are well-established effects of the affective component of questions. Such concepts as social desirability and threat to self-image were mentioned.

Some research (Cannell, Fisher and Bakker: 1965; and Cannell and Fowler: 1965) found that the more threatening the reason for hospitalization the less likely such hospitalization would be reported. A similar pattern was found in the likelihood that a chronic condition would be reported (Madow: 1967). Sudman reported validated data that showed under-reporting of undesirable events, such as being arrested for drunken driving, and over-reporting of desirable characteristics, such as having a library card.

Monteiro said that differences in male-female reports of disability days because of illness may reflect a greater reluctance of men to admit they were ill. Greenberg said that Sidney Cobb's data indicate that men who are unemployed are more likely to attribute disability days to illness rather than to being laid off. Both were referring to unpublished data. Apparently, it is more socially desirable to report illness than unemployment. In these and in previous examples, the result of social desirability forces is likely to be biased data.

Cues from the interviewer

The interviewer may give cues that affect respondent behavior. The types of cues that may be given are virtually limitless—involving verbal and non-verbal cues.

Marquis suggested that the number and types of probes an interviewer uses, and his pace, can communicate to respondents certain expectations for respondent behavior. It has been found (Cannell, Fowler and Marquis: 1968) as established in independent interviews with interviewers that their goals of accuracy and speed are communicated to respondents. Fuchsberg noted that the speed of the interview is one of the most important indicators to the participants that there is a rush to complete the interview.

Reaction to the difficulty of the task or question

When the respondent is given a difficult question or one that requires a great deal of effort to retrieve the information requested, he is liable to take a short cut that produces response error. The researcher is often unaware that this has occurred, because an acceptable, codable answer was obtained. Some data on task difficulty (Cannell, Fisher and Bakker: 1965 and Cannell and Fowler: 1965) show that hospital stays in the distant past and those that had low impact (because they were short or did not involve surgery) were less likely than others to be reported.

Personality or cultural patterns

It was suggested that some cultural patterns generate response tendencies. Reeder, for example, suggested that Mexican-American respondents in Los Angeles may have a yea-saying tendency. Others have suggested that certain ethnic groups tend to exaggerate or minimize their health problems.

Racial matching of interviewers and respondents

The issue of whether more valid data are obtained in interviews in which the interviewer and respondent are of the same race has been the subject of much discussion over the past several years. This subject is of sufficient importance in survey research that we have separated the discussion from other related topics.

There was general agreement among participants that whether the race of the interviewer and respondent was the same or different had no discernible effect on the data reported *except* where the interview focused on racial topics. This conclusion was stressed especially by Bradburn, LoSciuto, and Schuman who reported their own and other research to substantiate this position (Schuman and Converse: 1971). The evidence also generally supported the conclusion that racial matching made no difference in response rates (LoSciuto). However, Schuman found that black interviewers did obtain somewhat lower response rates in some white neighborhoods. It was suggested that studies showing black/white differences might well reflect less thorough training of the black interviewers.

From this discussion, the participants generally agreed that, while there may be some minor effects from racial differences between interviewer and respondent, much of the early concern over the issue has been dissipated. Black interviewers are generally as successful, both in response rates and in the data obtained, with white respondents as with black. Similarly, white interviewers can interview either race respondents.

When the topic of the interview is racially related, however, significant effects are reported (Schuman and Converse: 1971). Interestingly enough, it appears that matching the race of the participants is not always best. Greenberg reports the feeling of interviewers in a study covering sensitive topics of family planning. The interviewers felt that black interviewers obtained less accurate reports from black respondents because of their concern their responses might be spread to other members of the black community by the black interviewers. Weiss provides quantitative findings consistent with these impressions (Weiss: 1968).

The conference hoped that this discussion may help lay to rest the recurring topic of race matching of interviewers and interviewees. That race may reflect other characteristics that will affect interviewer results (education, sex, age, socioeconomic status, and

most important, training and experience) should not be overlooked.

Conclusion

There are other points that could have been raised that can or do affect the responses obtained. Marquis pointed out that we should not only consider interviewer status or respect as a source of bias, but also the possible positive uses of interviewer influences. This was an appropriate follow-up to his earlier characterization of response bias. In fact, most of the factors discussed in the section influence responses. In some cases we know how to minimize their effects by manipulating question wording or objectives. However, in many cases the possibility that an extraneous factor may influence a response cannot be eliminated. Rather than setting out on the almost endless task to eliminate all possible biasing factors, the solution would seem to be to strengthen and structure those forces that lead to the desired outcome, namely, accurately answering the questions asked. The goal should probably be to make giving accurate answers the dominate force in the interview. Some efforts to accomplish that goal are discussed in the final section of this session.

Methods for obtaining respondent cooperation

The topic of pay to respondents as a means to motivate good interview performance was discussed earlier. One of the criticisms of payment was that it was unclear what meaning the payment would have to respondents. No doubt the meaning would depend on the respondent's circumstances and the way the compensation is presented. A payment for time spent may have a positive effect whereas a perception of it as a "gift" may appear as a reward for giving the interviewer the answers he wants.

The researcher's goals, of course, are to induce the respondent to accept the task of being a good respondent—not simply going through the motions of giving some answers, but to attempt insofar as possible to give the answers that meet the researcher's objectives. Throughout the conference, and in this session in particular, the issue of how to achieve this was discussed. The desire to contribute to research, the relationship to the interviewer, and payment are examples of forces that operate on respondents in ways that may help to achieve this goal. However, the meaning is unstructured and the ways respondents react to them can vary widely.

Overall, though, we know that survey data are the result of a complex interaction among interviewers, respondents, question content, and interview procedures. We have to take respondents as they occur in the population, and question objectives may be difficult to change, although they can often be modified to reflect our understanding of what respondents are in fact willing and able to do. The main things we can change, however, are the interviewers' be-

haviors and the procedures they are asked to work with.

Cannell presented some findings of a research program designed to harness and structure the potential forces that can be brought to bear on respondents in a way that is much more directly related to achieving the research goals. The studies attempt to accomplish this by focusing on the information being reported and on the reporting process rather than on any personal affective response to the respondent.

Rapport may be (as Hyman noted years ago) not only unnecessary, but deleterious to the interviewer role. A more professional, task-oriented interaction may be better and more productive of better reporting (also see Dohrenwend et al.: 1968).

Woolsey and Gerson commented that Census interviewers are characterized by a very businesslike approach. Gerson believes this also has the advantage of reducing interview time, and thus, saving money.

Early studies examining the interaction patterns showed that much interviewer behavior was spontaneous; that is, was not part of his training or instructions but reflected particular interpersonal needs of the moment or the interviewer's interaction with others. These idiosyncratic behaviors were individual in nature, varied from interviewer to interviewer, and were outside the control of the researcher. The feedback techniques used were especially individual in nature and appeared to be a major potential source of interviewer variation in the completeness and accuracy of obtained responses.

Experiments were conducted and designed to control much of the interaction, especially the important feedback techniques. This was done by making the questions more self-contained and by specifying feedback statements for the interviewer to use. These statements focused explicitly on the process and content of the respondent activities of answering questions rather than on the rapport or interpersonal affective aspects of the interview.

Essentially three different kinds of strategies have been used. One of these involves giving respondents detailed instructions about what they are supposed to do—not just on individual questions, but for the interview as a whole. Instructions stressed the importance of accuracy, of reporting even minor events, and of encouraging the respondent to work hard to recall distant or insignificant events. Such instructions appear to improve reporting significantly.

A second strategy is designed to clarify the respondent's commitment to accuracy by making it explicit. When a respondent agrees to an interview, typically he or she probably does not know details of what has been agreed to. Some respondents are probably agreeing to go through the motions, while others are agreeing to provide accurate information. Interviewers may communicate differently about what they perceive the agreement to be. No doubt, some interviewers communicate the fact that all they want is

some answers to fill some blanks in the questionnaire.

In an experimental study of the effects of commitment, a written form was presented to respondents for their signature. The form stated that by signing the document, respondents were agreeing to give complete and accurate information to the best of their ability. They were told that if they could not make this commitment, they should not continue with the interview. The interviewer also signed the same form, committing himself in writing to the confidentiality of the data.

This procedure accomplishes at least three goals. It has the potential to eliminate the respondent that is a chronic problem for survey researchers: the one that appears to accept the task but in fact does not. Second, it clarifies and standardizes for respondents what they are agreeing to do. Third, by virtue of signing, the parties are in essence making a commitment, which becomes an additional force on them to honor it.

The results were that only a small percentage of respondents refused to sign the form, and the quality of reporting improved significantly.

Sudman reported a similar procedure used in a pretest of collection of income data. In this case, he used a lengthy introduction, that noted that the subject of income is a sensitive one; that some people have fears about IRS or possible misuse of the information. It assured respondents that answers will not be misused and advised respondents that they did have to answer questions. However, if they choose to answer the question, they were requested to do so accurately. Essentially, it tells the respondent that the researchers would rather have no answer at all than a poor one; and that by answering the question, a commitment is being made to do it to the best of the respondent's ability. Sudman reports that the rate of refusal to answer is lower with this instruction than without it, but that the quality of information appears to be considerably better than with the standard approach. He is presently conducting a field experiment that will provide data on this way of asking about income.

A third strategy studied by Cannell is the use of positive and negative feedback by the interviewer, depending on the respondent's answer.

In effect, the procedure rewards "good" performance and reacts negatively to inadequate performance. For example, a respondent was asked whether she had been sick or not feeling well at any time during the past two weeks. If illness was reported, this positive feedback was given—"That's the kind of information we need for this study." If the respondent gave a rapid "no" response, the feedback used was something like: "You answered quickly. Sometimes it's hard to remember these things. If you think about it again you may remember something." This strategy is another attempt to clarify what is expected of the respondent. Moreover, it makes it clear that the interviewer ex-

pects a certain level of quality and is not a passive person who will accept any level of performance. We know that interviewers communicate their expectations (Fowler: 1965); but they communicate different ones, in an unstandardized way. One important goal of the strategy is to standardize the expectations interviewers communicate and the way they communicate them. The result of the procedure will increase significantly the reporting of events or behaviors known to be commonly underreported.

An analogous experiment reported by Sudman also used directive probes. For those who denied ever using marijuana, for example, the probe was, "Not even once?"

These studies have not been fully analyzed and the full potential of these strategies has not yet been evaluated. However, they mark an important and promising avenue to improve the quality of data collection. Interviewing has its roots in the nondirective clinical interview. Being nondirective in terms of the content of the answers is, of course, essential; but being nondirective with respect to the quality of the answers is not. Interviewer variation has consistently been found to be very large. Examination of the interviewer-respondent interaction shows that the majority of the discussion between interviewer and respondent is not standardized and it deals with something other than the question and answer process.

One important aspect of these experimental efforts is to minimize interviewer behavior that is not standardized. By structuring transitions, instructions, and responses to answers, some of the sources of between-interviewer variation can be reduced. Moreover, the structuring is in the direction of clarifying respondent expectations and setting some standards for them.

In the conference discussion, it was pointed out that these procedures may lead to overreporting. Clearly, reinforcement can most easily be used when the direction of bias or error is known in advance. It is more difficult to apply when there is not a clear criterion for interviewers to use. However, the general instructions and strategies for enlisting commitment to accuracy can be applied to all kinds of reporting.

The tests of these procedures thus far have been limited to increasing the reporting of events or conditions commonly underreported. The criterion has been that more reporting is better reporting. For the items used, this assumption is well based in validity studies. However, these strategies require more development and testing. In particular, validation is needed with better criteria for accuracy and with a wider range of health items.

The training and procedures of interviewers have remained essentially unchanged for thirty years. Looking at the reporting problems we have at the moment argues that we sorely need to do a better job of eliciting high-quality performance from respondents. Work to date suggests that we can do better through structuring interviewer behavior and setting clear goals for

respondents. At this time, this is perhaps the most promising area of research to improve survey data collection methods.

References

1. Cannell, Charles F., Gordon Fisher and Thomas Bakker
1965 "Reporting of Hospitalization in the Health Interview Survey," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 6, Washington, D.C.: U.S. Government Printing Office.
2. Cannell, Charles F. and Floyd J. Fowler
1965 "Comparison of Hospitalization Reporting in Three Survey Procedures," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 8, Washington, D.C.: U.S. Government Printing Office.
3. Cannell, Charles F., Floyd J. Fowler and Kent H. Marquis
1968 "The Influence of Interviewer and Respondent Psychological and Behavioral Variables on the Reporting in Household Interviews," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 26, Washington, D.C.: U.S. Government Printing Office.
4. Carr, Leslie G.
1971 "The Role Items and Acquiescence," *American Sociological Review*, Vol. 36, No. 2 (April), pp. 287-293.
5. Chromy, J. and Daniel Horvitz
1974 "The Use of Monetary Incentives in National Assessment Household Surveys," *Proceedings of the American Statistical Association, Social Statistics Section*.
6. Colombotos, John
1969 "Personal Versus Telephone Interviews: Effect on Responses," *Public Health Reports*, Vol. 84, No. 9 (September), pp. 773-781.
7. Coombs, Lolagene and Ronald Freedman
1964 "Use of Telephone Interviews in a Longitudinal Fertility Study," *The Public Opinion Quarterly*, Vol. 28, No. 1 (Spring), pp. 112-117.
8. Dillman, Don A., Edwin Carpenter, James Christenson and Ralph Brooks
1974 "Increasing Mail Questionnaire Response: A Four State Comparison," *American Sociological Review*, Vol. 39, No. 5 (October), pp. 744-756.
9. Dillman, D. A., Jean Gallegos and James H. Frey
"Reducing Refusal Rates to Telephone Interviews," Forthcoming in *The Public Opinion Quarterly*.
10. Dohrenwend, Barbara S., John Colombotos and Bruce P. Dohrenwend
1968 "Social Distance and Interviewer Effects," *The Public Opinion Quarterly*, Vol. 32, No. 3 (Fall), pp. 410-422.
11. Fowler, Floyd J.
1965 "Education, Interaction and Interview Performance," Unpublished Doctoral Dissertation, the University of Michigan.
12. Fowler, Floyd J.
1973 "Final Report: Vermont Health Utilization Survey," Survey Research Program, Boston, Massachusetts, May.
13. Haberman, Paul W.
1963 "Differences Between Families Admitting and Denying an Existing Drinking Problem," *Journal of Health and Human Behavior*, Vol. 4, pp. 141-145.
14. Haberman, Paul W.
1973 *Study on Sanitary Criteria for Salt*
1974 *Water Bathing Beaches: Pretest of Illness Inquiry System and Site Selection, Part 1, 1972, November 1973, Part 2, 1973, January 1974, Office of Research and Development, U.S. Environmental Protection Agency.*
15. Henson, Ramon, Aleda Roth and Charles F. Cannell
1974 "Personal Vs. Telephone Interviews and the Effects of Telephone Reinterviews on Reporting of Psychiatric Symptomatology," Survey Research Center, Institute for Social Research, The University of Michigan, October—Also Submitted to the *Journal of Applied Psychiatry*.
16. Hochstim, Joseph R.
1967 "A Critical Comparison of Three Strategies of Collecting Data from Households," *Journal of the American Statistical Association*, Vol. 62, No. 319 (September), pp. 976-989.
17. Hyman, Herbert H., et al
1954 *Interviewing in Social Research*, Chicago: University of Chicago Press.
18. Institute for Survey Research
1975 "Final Report for a Study of Supplementary Security Income Redetermination," Submitted to the Social Security Administration by the Institute for Survey Research, Temple University.
19. Klecka, W. R. and A. Tuchfarber, Jr.
1974 "The Efficacy, Biases, and Problems of Random Digit Dialing," Prepared for

delivery at the annual conference of the American Association for Public Opinion Research, Bolton Landing, N.Y., May 31-June 2, 1974.

20. Madow, William C.

1963 "Interview Data on Chronic Conditions Compared with Information Derived from Medical Records," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 23, Washington, D.C.: U.S. Government Printing Office.

21. Marquis, Kent H., Charles F. Cannell, and Andre Laurent

1972 "Reporting of Health Events in Household Interviews: Effects of Reinforcement, Question Length, and Reinterviews," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 45, Washington, D.C.: U.S. Government Printing Office.

22. Mooney, H. William, Beatrice R. Pollack and Leslie Corsa

1964 "Uses of Telephone Interviewing to Study Human Reproduction," *Public Health Reports*, Vol. 83, No. 12 (September), pp. 1049-1060.

23. Schuman, Howard and Jean M. Converse

1971 "Effects of Black and White Interviewers on Black Responses in 1968," *The Public Opinion Quarterly*, Vol. 35, No. 1 (Spring), pp. 44-68.

24. Thornberry, Owen and H. D. Scott

1973 "Methodology of a Health Interview Survey for a Population of One Million," Rhode Island Health Services Research, Inc., Paper Presented at the 101st Annual Meeting of the American Public Health Association, San Francisco, California, November.

25. Weiss, Carol H.

1968 "Validity of Welfare Mothers' Interview Responses," *The Public Opinion Quarterly*, Vol. XXXII, No. 4 (Winter), pp. 622-633.

Summary and Conclusions

This session of the conference focused on the interaction between the interviewer and respondent, and several positive and negative results from these interactions were discussed. That the task given a respondent is often too difficult for him to perform adequately and willingly was given attention during this session as well as in the one that preceded it.

Telephone interviewing received special attention since the method of data collection is being utilized increasingly because of its lower costs compared with

personal interviews. The danger that some segments of the population will be excluded from the sample if the telephone is used as the exclusive method must be given careful consideration.

There is considerable evidence that the quality of the data from telephone interviews is comparable to personal interviews, although more evidence on this issue needs to be obtained.

Compensating respondents evoked considerable discussion, the general conclusion being that there was no good evidence that financial or other rewards improved the response rates, and there was some concern that such techniques might, in fact, introduce biasing forces. It was generally agreed that unless special demands were made on the respondent, in terms of time or work load (such as keeping extensive diaries), compensation should not be made.

Some of the complex issues of response sets were discussed. Since this is both complex and diverse in nature, no firm conclusions were reached. Rather, the nature of the discussion and the lack of adequate research data suggest that the area deserved much more research attention. Diverse topics were considered, ranging from the effects of relative status between interviewer and respondent, the form of the question, the nature of the subject of the question, and personality and cultural response patterns. On racial matching of interviewers and respondents, the group felt that the evidence was sufficiently clear to conclude that when the topic of the survey or individual questions were racially related, racial matching was important. For non-racial issues, racial matching was unimportant and not a potential source of bias.

The session concluded with descriptions of current research designed to improve the quality of research data by achieving better respondent performance. Several studies were described in which interviewer techniques, question designs, and other procedures were used to improve reporting. Several such techniques had significant effects on respondent performance. The general conclusion was that research on respondent performance and techniques for improving it was particularly promising and should be encouraged.

Needed Research

There are five areas of research that emerge from, or are directly related to, the discussion in this section.

1. The increased use of the telephone for Government-sponsored research is clearly a new and important development. There are three clear areas of uncertainty regarding the use of telephone interviews:

a) How to use telephones as the basic data collection modality for a one-time survey without serious concerns about the quality of the sample. While random digit dialing and household based samples using combinations of telephone and personal interviews are promising and ap-

pear to be successful, there is a lack of clear guidelines for when and how telephones can be used. The issues include when—for which populations, under what conditions—replying on the telephone alone is biasing, what kinds of supplemental procedures can be used to avoid these biases, and the circumstances under which combinations of telephone and other procedures are not cost effective.

- b) There is a need for research on what special interviewer techniques or behaviors are required for telephone interviewing.
 - c) There needs to be more extensive testing of the effect of telephone procedures on the quality of data. There have been no conclusive experiments in which the quality of telephone data has been independently validated. Although aggregate data on the telephone approximate personal interview data for standard health interview items, the application of telephone procedures to a broad range of subject matters—particularly those whose social desirability bias is an issue—has not been fully evaluated.
2. The role of interviewer behavior on interview results is a critical area for research. We know there is a great variation by interviewer without fully understanding the reasons or what to do about it. There is some promising preliminary work that leads one to suspect that interviewer variation can be decreased and interviewers can be used to greatly improve respondent performance. However, this work needs to be much more developed. We are not close to being able to prescribe specific interviewer procedures.
 3. Although there is uncertainty about what the interviewers conduct should be, there are many procedures that are well documented as being essential to produce reliable data. However, there is great variability in the training and monitoring procedures used to implement these procedures.

The importance and content of training, the efficacy of different amounts or kinds of training, and efficacy of different quality control strategies have not been evaluated adequately. Such evaluation is critical as we attempt to set guidelines for better methodology.

4. Related to Number 3 above is the problem of adequately reporting on the quality of data collection. Researchers report sampling error estimates as if they were the only or primary source of error in survey data. They report response rates. However, the way interviewers do their job has been shown to account for 50 percent of variance in the data's accuracy. It is highly desirable to develop standard indices for the quality of data collection. This means determining what the relationship is between various indices of interviewer behavior and the quality of data that results.
5. Of all the demographic issues raised, perhaps the most pervasive is the impact of status differences between interviewer and respondent. This stems, in large part, from the fact that most established interviewing staffs are well educated—at least high school graduates, and more commonly college educated. The consequences of the fact that in most survey projects a substantial number of interviews are conducted by interviewers who are obviously of higher status than their respondents is not well documented. Furthermore, because that is likely to remain the case, we need research on how best to deal with this situation; we need to know in what circumstances is status difference between interviewer and respondent important; and what interviewers can do to counteract the negative consequences.
6. Additional research on "Respondent Compensation" is needed, especially on varying levels and types of compensation. Current evidence on this subject is mixed, i.e., compensation payments neither improve the response rate and the quality of the data, nor do they have a deleterious effect.

VALIDITY PROBLEMS

Bernard G. Greenberg, Ph.D., Chairman
Monroe G. Sirken, Ph.D., Rapporteur

24 Introduction

At the outset, the Chairman explained the origin of this particular session and summarized the reasons for dividing the time as shown in the outline of the Conference Agenda. The idea started out as a session that would be devoted to the problems and methods of collecting data on highly sensitive and/or confidential questions. The purpose was to promote a discussion of randomized response and other survey methods that are useful in minimizing mean square error, where the latter term is defined as the sum of the square of the bias plus the sampling variance.

The bias refers to the error in estimating the true mean of a distribution of a continuous variable or the proportion of some attribute measured on a dichotomous or multichotomous scale. The reason for the bias involved here is the lack of cooperation that results in nonresponse or untruthful reporting by those who appear to participate fully by providing a reasonable, but not completely correct, response. It is worth noting that the bias not only affects the measurement of central tendency, but that frequently the impact is even greater upon the magnitude of the mean square error. There is no simple or direct way to measure bias; otherwise, one could adjust estimates of mean and standard deviation by correcting for such bias. The goal is to eliminate or reduce to a minimum the effect of this bias in reporting. Thus, the topic of sensitive questions, listed as the third item in the Agenda, was basically the starting point.

As the conference planners considered the problem of bias, it became evident that more than simply sensitive questions were involved since bias may result from other sources as well. Thus, the scope of the session was broadened to consider validity in general. In the connection, the usefulness of checks for accuracy of response by utilizing simultaneous record checks when data from several sources are collated and compared. Data should not be accepted at face value and supplemental sources of information should be used whenever records can be checked in this manner. Further consideration of record checks led to the realization that the process would sometimes involve the problem of matching and linkage of records. Cur-

rently, these are most often carried out by computers using material from data banks and other large data registers. The problems of matching and techniques of linkage comprise a whole subject area requiring full attention in itself, but interest was primarily in the uses of linkage as a means to verify data. This resulted, therefore, in the second item listed on the agenda for this session.

In any discussion of record linkage, the question arises as to whether such matching may be a violation of the confidentiality of the data. The use of a record for a purpose not intended when the information was originally collected raises invasion of privacy questions. Thus, before realizing it, the problem of obtaining answers to sensitive and potentially stigmatizing questions had broadened into the four areas in the agenda, viz., confidentiality and the invasion of privacy problem (IPP), record linkage to establish validity, methods of obtaining answers to sensitive questions, and the use of simultaneous record checks to ascertain validity. Discussion is planned in that sequence.

Confidentiality and Invasion of Privacy

Before inviting participation from others, the Chairman introduced the subject of confidentiality and invasion of privacy. It seems ironic that at a time when society needs more data on the personal and family life of its citizens in order to plan, administer, and evaluate social programs of all kinds in health, education, and welfare, there is a growing tendency to view the collection of such data as an evil in itself. Undoubtedly, there have been instances of abuse and misuse of personal data files by persons doing so either intentionally or without realizing it. Thus, data on race in the field of health have come under criticism (Terris: 1973) and then later were defended (Greenberg and Cassel: 1974). At any rate, the potential for abuse and misuse of personal and confidential data always exists; therefore, there has been a growing tendency to prevent the collection of such data or to withhold its use to a narrow or limited audience. In some context, any demographic variable involving age, race, sex, place of origin, education,

marital status, and others can be viewed by some as confidential information and potentially embarrassing. Thus, although statisticians and survey researchers have been faced with problems of confidentiality ever since the first census in 1790, the individuals and agencies involved were aware of the highly sensitive nature of their data files and took unusual precautions to protect individuals and groups. The data were treated in a scientific manner and caution was exercised not to allow commercialization of the results or embarrassment and risk to any individual when data were disaggregated into minute components.

The advent of computers, credit cards, and the growth of large data banks on all types of persons and their characteristics began to chip away at the sanctity of the data and the methods previously considered adequate as safeguards. Within the last few years there has emerged a curious alliance of two large groups anxious to restrict the compilation and use of such data files. On the other hand, there are liberals who are concerned with the protection of civil liberties and the fear of big brother in government repressing the rights of individuals. Data files of the FBI and CIA may justify this concern. On the other side of the spectrum are those who would generally be considered politically conservative and who have been opposed to any data collection procedure that might impinge upon their freedom to act for their own interests. Thus, questions about family planning were early viewed as an invasion of the privacy of the bedroom. The interesting case that occurred prior to the 1970 census when a Congressman from Ohio tried to limit the type of data collected by the Bureau of the Census is another illustration of this concern. The item that triggered this reaction was the proposed question about a family's bathroom and whether it was private or shared by others.

A recent piece of legislation enacted in 1974 bears directly on the question of confidentiality and invasion of privacy in data collection efforts. It is called the Privacy Act and became effective on September 27, 1975.

The provisions of the Act were reviewed briefly (Jabine). It was pointed out that the Act applied to record systems directly controlled by Federal agencies and to systems operated for Federal agencies under grant or contract with private agencies. Retrievable records, both administrative and survey types, that identify individuals are the focus of this legislation. The Act includes the following provisions:

- 1) An inventory of data systems will be published so that the public will be advised of the existence of all data systems that meet the provisions of the Act. The compilation of a comprehensive inventory of eligible data systems is now under way.
- 2) The prospective respondents must be informed about the authority for collecting the data, whether mandatory or voluntary, and the use to be made

of the data gathered in the survey. The confidentiality provisions of pending data systems will be made known to the public by notices that will be inserted in the Federal Register annually or more frequently if changes are made in the confidentiality provisions.

- 3) Every individual has a right to know what information is contained in his data records in the Federal agency and to request corrections in his record in these data systems.
- 4) There are limitations on the transfer of data in record systems from one Federal agency to another for statistical purposes. A release from each individual is required unless either of these conditions is satisfied:
 - a) A notice in the Federal Register states that release of the information to specified agencies for specific purposes is intended. Transfers to other agencies are possible if the purpose is compatible with the purpose with the reasons the data were originally collected.
 - b) Transfer of information to the Census Bureau will be permitted without prior notification to the respondent if the information being released is related to the Census program.

Discussion by participants centered on the provisions of the Act and their interpretation. It was pointed out that some provisions were subject to various interpretations and that it was not clear that the Act applied to all Federal agency data systems financed under contract with private and non-Federal agencies. It is not clear how the Act applies to data collected in the past. The view was expressed that the Act is a threat to statistical collection and compilation operations and that it is doubtful it would permit repeating some important statistical studies involving transfer of information from one agency to another such as the Birth Registration Test and the University of Chicago National Study of Social Class Differentials in Mortality. Also, there is the danger that the Privacy Act will serve as model legislation for state and local governments. It was proposed that professional groups acting collectively rather than independently, exert influence to modify the legislation. An avenue for such action might be through the National Research Council, which has a committee (chaired by Dr. Alice Rivlin) that is investigating the negative consequences of overly restrictive legislation on linking data files. Furthermore, the Committee on National Statistics, National Academy of Sciences, has an arrangement with the Bureau of the Census to convene a panel to investigate the effects of the Privacy Act.

This problem is not peculiar to the United States. Several Western European countries have in recent years adopted legislation to protect the confidentiality and privacy of individual records. In Sweden this has restricted the scope of statistical studies. Dalenius (1974) has presented an overview of the invasion of

privacy problem and discusses statistical techniques for overcoming this problem.

The effect of the Act with regard to Social Security numbers was mentioned. The opinion was expressed that the Act did not ban the use of the Social Security number on non-SSA records but makes the reporting of the number a voluntary matter for the respondent. Any universal identifier as a requirement is banned by the Act.

The discussion on confidentiality and invasion of privacy may be summarized as follows:

- 1) Recent legislation has generated a conflict of interest between the need to protect the confidentiality of individual records and the need to produce essential statistics required for social and economic program planning and evaluation. This conflict of interest presents a real but unnecessary threat to one producing essential statistics.
- 2) This conflict of interest is not an inherent feature of data systems. As a matter of fact, traditionally, the major statistical Federal agencies have been the most outspoken proponents of and contributors to the policy and practice of assuring and providing confidentiality of information about individuals.
- 3) Recent legislation has contributed to the problem by its lack of specificity and clarity. Thus, the legislation is open to various interpretations and it is vague about matters of its implementations. Its primary weakness, however, is that it fails to make clear a distinction between statistical records, administrative records used for programmatic and regulatory purposes, and administrative records used for research purposes.
- 4) Several conference attendees proposed that professional societies represented by statisticians, social scientists, and related groups, should seek to amend the Privacy Act and influence the implementation of its provisions so that a clear and viable distinction is made between record systems used for statistical purposes and those used for program and regulatory purposes while, at the same time, preserving features of the Act that strengthen the confidentiality of records of individuals.

Record Linkage

This subject of record linkage (Waksberg) began by stressing the technical difficulty of conducting linkage studies, especially from the viewpoint of matching records for the same individuals in different data sets. Alternative objectives for conducting linkage studies were described as follows:

- 1) To evaluate statistics that are generated by a data system. For example, the Census Bureau has conducted coverage studies involving linkage of birth records and Census records, and income studies involving IRS and Census records. Establishing posi-

tive matches between records in different systems is always a difficult problem, and there are errors due to mismatches and nonmatches. Frequently, it is difficult to decide which of the matched records is the more valid one.

- 2) To supplement the statistics obtained from a data system. In the Medical Economic Study being conducted by Johns Hopkins University and Westat under contract with the National Center for Health Statistics, the household expenditures on medical care being collected from a panel of consumers in a household sample survey are being supplemented by data obtained from the records of medical sources providing the care and the records of health insurance companies making the third party payments.
- 3) To obtain outcome statistics to evaluate a non-statistical program. For example, the effect of Employment Training Acts is being evaluated on the basis of the future earnings of trainees as reflected in Social Security records of the trainees.

The Chairman recommended that the group concentrate on the first type of study and address the question, "How do you decide which record is correct?"

It was observed that there is no universal rule. In many studies, however, the answer is reasonably clear. For instance, utility costs based on the records of utility companies are probably better than those based on responses to a household survey. Similarly, physician costs based on doctor records are probably more valid than those reported in household surveys. There are instances, however, in which one cannot assume that records of physicians are more valid than responses in household surveys. For example, for pregnancies involving a fixed cost for a specified regimen of care, the records of the obstetrician may fail to list all visits since the patient is not charged separately for each of them.

Several participants introduced their own experiences in trying to determine validity of records. In the pretest of teen-age drug use (Haberman), alternate labeling methods were investigated. More drug users were enumerated when the persons provided their names, which was the least anonymous of the labeling methods. Records being used for validation purposes may be plagued by the same problems as the statistics being validated. For example, the D.C. Drivers Test (Boisen) attempted to use drivers' licenses to estimate undercoverage of black males in critical age groups in the Census. However, the addresses on the drivers' licenses were subject to gross inaccuracies.

There is reason to be suspicious about equating an increase in frequency of reporting with greater validity. For example, studies of college groups (Boruch) indicate that there is overreporting of marijuana use and driving while drinking and other types of behaviors approved by peer groups. Although

the guidelines are not entirely clear the Alcohol and Drug Abuse Acts of 1970 and 1971 empowered the HEW Secretary to grant testimonial privilege to social researchers working on these topics.

There are ways of checking physician records to determine accuracy in reporting. One method is internal consistency. For example, for persons with allergies, the standard practice (Sudman) is to require an allergy shot once every week or two. If the patient reports on a regular basis for three months for a weekly shot and the physician's record indicates sporadic visits by the patient, one suspects that it is the physician who is in error. A study was done in Saskatchewan (Fuchsberg) comparing household interview survey data with physician records. It was discovered that 15 percent of physicians' claims were not filed and, hence, never appeared in the record system.

The use being made of records can sometimes be a guide to the records validity. For example, it is socially desirable for a physician to specify his teaching hospital affiliation when reporting information for AMA records, even though he may have no such affiliation (Monteiro).

Sometimes underreporting may be due to matching problems. In one study, (Cannell) about 15 percent of the admissions were not reported by hospitals. By repeatedly returning the unmatched names to the hospital, the underreporting was reduced from 15 percent to 2 percent.

There is a growing concern about the effect the invasion of privacy and record linkage problems will have on response rates. For example, in one study (Woolsey) hospital records are being used to estimate the incidence of fairly rare diseases. The plan is to supplement the information in the hospital records by conducting surveys with the patients and their families. Thus, the names and addresses of patients as recorded in the hospital files are needed. Some hospital authorities have concluded that family authorization is needed before their hospitals can participate in the study. One solution would be to have the hospitals serve as the agent for collecting the information from the families; another solution would be to have the hospitals request authorization from the families. (Neither solution seems to be ideal because of inherent problems in both.) A legal type authorization form versus an informal letter approach is being considered.

In concluding the topic of record linkage, the point was made that perhaps the group was being overly pessimistic about the use of validation studies. For example, the factors associated with disease etiology are rarely determined by a single epidemiological study. Similarly, it may take a combination of several validation studies before conclusions can be reached about bias errors. In its validation studies, the Census Bureau depended on several studies to estimate response errors.

Sensitive Questions

The Chairman mentioned that at least five techniques have been found useful in gathering data on sensitive questions, protecting data confidentiality during transmission over telephone lines and/or while in storage on computers and in data banks, or in restricting the interpretation of published data so that confidentiality is retained. These methods are by no means a complete listing because they omit such obvious techniques as anonymous replies, sampling of variates, use of interval measurements, and other procedures.

Randomized Response

The first technique discussed was randomized response, a technique developed only ten years ago by Stanley Warner (Warner: 1965). The term randomized response is a misnomer because it is really a response to a randomized question. To illustrate use of the technique in one of its simplest forms, the Chairman demonstrated its application. His objective was to ascertain what proportion of the conference participants had cheated on their Federal income tax in the year 1974. For purposes of definition, cheating was defined as the underreporting of income, such as dividends, interest, honoraria, or consultant fees by \$25 or more, and/or the overstatement of deductible items, such as medical expenses, business expenses, or charitable contributions by a like amount.

The participants were asked to take a coin from their pocket and proceed as follows:

"Toss the coin in the air and if the result is 'heads', keep that fact to yourself but answer the sensitive question 'Yes' or 'No'. The sensitive question is whether or not you cheated on your income tax last year. If the result of the coin toss is 'tails', answer the nonsensitive question in the same way. The nonsensitive question is whether your mother was born in the month of April. (If you do not know your mother's exact month of birth, substitute your own month of birth but keep the fact to yourself.)"

The Chairman then wrote on the blackboard the same instructions.

HEADS I cheated on my Federal income tax last year.

(Yes or No)

TAILS My mother was born in the month of April.

(Yes or No)

All those who wished to reply "Yes" raised their right hand, and fourteen were counted. Those who wished to reply "no" then raised their hands and thirty-six were counted.

14 Yes

36 No

—
50 Total = N

The Chairman explained that if there had been no cheating on income taxes last year, the only persons who would have raised their hands would have been those who had a coin turn up "tails" and whose mother was born in April. The expected number of such persons is approximately $(\frac{1}{2} \times \frac{1}{12})$ of 50, or slightly over two. Thus, there were $14-2=12$ persons who admitted to cheating as defined. To convert this number to a percentage, only one-half of the 50 persons would be expected to have a coin turn up "heads" and were thus supposed to answer the sensitive questions. Hence, the proportion of cheaters is calculated as follows:

$$\frac{14-2}{\frac{1}{2} (50)} = \frac{12}{25} = 48\%$$

28

Several questions were raised by the conferees immediately following this demonstration. Would the respondent believe that randomized response did not violate his privacy? Would he not be equally as willing to reply anonymously on a piece of paper placed in a sealed envelope? Is there any comparison of cooperation between these two methods? Does tossing a coin create more suspicion among respondents than other randomizing devices? Has there been a validation of responses?

The Chairman responded to some of the queries before inviting general discussion from the floor. He said he knew of no studies comparing the randomized response versus the anonymous reply placed in an envelope but expressed a personal preference for the coin toss because of a possibility that an interviewer might open the envelope and record the respondent's name after leaving the interview. With respect to validation, he mentioned a randomized response survey on annual income conducted in North Carolina. The results for both black and white families were within a few dollars of the averages published by the Bureau of Labor Statistics for the southeastern part of the United States.

The role of education of respondents and the possible effect of education on respondent cooperation was raised. The (Brown and Harding: 1973) study of drug use among officers and enlisted men was reviewed (Horvitz). In all cases of drug usage except marijuana, the reported use was greater by randomized response than by anonymous questionnaire. The increase in reported use was greater among officers and this may be either because they understood the method better or they felt more threatened by the possibility of apprehension through the anonymous questionnaire. There is no question that the more threatening the respondent perceives by the question, the more value there is attached to the procedure of randomized response.

In a study of induced abortion in Taiwan, Chow and others (I-Cheng, Chow and Rider: 1972) used randomized response among the general population in that country. In fact, they used complicated sam-

pling devices consisting of a volumetric flask containing colored balls and a cloth bag containing colored stones. The results were similar to all the other studies of induced abortion in such populations. The rates estimated by randomized response were high and, as indicated, in line with what one would expect to have occurred if the truth could be ascertained.

A report was made (Sudman) on studies that suggest in special situations in which randomized response reduces underreporting, although it does not eliminate bias. Samples were selected from Court records of people who had declared bankruptcy, and from persons arrested for drunken driving, and also from the general population. Comparisons of underreporting were made between randomized response, face-to-face interviews, telephone interviews, and self-administered questionnaire. In the case of drunken driving, randomized response was best although there was still some underreporting. On bankruptcy, there was zero underreporting using randomized responses. But, when dealing with socially desirable attributes such as voting, randomized responses did not appear to work at all. In summary, not all response errors are eliminated by randomized response. Moreover, whenever data are to be classified by other variables, such as age, race, or sex, the sacrifices are even greater because the sample sizes are reduced.

Questions were raised as to whether or not the respondent might feel he was giving away his privacy by having to report "Yes" and whether there was an ideal nonsensitive question. In reply, the point was made that the nonsensitive question undoubtedly affects the respondent cooperation and influences the variability of the estimate. If the frequency of the nonsensitive question is symbolized by π_y , the respondent member of the sensitive group is given the maximum protection when $\pi_y=1$. Thus, in the demonstration concerning cheating on income taxes, if the coin toss was tails, the respondent could have been instructed simply to reply "Yes." In that way, at least 50 percent of the replies would have been in the affirmative and the respondent would have the maximum protection were he in the sensitive class. Of course, a "No" response would mean the respondent is answering the question on cheating and privacy is thereby lost.

If $\pi_y=1$, protection for the sensitive group member is greatest but the variance is also large. Conversely, if $\pi_y=0$, it is almost the same as if the direct question had been used because a "Yes" reply indicates membership in the sensitive class. Of course, in the latter case the sampling variance would be at a minimum. The Chairman felt that it is desirable to try to select π_y at approximately the same level of frequency as estimated for the sensitive question. That choice provides adequate protection from respondent suspicion and is close to the minimum variance.

More research on this issue is needed on this subject, however, as well as on respondent perception

and how it relates to his cooperation and willingness to tell the truth. Some persons might perceive a difference in a sampling device that was using the sensitive question 50 times out of 100 versus one with a probability of one-half from a coin toss.

It was pointed out that there was a need for a rough idea of the magnitude of the bias in a direct question approach to decide whether the increased sampling variance was worth the sacrifice. This issue was discussed in some of the first few articles on randomized response published in the *Journal of the American Statistical Association*. It was shown there that one need not have more than 5 to 10 percent untruthful reporting of a binomial variable to more than compensate for sacrifice resulting from the randomized response technique. The opinion was expressed that bias will vary with the socioeconomic status of the individual and how threatening the question is to him. More research is needed to compare the different methods so as to measure the bias with sufficiently large groups of respondents of differing backgrounds and with a wide variety of sensitive questions.

Reference was made to a study in which a validity comparison confirmed randomized response technique at an early stage of its development. This was a study of illegitimacy in North Carolina in which sample households, obtained from birth records on file with the state health department, were visited and randomized response used to determine whether or not an illegitimate birth had occurred. The proportion of illegitimate births in the sample was known in advance and was used for comparison with the randomized response estimate. In white households the correct answer was 7.7 percent illegitimate births whereas randomized response estimates were 7.4 percent was estimated. Black households were purposely selected to yield an illegitimacy rate of around 50 percent. The true value among the birth samples was 45.4 percent and through randomized response 42.3 percent was estimated. This latter was in a sample size of less than 100.

The possibility was raised as to whether or not a direct question might have done just as well when dealing with illegitimate births. In response to this, another study was cited in which a direct question was asked about births occurring in households. Only households in which a birth had actually occurred were included in the sample. Among illegitimate births, over 50 percent of the respondents reported no birth had occurred in that household. This, of course, is a very large bias.

The point was made that in almost all surveys reporting confidential data the degree of bias is not shown. If one reports the response rate at 85 percent, we do not know if this is high or low. It was also observed (Dalenius) that one advantage of randomized response would be if a court were to subpoena survey records, nothing could be used against a par-

ticular respondent. This is another reason randomized response is so valuable; moreover, developments in the technique since 1965 have been tremendous. Finally, there is no reason randomized response cannot be used in conjunction with other methods, including direct response.

In summary, the consensus seemed to be that randomized response has a lot to offer in those special situations where the respondent may feel threatened with an invasion of his privacy. More research and applications need to be undertaken on the use of sampling devices, new designs, use of innocuous questions, use in mail surveys, and the role of the interviewer. The interviewer has a more important influence in randomized response than with structured interview schedules because he or she must not only be convinced of the value of the method but be prepared to answer questions to allay any respondent suspicions.

Coding Designs

Coding designs are methods useful to collect data as well as to protect their confidentiality during transmission over telephone lines or storage in computers. The coding procedure, since data are stored in binary sets, is a series of 0's and 1's. The same sequence is used to decode as to code the data if binary sets are used as in a computer.

The coding designs can be combined with randomized response so that the sequence of 0's and 1's may be random as long as the program of generating them is kept secret. This influences the calculation of correlation coefficients between sets of data. It was mentioned that about 1971 there was a Ph.D. thesis by William Barksdale at the University of North Carolina that discussed this problem of correlated data in randomized response.

These coding techniques are useful where many persons have access to data in the computer. The point was made that a cryptic device is useful to protect confidential data inhouse. Also, in publishing data, sometimes data in a cell may be subjected to randomized response as long as the marginal totals are retained. (Reference was made to a report in the Office of Education on this subject).

It was also pointed out that sometimes a face sheet with identifying data can be stored separate from the data sheets as long as there is a linkage file. Attending Census Bureau representatives confirmed their desire to protect the privacy of data transmitted over telephone lines.

Weighing Designs

Weighing designs are survey techniques that can also be used to collect sensitive data. The original work on weighing designs done by Hotelling many years ago weighed small objects on a balance scale. With this method instead of measuring separately

the weight of objects X and Y in turn, with error of σ for each weighing, one can obtain as much information in two weighings as from four. The ingenious device is to weight X+Y, and then X-Y. The weight of X is one-half the sum of these two results, and the weight of Y is one-half the difference of the two results. The result is that each estimate has the same standard error as if it were based on the mean of two direct weighings (Wallis and Roberts: 1956, Banerjee: 1975).

This concept is easily transferred to collecting sensitive data where Y may be a threatening question and X not. Thus, suppose one-half of the sample respondents are asked:

30

"How many times did you go to the movies during the last month? *plus* How many abortions have you had during the past year?" The other half of the sample respondents are asked: "How many times did you go to the movies during the last month? *minus* How many abortions have you had during the past year?"

Obviously, these are only illustrations of the technique and one has to use care to choose X so that it is always greater than Y to avoid negative numbers.

There are many variations on this technique that can be combined with randomized responses, such as Federer and his colleagues at Cornell have done. They used balanced incomplete block designs. Thus, they were interested in obtaining estimates of seven variates so that each group of seven respondents was asked to report the totals of three questions, as follows:

$$\begin{aligned} Y_1 &= X_1 + X_2 + X_4 \\ Y_2 &= X_2 + X_3 + X_5 \\ Y_3 &= X_3 + X_4 + X_6 \\ Y_4 &= X_4 + X_5 + X_7 \\ Y_5 &= X_5 + X_6 + X_1 \\ Y_6 &= X_6 + X_7 + X_2 \\ Y_7 &= X_7 + X_1 + X_3 \end{aligned}$$

That is, the first respondent in each group of seven was asked to report the sum of variates X_1 , X_2 , and X_4 . The second respondent was asked to report the sum of X_2 , X_3 , and X_5 , and so on. The interviewer does not know which three questions the respondent has added for his answer since the latter drew one of the seven possibilities at random. The estimating equations are straightforward (Smith, Federer, and Raghavarao: 1974, Raghavarao and Federer: 1973).

One participant (Waksberg) observed that what worried him in estimating the Y question in the simple design with (X+Y) is that the variance of the nonsensitive X question may be so much greater and one is not really reducing the variance of Y. One should pick an X with a low variance in the population.

Contamination or Error Inoculation Methods

The Chairman expressed regret that Dr. Boruch had to leave the conference early, because the latter

has contributed many ideas and applications to this area in which contamination procedures are purposely introduced to mask the true value of an observation. The method can be used to inject error by the respondent in his reply so as to protect its confidentiality. The contamination might also be used in data storage in computer files or whenever confidential data are published for small cells or areas that might be easily identified.

Dr. Horvitz commented that in one version of this technique the interviewer directs the respondent to use some randomization choice in order to determine whether to lie or tell the truth when replying to a sensitive question. One random choice might be simply to lie when answering a sensitive question, whereas the other random choice would be to tell the truth. What happens is that false negatives and false positives occur, and one has to correct for them in estimating the true proportion. Thus,

$$\pi = \frac{\rho - \alpha}{1 - \alpha - \beta}$$

Where α = false positive rate
 β = false negative rate
 ρ = reported proportion
 π = true proportion

It was observed that this is somewhat similar to what Dr. Kenneth Poole reported in his recent article (Poole: 1974). Poole was interested in the distribution function of a continuous variable and used income distribution as an illustration. He combined the contamination with randomized response by asking the respondent to multiply the true response by a random number and to tell the interviewer only the final result. A similar technique involves adding or subtracting a random number, with mean zero, to the true response and reporting only the contaminated algebraic sum. A question was asked whether this is not a weighing design and, if not, what the difference is between a weighing design and contamination.

The Chairman stated there was a structural similarity but that a weighing design involves the reporting of sums of several components, not individually identified, without any inoculation of error. In the contamination procedure, one adds a contaminant at random. For example, suppose the respondent's income is \$20,000. The respondent is asked to choose a random number between 1 and 5 from some device. If he selects 3, the reported answer is \$60,000 and the interviewer would not know whether the true value is \$12,000, \$15,000, \$20,000, \$30,000 or \$60,000. The contaminant is selected at random.

As noted earlier, Dr. Boruch was not able to be present to emphasize the value of this procedure or to amplify its applications in certain instances of sensitive data. Boruch has compared the method to randomized response both in a theoretical sense and in actual field trials. The design can also be made more complicated in ways other than by simply inoculating false positive and false negative errors. Others

have examined some of these designs and two handy references are the papers by Warner (Warner: 1971) and Greenberg, *et al* (Greenberg, Horvitz and Abernathy: 1974). Before leaving Dr. Boruch gave the Chairman a copy of his latest effort in this area. The report involves the use of the technique to preserve data file confidentiality (Campbell, Boruch, Schwartz, and Steinberg: 1974).

Network Surveys of Rare and Sensitive Conditions

Sirken discussed network surveys as a method of dealing with sensitive questions by protecting confidentiality.

The health and related conditions about which respondents are sensitive and feel threatened when asked about them in population sample surveys, are often rare conditions. Thus, survey estimates of these conditions are not only subject to substantial underreporting but to large sampling errors as well. Various design strategies have been proposed (Sirken: 1970) for estimating rare health conditions, but few strategies have been proposed for estimating conditions that are both rare and sensitive. In these remarks, Sirken described briefly (1) the network survey methods for controlling both the sampling error and response bias and (2) an interesting application of this method to a household sample survey of substance use that was recently conducted for the Michigan Office of Drug Abuse (Sirken: 1975).

The essential design feature of the network survey of substance use is that the drug user is permitted to be enumerated at more than one enumeration unit. To adjust for the contingency that not all drug users are eligible to be enumerated the same number of times, network estimators require ancillary information that is not needed by the estimators of conventional surveys since the latter would not permit the same drug users to be enumerated more than once. Several unbiased network estimators have been reported (Birnbaum and Sirken: 1968, Hsieh: 1970). One of these, the multiplicity estimator, weights every enumerated drug user by the inverse of the number of enumeration units where the user is eligible to be enumerated. The ancillary information needed to calculate the counting rule weight is usually collected from the person who reports the drug user in the survey. For example, if the household survey adopted a counting rule that made drug users eligible to be reported by their friends, the person in the survey who reported a friend as a drug user would also report the number of the user's friends.

The Michigan Survey of Substance Use estimated the prevalence of substance use during the preceding year for alcohol and 15 different kinds of drugs. Two sets of estimates were produced. One set, referred to as the conventional or self-estimates, was based on questions in which the sample persons reported their own use. The other set, referred to as network or friends estimates, was based on projective questions

to which the sample persons reported the percentages of their friends who used each substance. The friends estimator of substance use was the average of the percentages of friends users reported by sample persons in the survey. The friends estimates were between 50 and 200 percent higher than the self-estimates for each of the 10 nonprescribed and illicit drugs but somewhat smaller than the self-estimates for alcohol and for 4 of the 5 prescribed drugs. The sampling variances of the friends estimates were uniformly (25 to 50 percent) smaller than the sampling variances of the self-estimates.

The friends estimates are puzzling. Why are they larger than the self-estimates of nonprescribed and illicit drugs? In this connection, one can note that the question about friends use preserve the anonymity of the drug users since their identities are not divulged by the friends who report them in the survey. Hence, the questions on nonprescribed and illicit drug use by friends might be less threatening and, hence, less subject to underreporting than the questions on self-use of these drugs. However, this does not explain why the friends estimates appear to be about the right order of magnitude. A possible explanation was offered along the following lines.

The friends estimator, being a network estimator, would be unbiased if every user enumerated in the survey were weighted by the inverse of the number of times he was eligible to be enumerated. In this case, the number of times a user is eligible to be enumerated is equal to the number of the user's friends. In the Michigan Survey, however, the estimator weighted the enumerated users by the inverse of the number of the respondent's friends. Thus, a sufficient condition for the estimator based on friends use in the Michigan Survey to be unbiased is that the friends of drug users each have about the same number of friends and this number is equal to the number of friends of the drug user. The condition would be satisfied, if, for example, the friends of a drug user were friends of each other and none of them had any other friends. The fact that the friends estimator has smaller sampling variance does not necessarily imply that it is superior to the self-estimator because (1) the estimates, based on both estimators, are subject to measurement errors that would arise in conducting the surveys, and (2) for fixed sample size the survey costs are greater for the friends estimator than for the self-estimator. Selected experiments need to be conducted to estimate the mean square error and cost components associated with the two estimators to determine the conditions under which one or the other estimator is indicated. Since the preliminary findings from the Michigan Survey findings are intriguing, they deserve to be investigated, replicated, and hopefully improved.

Dr. Eckerman suggested that it was possible that the friends estimator overstates drug use.

This may simply be a function of lack of knowledge and misapprehension regarding usage by others. Chanck (1932), in an early study of norms in a rural community, coined the term "pluralistic ignorance" to account for the fact that while many household respondents actually deviated from church instilled norms prohibiting card playing and use of alcohol and tobacco they at the same time contended their neighbors and friends adhered to these norms. We may be encountering a similar phenomenon in the drug abuse field but with people overestimating rather than underestimating their friends' drug usage. He suggested that research was needed to investigate this matter. Sirken agreed, but added that it is generally believed that population surveys underestimate the prevalence of non-prescribed and illicit drug use.

32

Record Checks

The Chairman asked two or three participants to discuss the question of how to use record checks to establish validity.

Mr. Shapiro noted that the Health Services Research and Development Center of The Johns Hopkins University, in collaboration with Westat, Inc., is testing alternative survey methods for collecting information on medical utilization and expenditures under a contract with the National Center for Health Statistics. A household panel is being requested to maintain diaries on health care experience over a six month period. Two experimental variables are being tested: periodicity of reinterview (monthly vs. bi-monthly) and type of contact (in-person vs. telephone). The cost-effectiveness of the alternative strategies is being measured using several criteria including accuracy and completeness of household data determined through comparisons with data records of health care providers and third party payers.

Dr. Federspiel discussed the Medicaid Program in Tennessee. The primary objectives of the project are to ascertain validity and to get some idea of the extent of improper prescribing of drugs. The data files are being studied. The file of medical service claims that contain the diagnoses for services provided has been matched with the file of prescription claims that identifies the purchased drug. The matched records have disclosed inappropriate prescribing of some drugs in the validation process.

Mr. Jabine reported that record check studies (Steinberg: 1973, Scheuren, Bridges, and Kills: 1973, Scheuren, Kills, and Oh: 1973, Leamer: 1974, Robbins and Siegmund: 1974, Dyer: 1974), involving inter-governmental agency data linkage (the Census Bureau, Internal Revenue Service, and the Social Security Administration) were conducted to improve the quality of statistics on income distribution. The studies have linked income data reported in the Current Population Survey with earnings and benefit data reported in Social Security records. The Pilot Link

Study was conducted in 1963, and the Exact Match Study was conducted in 1973. The Social Security number was one of the variables used to match records in the two data systems. Another record check study currently underway involves linking of reports of Social Security income payments in the March 1975 Current Population Survey with reports of earnings records in the files of the Social Security Administration.

It was further observed that a selected bibliography has been completed on the matching of person records from different sources (Garey and Hwang: 1974).

References

1. Bailey, M. B., Paul W. Haberman and J. Sheinberg
1966 "Identifying Alcoholics in Population Surveys: A Report on Reliability," *Quarterly Journal of Studies on Alcohol*, Vol. 27, pp. 300-315.
2. Banerjee, Kali S.
1975 *Weighing Designs*, New York: Marcel Dekker, Inc.
3. Birnbaum, Z. W. and Monroe G. Sirken
1968 "Designs of Sample Surveys to Estimate the Prevalence of Rare Diseases: Three Unbiased Estimates," *Vital and Health Statistics*, National Center for Health Statistics, Publication No. 1000, Series 2, No. 11, Washington, D.C.: U.S. Government Printing Office.
4. Brown, George H. and Francis D. Harding
1973 "A Comparison of Methods of Studying Illicit Drug Usage," Technical Report 73-9, Human Resources Research Organization, April.
5. Campbell, Donald T., Robert F. Boruch, Richard D. Schwartz and Joseph Steinberg
1974 "Confidentiality—Preserving Modes of Access to Files and to Interfile Exchange for Useful Statistical Analysis," Appendix A., *Report of National Research Council Committee on Federal Agency Evaluation Research*.
6. Dalenius, Tore
1974 "The Invasion of Privacy Problem and Statistics Production—An Overview," *Statistik Tidskrift*, Vol. 3, pp. 213, 225.
7. Dyer, Alan R.
1974 "Hypothesis Testing Procedures for Separate Families of Hypothesa," *Journal of the American Statistical Association*, Vol. 69, No. 345 (March), pp. 140-150.
8. Garey, M. R. and F. K. Hwang
1974 "Isolating a Single Defective Using Group Testing," *Journal of the Amer-*

- ican Statistical Association, Vol. 69, No. 345 (March), pp. 151-155.
9. Greenberg, Bernard G., Daniel G. Horvitz and Joseph R. Abernathy
1974 "A Comparison of Randomized Response Designs," *Reliability and Biometry* by Frank R. Proschan and R. J. Serfling, SIAM, pp. 787-815.
 10. Greenberg, Bernard G. and John C. Cassel
1974 Letter to the Editor on "Desegregating Health Statistics," *American Journal of Public Health*, Vol. 64, No. 2 (February), pp. 98, 170-171.
 11. Haberman, Paul W. et al
1972 "High School Drug Behavior: A Methodological Report on Pilot Studies," S. Einstein and S. Allen, Editors, *The Proceeding of the First International Symposium on Student Drug Use*, Farmingdale, N.Y.: Baywood Publishing Co., pp. 103-121.
 12. Hsieh, N. C.
1970 "Some Estimator Techniques for Utilizing Information from Elements Not In the Sample." Survey Research Center, UCLA.
 13. I-Cheng, Chi, L. P. Chow, and Rowland V. Rider
1972 "The Randomized Response Technique as Used in the Taiwan Outcome of Pregnancy Study," *Studies in Family Planning*, A Publication of the Population Council, Vol. 3, No. 11, pp. 265-269.
 14. Johns Hopkins Medical Institution
1975 "Analytic Plan for the National Center for Health Statistics Medical Economic Study," The Health Services Research and Development Center, Unpublished.
 15. Leamer, Edward E.
1974 "False Models and Post-Data Model Construction," *Journal of the American Statistical Association*, Vol. 69, No. 345 (March), pp. 122-131.
 16. Poole, Kenneth W.
1974 "Estimation on the Distribution Function of a Continuous Type Random Variable Through Randomized Response," *Journal of the American Statistical Association*, Vol. 69, No. 348 (December), pp. 1002-1005.
 17. Raghavarao, D. and W. T. Federer
1973 "Application of BIB Designs as an Alternative to the Randomized Response Method in Survey Sampling," *BU 490 M in the Mimeo Series of the Biometrics Unit*, Cornell University, Ithaca, New York.
 18. Robbins, Hubert and David O. Siegmund
1974 "Sequential Test Involving Two Populations," *Journal of the American Statistical Association*, Vol. 69, No. 345 (March), pp. 132-139.
 19. Scheuren, Frederick, Benjamin Bridges and Beth Kills
1973 "Report No. 1: Subsampling the Current Population Survey: 1963 Pilot Link Study," DHEW Publication No. (SSA) 74-11750 (August).
 20. Scheuren, Frederick, Beth Kills and H. Locke Oh
1973 "Report No. 2: Coverage Differences, Noninterview and Nonresponse, and the 1960 Census Undercount: 1963 Pilot Link Study," DHEW Publication No. (SSA) 75-11750 (December).
 21. Sirken, Monroe G.
1970 "Survey Strategies for Estimating Rare Health Conditions," *Proceedings of Sixth Berkeley Symposium of Mathematical Statistics and Probability*, University of California Press, pp. 135-144.
 22. Sirken, Monroe G.
1975 "Evaluation and Critique of Household Surveys of Substance Use," Chapter IV, *Alcohol and Other Drug Use and Abuse in the State of Michigan*, Office of Substance Abuse Services, Michigan Department of Public Health.
 23. Smith, L. L., W. T. Federer and D. Raghavarao
1974 "A Comparison of Three Techniques for Eliciting Answers to Sensitive Questions," *BU-525-M, Mimeo Series, Biometrics Unit*, Cornell University Ithaca, New York, and 1974 *Proceedings of the Social Statistics Section*, American Statistical Association.
 24. Steinberg, Joseph
1973 "Some Observations of Linkage of Survey and Administrative Record Data," DHEW Publication No. (SSA) 74-11750.
 25. Terris, Milton
1973 "Desegregation Health Statistics," *American Journal of Public Health*, Vol. 63, No. 6 (June), pp. 477-480.
 26. Wallis, W. Allen and Harry V. Roberts
1956 *Statistics, A New Approach*, Glencoe, Illinois: The Free Press, pp. 477-478.
 27. Warner, Stanley L.
1965 "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American Statistical Association*, Vol. 60, No. 309 (March), pp. 63-69.
 28. Warner, Stanley L.
1971 "The Linear Randomized Response Model," *Journal of the American Statistical Association*, Vol. 66, No. 336

Summary and Conclusions

Persons collecting data in surveys should always be on guard to ascertain the validity of the responses. Bias may occur not only because the information involves a sensitive area of questioning but also because the respondent may not know or remember the true facts.

Sensitive questions can be ordinally scaled into four categories as follows:

1. *Illegal actions or behavior* (e.g. cheating on income tax, driving under the influence of alcohol, speeding).
2. *Not illegal but socially deviant behavior* (e.g. specific unusual sex practices, alcoholism or other drinking patterns frowned upon by society).
3. *Embarrassing facts* (e.g. bankruptcy, failure in school, dishonorable discharge).
4. *Confidential data wherein privacy is sought* (e.g. earnings, voting behavior, history of illness).

In attempting to validate data, one technique is to use other available records. In such cases, one must be careful not to violate the provisions of the Privacy Act of 1974 especially since it is becoming possible to link more records by means of computers and the use of universally used identifying numbers such as found on Social Security cards and drivers' licenses.

It was agreed that some provisions of the Act threaten to stop many of the kinds of legitimate research and validation procedures statisticians have used for decades. There was consensus that statisticians and survey researchers should attempt, through their professional organizations in having objectionable portions of this legislation modified to permit research that cannot harm an individual either directly or as a member of a group. Specifically, the point was made that a distinction should be drawn between record systems used for statistical purposes (e.g. vital statistics, census) and those used for program and regulatory purposes. Suggestions were made to channel these efforts through the National Research Council and the Committee on National Statistics of the National Academy of Sciences.

Record linkage as a means to check validity was discussed as were some of the alternative uses of linking records from various systems. Concentrating on the usage of record linkage for the purposes of confirming validity, it was pointed out that frequently one is faced with having to decide which record is the correct one. One should not assume that the record with the larger number of undesirable attributes, or greater frequency of asocial behavior, is automatically or always the correct one. Also, linkage of records has another problem in the number of mismatches and nonmatches. It was the consensus that matching records should be continued in many studies and a

larger number or a combination of sources should be sought.

In the area of sensitive questions, there was discussion of five procedures found to be useful to encourage respondents to cooperate and to reply more truthfully. The first technique discussed at length was the randomized response in which the question itself is selected by the respondent, and the data gatherer does not know from the answer which question is chosen. The usefulness of this method was considered in various settings and it was generally agreed that more field trials should be conducted to ascertain the method's applicability in the four previously listed categories of sensitive questions. Especially needed are studies of the different factors, such as age, education, and economic status that influence the respondent's comprehension and cooperation in randomized response. In addition further studies are needed on the sampling device used to select the question although preference seemed to fall on the toss of a coin. The interviewer's influence in the use of randomized response and its applicability to questionnaires in mail surveys were also found to need further exploration.

Two other methods used to gather data on sensitive issues were the coding designs and weighing designs. The former is interesting because it is also easily adaptable to the protection of the confidentiality of data stored in computers or transmitted over telephone lines. The weighing designs are not a new concept and have a structural similarity to balanced incomplete blocks in experimental design.

The contamination or error inoculation method was also discussed because it too can be adapted to protect the confidentiality of stored or transmitted data. The method has interesting possibilities, particularly with educational and psychological measures, and further research should be conducted to study different variations of the technique as well as field trials of its usefulness.

The fifth and last technique to collect data on sensitive issues that was discussed is the network survey involving multiple respondents. It has a certain resemblance to the sociograms and sociometric analysis to study each respondent's relationship and/or rating to others in his or her network. The consensus was that this is another area needing further exploration to ascertain not only how to improve the technique of estimation but also field trials to learn where the procedure may be useful.

Needed Research

1. There is increasing evidence that several new techniques in survey research can reduce and perhaps eliminate the bias caused by untruthful reporting or the refusal to answer questions about sensitive issues. One of the most potentially fruitful is that randomized response and further research is needed to conduct field studies to establish how well the method overcomes bias in the

- four categories of sensitive questions enumerated in the Summary and Conclusions. Particular attention should be focused on the interviewer's influence as well as the effect of age, and socio-economic status of the respondent.
2. The validity of data gathered by surveys and special studies should always be examined by checking various records to obtain information from other sources. Research is needed so that survey users can learn about new techniques of linkage in order to match records for validity. Moreover, additional studies should attempt to illustrate how some of the infrequently used sources of official statistics might be developed for establishing validity of response. These needs are particularly great when the data involve sensitive issues such as those enumerated in the four categories mentioned in the Summary and Conclusions.
 3. The use of weighing designs, contamination of data, coding systems, and network surveys using multiple respondents are also useful techniques in learning about sensitive issues. More research is needed to compare the efficiency of these methods vis-a-vis the direct question and randomized response. It is especially important to ascertain which procedure is optimal under specific circumstances.
 4. Further research in randomized response is needed to establish the reaction and perception of respondents to the method and to ascertain the amount of risk or jeopardy they are willing to tolerate before refusing to cooperate or resorting to untruthful replies. This relationship will vary with the degree of sensitivity as scaled in the four categories listed in the Summary and Conclusions.
 5. Additional studies should be made of the acceptability of various sampling devices used in randomized response and the contamination methods. These include decks of cards, coins, dice, sealed transparent plastic boxes, the random number target, or the volumetric spherical flask with colored balls invented by research workers at Johns Hopkins University.
 6. More studies should be made on the use of the randomized response involving quantitative variables.
 7. Further research is needed on the usefulness of randomized response and contamination methods in mail questionnaires, telephone surveys, and situations other than the personal interview.
 8. Studies need to be made on how to establish the most correct record when multiple record checks are instituted. This problem is especially acute for questions that may involve sensitive items. Special example should be developed to illustrate how to use longitudinal studies, additional or supplemental records, specific panels, and other respondents to ascertain validity. Also, what will be the involvement of respondents themselves to develop techniques to improve validity. This need for validity checks is especially important when overreporting may be operative, and the fallible assumption is sometimes made that the source showing greater use of frequency is automatically judged to be the correct one. Various hypotheses need testing according to the categories of varying sensitivity as enumerated in the Summary and Conclusions. In some categories, overreporting may be more serious than underreporting as the source of the bias.
-

TOTAL SURVEY DESIGN

Daniel G. Horvitz, Ph.D., Chairman
Kirk Wolter, Ph.D., Rapporteur

36 Introduction

The chairman opened the session by defining and then discussing "total survey design" in some detail. Total survey design (TSD) is a concept that implies a balanced allocation of survey resources among the different error components in order to minimize the total error of estimate. For example, the researcher who invests a portion of a given survey budget so as to reduce bias arising in the measurement process rather than using the entire budget to reduce the sampling error by increasing the sample size, is attempting to apply the total survey design concept. If the particular budget allocation results in the smallest total survey error achievable for the given survey conditions and budget, then the survey researcher is successfully applying the TSD concept.

To use the TSD concept, an error model is required that can be applied to surveys in general. Such a model must be able to include all of the different error components that arise in surveys. The Bureau of the Census model developed by Hansen, Hurwitz, and Bershada (1961), and referred to earlier in the Conference, is such a model. It includes separate components of error such as the pure sampling error variance, the simple response variance (a measure of response reliability or response consistency), the correlated response variance (most often associated with interviewers), the interaction of the response error components with the sampling error,¹ and the bias or net systematic error. This model was originally developed for dichotomous variables and simple random sampling. Koch (1973) has recently extended this model to the multivariate case and for continuous as well as qualitative variables. The immediate value of this extension is twofold. First, it includes complex bivariate estimators, such as a regression and correlation coefficients and ratio estimation. Second, it is not confined to simple random sampling, but may be applied to unequal probability sampling designs.

¹ For example, this component can arise when those respondents whose exact measure of the variable of interest is less than average tend to underreport their exact measure and those whose exact measure is greater than average tend to overreport their exact measure.

Total survey error models also have extremely important long-range significance. They provide a basis or common frame of reference for putting into proper perspective methodological research concerned with improving the quality of surveys. Thus, alternative survey procedures (sample designs and measurement designs) can be partially evaluated by comparing the relative magnitudes of the different components of error in the total error model. As indicated by the TSD concept, complete evaluation of alternative survey procedures requires a cost function and knowledge of the cost components, as well as knowledge of the various error component parameters in the model. It follows that a total error model provides a basis for evaluating or adding up the value or meaning of all the survey methodological research conducted to date. Such a summation would quite likely reveal significant gaps in our knowledge of survey errors. Nevertheless, the error and cost models together with those estimates of model parameters that are available for a given survey strategy, will provide the feedback mechanism so essential to more cost-effective choices of future survey strategies.

Survey designs that permit the total mean square error to be estimated require a method for estimating the correlated response variance component. When data are subject to correlated response errors, that component is not included in the usual variance estimates. Also, a separate procedure for estimating the net bias is often required, although certain sources of bias (or adjustments for bias) can be measured as part of the regular survey design.

Two additional references are Bailer (1975) and Lessler (1974). The first of these provides an excellent discussion of the various error components in the Bureau of the Census model and their magnitudes for selected 1970 census variates. The second reference provides a basis for making rational survey design decisions for the case (mentioned above) in which some investment is made to eliminate (reduce) bias by using inexpensive but imperfect measurements on all respondents and costly but accurate measurements on a subsample of respondents.

Questions for Discussion

The chairman then asked the session participants for discussion of the following question: if minimization of the total error of estimate for a given survey resource level is a valid survey design goal, what are the implications of this goal for the design and conduct of methodological studies?

A second discussion was concerned with the utility of a survey error parameter (components) computerized information system; that is, a system based on a total survey error model and a standardized set of error component definitions acceptable to both social scientists and statisticians. Initially, the information system would contain estimates of the error model parameters for survey measurements reported in the literature. Once established, the information system would be available to the survey research community in general, which in turn would contribute new data on error components and costs from future surveys and methodological studies.

Conceptually, the error components estimates would be stored in an n -dimensional matrix with $(n-1)$ of the dimensions providing essential descriptive information of the specific measurement design, i.e., type of population or subpopulation, context of survey, sample design (e.g., stratification, size of sample), variable measured, exact wording of question, method of measurement (e.g., personal interview, mail questionnaire, telephone interview), and relevant cost data. The remaining dimension would contain the specific error component parameters such as the sample design effect, simple response variance, and bias.

Total Survey Design Discussion

The first point made in response to the general question (Jabine) was that substantial expenditures of time, money, and manpower are required to produce accurate estimates of the components of the total mean square error. This was illustrated by a study conducted at the Bureau of the Census (Jabine and Tepping: 1973) the purpose of which was to estimate each of the components of the total mean square error for certain occupation and industry items. In that study accurate estimates of all the components were produced except for the bias term. To estimate the bias, record checks were performed. However, despite great effort and substantial expenditures, the resulting standard errors of the bias estimates were too large to admit inferential statements. Consequently, the authors were unable to evaluate the magnitude of the bias of the variable components of error.

It was pointed out that the principal uses intended for the survey data must be considered in applying the TSD concept (Woolsey) and, in fact, should be a determining factor in the allocation of funds to control the various error components. For

example, if one is estimating the change in unemployment over a certain time period, the allocation may differ greatly from the case in which geographical comparisons are to be made within a given survey.

The chairman agreed with this comment, adding that the discussion of the acquiescent respondent the previous day illustrates the point well. Dr. Carr had suggested a measure of acquiescence was needed for every respondent to adjust for the distortion or error it introduced into data on anomie. Since the principal purpose of a survey might be to estimate the correlation between anomie and, say, social status, it is quite possible that a TSD approach would suggest that it would be more cost-effective to measure acquiescence on only a subsample of the respondents and then use that data to adjust the estimate of the correlation of interest.

Many trade-offs are involved in any TSD strategy. One trade-off discussed (Fowler) arises when considering different reporting periods for which data are to be obtained. For example, if the characteristic of interest is the number of recent visits to a physician, the statistician may encounter considerably different non-sampling errors depending upon how "recent" is defined whether to refer to the past week, the past month, or the past year. A decision to use the past week could elicit highly accurate data, but might very well require a considerably larger sample, and hence additional cost, to achieve a sufficient number of physician visits for analysis purposes. From a cost standpoint, it may be better to ask for physician visits in the past year, despite the increase in response bias.

In spite of the noted difficulty in applying the total survey design concept, it was remarked (Dalenius) that it provides the only rational approach available to the statistician. Other approaches only produce very special results at best. The overuse and misuse of the word "optimum" is illustrative of this. In reality there is no theory for an "optimum sampling design." There are merely local optima that apply in very special cases, i.e., optimum allocation in stratified sampling.

Another important point stated (Jabine) at the session was that the statistician or health services researcher must first define the variables with which he is attempting to optimize his design strategy. This is a necessary precondition to the construction of a total survey design. In this context, the possible lack of measures of the components of error appropriate to surveys concerned with estimating changes over time was noted.

Finally, it was pointed out (Bradburn) when there is an external validity criterion, the response error may include a bias term, i.e., a deviation from a true value. When there is no such criterion, as in the case of attitudes, only response variation can be measured.

Total Survey Design Matrix

The method of conceptualizing the TSD information system concept suggested by the chairman was then discussed. This, as described earlier, would involve a large matrix whose cells would contain the cumulative past history and experience of survey researchers. Such a matrix would identify untapped areas for further research as well as provide guidance in the design and analysis of future surveys.

It was suggested (Eckerman) that the so-called design matrix was a useful method to conceptualize total survey design that a valuable first step would be to lay out the dimensions of the matrix. It was urged that this be accomplished initially through a systematic inventory of what is known about the methodological issues.

Berelson and Steiner's book entitled "Human Behavior: An Inventory of Scientific Findings," while perhaps too general for a person trained in social psychology, is no doubt very useful for the uninitiated as a means of obtaining an overview of the field. In a similar way, a recommendation might be forthcoming from this conference toward the assemblage of such an inventory in the field of survey research as a means of systematizing for the beginner what is known.

It is apparent, in a compendium of surveys of drug abuse developed by William A. Glenn, that the level of sophistication extant among members of this conference is unfortunately infrequently found in newly burgeoning areas of research. There should be some ready means of familiarizing researchers, new to the field, with the intricacies of survey research. The Census Bureau's Technical Report No. 34 is a useful beginning but involves, in just the first volume, over 25,000 entries. An inventory might be a means of highlighting the most important and relevant findings of the past few years.

Another important point mentioned (Jabine) with regard to the design matrix was the need to standardize terms. It was felt that with a recognized standard set of terms, the matrix would serve as an information system from which data and insight into the survey design problem could be retrieved.

Reference was made to the recent effort to standardize background items used in survey research (Eckerman). This task has been carried out under the direction of the Center for the Coordination of Research on Social Indicators and a report is available (Reeder). An effort designed to evaluate commonly used instruments and scales in surveys in terms of their reliability and validity apparently ran into many problems (Hensler), relatively few instruments could be evaluated. Nevertheless, the results are to be published this summer as a monograph entitled Health Surveys Reference Index (Reeder).

In discussing the TSD matrix, the participants found it useful to distinguish between two fundamen-

tally different worlds of health services researchers (Shapiro).

First, there are those researchers who are involved with repetitive, large-scale surveys. Allied with this group are the methodologists whose efforts are directed at improving survey results and eliminating measurement error. It is perhaps with this group that the prime responsibility rests for filling out the cells of the design matrix.

The second group of researchers is those who are involved with "one-shot surveys." This group suffers in that monetary resources are almost never available to investigate fully or estimate each of the components of the total mean square error. Moreover, in many cases this would be beyond their capabilities even if resources were available.

A discussion of the dimensionality of the TSD information matrix followed along with its implications for future methodological research (Marquis). The principal implication of this discussion was that methodological research should not be done unless the results could be entered into the information matrix. Each successive study should add to the store of knowledge by filling in empty cells. It was recognized that some studies would not necessarily provide data along all n-dimensions of the matrix, but that the findings might still satisfy admissibility criteria along providing only conditional results.

Funding Problems

One of the participants (de la Puente) remarked that precious little money is allocated for the design of "one-shot" surveys. The chairman remarked that this is one of the prime motivational features of the total survey design matrix. It is exactly because the ad hoc survey designer does not have money to investigate errors that may arise in his design that he needs guidance from other sources. The TSD matrix would provide this guidance by making available the cumulative experience of others working with similar designs and data.

Several participants (Sirken, Shapiro, and others) expressed concern that no one organization has the prime responsibility for funding or performing original research on methods of evaluating and dealing with nonsampling errors. To cope properly with the nonsampling error problem requires detailed and expensive planning, yet very little research money has been made available exclusively for this purpose.

Another viewpoint was also expressed (Greenberg and Waksberg). This view was that any large agency has the responsibility to study survey methodological problems. Although this view has been expressed many times, the combined experience of the conference participants was that no strong trends in this direction are apparent within the federal statistical establishment, except for such agencies as the Bureau of the

Census and the National Center for Health Statistics.

It was noted (Cannell) that the primary concern of many health services researchers and most funding agencies is with the items of costs, sampling errors, and response rates. In this connection, it was suggested that if valid estimates of additional components of the total mean square error were available, these would provide researchers with persuasive arguments in their attempts to obtain research funds.

A vital element of the total survey design matrix then must be cost. Survey designers need to know the cost trade-offs involved in their design alternatives. A design matrix would provide this information by acting as a "shopping list." This list would inform the methodological researcher of the extent of the bias, simple response variance, correlated response variance, and sampling variance he is buying with a particular survey procedure.

The comment was made (Dalenius) that uniform definitions of costs would be required before the cost "shopping list" could be formulated. Presently, uniform international cost definitions do not exist.

Hansen, Hurwitz, Bershah Model

As discussed in the introduction, one of the early attempts to provide a mathematical model for non-sampling errors was the Hansen, Hurwitz, and Bershah (1961) model. That model, amplified by Hansen, Hurwitz, and Pritzker (1964), has been used extensively at the Bureau of the Census.

The comment was made (Woolsey) that even in the absence of a total survey design matrix, the mere existence of the Hansen, Hurwitz, Bershah model has provided survey statisticians with valuable insight into the design process. By making educated guesses of the costs and the magnitudes of error, the survey designer has been able to make sensible choices amongst design alternatives by reference to the measurement error model.

In a related comment, it was noted (Sirken) that although the Hansen, Hurwitz, Bershah model identifies the parameters associated with measurement error, the survey profession has not as yet provided estimates of these parameters. This was illustrated by noting that nonresponse percentages are frequently quoted while the magnitude of the biasing effect of nonresponse is rarely known (Schuman).

A contrary point of view, was expressed that estimates of some measurement error parameters have been presented. The important question then is how much does the aggregate survey profession know regarding nonsampling error levels?

Presentation of Nonsampling Errors

Although much of the discussion centered on problems of funding, it was also suggested (Waksberg) that perhaps it is not yet known how reports

of nonsampling error levels should be presented for specific surveys. This problem has been discussed without resolution at the Census Bureau, where reports of separate components of measurement error have been published, but where a total mean square error figure has yet to be produced.

Additionally, it was noted (Jabine) that the Census Bureau possesses numerous unpublished estimates of certain components of the mean square error. It would be useful to health services researchers and the survey profession in general if Census published more of these estimates in conjunction with survey results.

With regard to the presentation of errors, a call (Dalenius) for more honest reporting was made. Often survey results are claimed to be statistically significant when in fact the estimated standard errors are so large as to render such statements false. Moreover, full discussions of all the errors of estimate are rarely presented in the survey literature. By failing to account for all error sources, the survey researcher understates the total level of error and makes inferential statements based on the understated errors. More honest reporting throughout the profession would be helpful in this regard.

Other Comments

It was noted (Dalenius) that in spite of funding difficulties and the general lack of methodological research on nonsampling errors, great progress has occurred in the past 25 years. Illustrative of this is the Birnbaum and Sirken (1950) paper that dealt with the optimum size of nonresponse. This helped researchers to understand the trade-offs inherent in a total survey design.

Also with regard to the nonresponse problem, it was pointed out (Fowler) that perhaps health services researchers should make a more concerted effort to obtain information about nonrespondents than is currently practiced. It was felt that many nonrespondents would provide information related to the principal objectives of the survey even though refusing a formal interview. Such information could be quite valuable in assessing the bias due to nonresponse.

It was remarked (Carpenter) that many forms of response error may result from the respondents' lack of perception. In some studies (Hamblin: 1971), this lack of perception has been shown to be physiological. In such cases, the bias may be predictable. Once measured, this perception bias could be adjusted for future surveys in which similar general conditions prevailed.

Of course, it does not necessarily follow that the magnitude of error observed in one survey will be replicated in another survey. Nevertheless, it seems clear that many kinds of error will repeat themselves, and to deal with these sources it is necessary to build a reserve of past information upon which to base future decisions. The chairman observed that this is the

motivation for the construction of the total survey design matrix.

References

1. Bailar, Barbara A.
1975 "Some Sources of Error and Their Effect on Census Statistics," Paper Presented at the Annual Meeting of the Population Association of America, April 18, Seattle.
2. Berelson, Bernard and Gary A. Steiner
1964 *Human Behavior: An Inventory of Scientific Findings*, Harcourt, Brace and World, Inc.: New York.
3. Birnbaum, Z. W. and Monroe G. Sirken
1950 "Bias Due to Nonavailability in Sampling Surveys," *Journal of the American Statistical Association*, Vol. 45, No. 249 (March), pp. 98-111.
4. Glenn, William A.
1975 *Recent Surveys of Nonmedical Drug Use: A Compendium of Abstracts II*. Final Report, Contract No. HSM-42-72-169, Prepared by Research Triangle Institute, Research Triangle Park, North Carolina, for National Institute of Drug Abuse, Rockville, Maryland, March.
5. Hamlin, Robert L.
1971 "Mathematical Experimentation and Sociological Theory: A Critical Analysis," *Sociometry*, Vol. 34, No. 4 (December), pp. 423-452.
6. Hansen, M. H., W. N. Hurwitz and M. A. Bershad
1961 "Measurement Errors in Censuses and Surveys," *Bulletin of the International Statistics Institute*, Vol. 38 (Part II), pp. 359-374.
7. Hansen, M. H., W. N. Hurwitz and Leon Pritzker
1964 "The Estimation and Interpretation of Gross Differences and the Simple Response Variance," *Contributions to Statistics Presented to Professor P. C. Mahalanobis on the Occasion of his 70th Birthday*, C. R. Rao, Editor, Calcutta: Pergamon Press, Ltd., pp. 111-136.
8. Jabine, Thomas and Benjamin Tepping
1973 "Controlling the Quality of Occupation and Industry Data," presented at the Meeting of the International Statistical Institution in Vienna, August.
9. Koch, Gary G.
1973 "An Alternative Approach to Multivariate Response Error Models for Sample Survey Data with Applications to Estimation Involving Sub-class Means," *Journal of the American Statistical Association*, Vol. 68, No. 344 (December), pp. 906-913.
10. Lessler, Judith T.
1974 "A Double Sampling Scheme Model for Eliminating Measurement Process Bias and Estimating Measurement Errors in Surveys." Institute of Statistics, *Mimeo Series No. 949*, Chapel Hill, N. C., The University of North Carolina.
11. Schanck, R. L.
1973 "A Study of a Community and Its Groups and Institutions Conceived of as Behaviors of Individuals," *Psychology Monograph*, Vol. 43, No. 2.
12. U.S. Bureau of the Census
1974 *Indexes to Survey Methodology Literature*, Technical Paper No. 34, Washington, D.C.: U.S. Government Printing Office.

Summary and Conclusions

Although the Total Survey Design (TSD) session generated some interesting discussion, it seems clear that the concept needs considerable further discussion in the literature in terms readily understood by survey practitioners. The possible implications of TSD to the design and conduct of methodological research are not immediately obvious and were not really discussed in any great detail in the TSD session. Thus, the question remains unanswered whether it is useful to carry out methodological studies that concentrate on a single component of error, thereby producing results that are highly conditional.

The need to measure the cost components associated with a given measurement design in order to be able to make use of the findings in the TSD sense requires emphasis and reemphasis. The lack of detailed cost data for specific alternative measurement designs that are readily available for use by the survey research community is viewed as serious. For example, the TSD concept can be used very effectively to choose the appropriate length of recall period for reporting utilization of health services, say, provided methodological studies were carefully designed to measure the variable error components as well as bias components for different reference periods covering a sufficiently wide range of alternatives. Clearly the costs associated with collecting sufficient data for annual statistics, for example, as well as the magnitude of the error components, can contribute significantly to the ultimate choice of recall period.

There was no clear consensus by the participants concerning the utility of the suggested information

matrix for survey error components and cost components. On the one hand, the value of such information can only be assessed in a context in which the need for better decisions among alternative survey strategies, and the role of the TSD concept in meeting that need, is widely recognized and accepted. On the other hand, the very act of establishing an information system for use by the survey research community could substantially expedite the process of improving the overall cost-effectiveness of survey designs.

The TSD session generated discussion concerning the lack of adequate funds committed to research on health survey methods. The growing need for more detailed and accurate data on health care needs and the utilization of health care facilities and services

emphasizes the survey research budgetary concerns of those participating in the session.

Needed Research

1. Implementation of research priorities in other sections of this report:
 - a. Necessitates the consideration of TSD as an integral part of their respective protocols; and
 - b. Is essential to the orderly development of TSD theory.
2. Additional research is needed in the development of methodologies for improved simulation and modeling techniques to determine the cost-effectiveness of various sampling design mixes and measurement strategies.

INDEX OF SUBJECTS

- 42
- Abortion 30
 - Induced 28
 - Accuracy 24
 - Acquiescence 11, 16, 17, 37
 - Advance Letters and Preparation 14, 15
 - Administrative Records 26
 - Age 28
 - Interviewer 17
 - Respondent 10, 17
 - Aided Recall Method 8
 - Alcohol & Drug Abuse Acts of 1970 & 1971 27
 - Annual Income 27
 - Anomie 11, 17, 37
 - Anonymous Questionnaires 28
 - Anonymous Replies 27, 28
 - Balanced Incomplete Block Design 30
 - Bankruptcy 28
 - Baseline Interviews 8
 - Bias 2, 8, 19, 24, 28, 29, 36, 37, 39, 40
 - Binomial Variable 29
 - Birth Registration Test 25
 - Black
 - Respondent, Low Income 11
 - Interviewers 18
 - Bounded Recall 8
 - Bureau of the Census 8, 25-27, 29, 32, 36, 37, 39
 - Calibration Techniques 9
 - Census Validation Studies 27, 32
 - Center for Health Administration Studies 8, 10
 - City Directories 14
 - Coding 2
 - Designs 29
 - Techniques 29
 - Coin Toss 22, 28
 - Commitment 1, 19, 20
 - Committee of National Statistics,
 - National Academy of Science 6, 25
 - Compensation (to Respondents) 3, 16-19
 - Complexity 1, 3, 7, 9, 10, 12
 - Computerized Information System 37
 - Computers 5, 16, 25
 - Conditional Results 38
 - Confidential; Confidentiality 24, 25, 27, 29-31
 - Data 3, 6, 24-27, 29
 - Information 24-26
 - Protection of 26
 - Provisions 24
 - Contamination; Contamination Procedures ... 29, 30
 - Continuous Variable 24
 - Cooperation 8, 10, 11, 16, 19
 - Correlated Response Variance 36, 39
 - Cost 1, 2, 5, 7, 9, 13, 16, 36, 37, 39, 41
 - Components 1, 2, 6, 36, 41
 - Function 36, 37, 39
 - Cultural Effects 10, 11, 18
 - Curanderos 11
 - Current Medical Survey 8
 - Current Population Survey 8, 13, 32
 - Data 4, 5, 16
 - Banks 25, 27
 - Files 25, 30
 - In the Computer 16, 29, 30
 - D.C. Drivers' Test 26
 - Design Effects 31, 38, 39
 - Diary; Diaries 2, 8, 9, 12
 - Dichotomous, Dichotomous Variables 24, 36
 - Direct Questions 29
 - Direct Response 29
 - Distribution Function of the
 - Continuous Variable 30
 - Doctor Records; Physician Records 26
 - Drivers' License 26
 - Driving While Drinking 26
 - Drug; Drugs
 - Illicit Use of 32
 - Improper Prescribing of 32
 - Use 26, 28
 - Users 31
 - Drunken Driving 28
 - Education 10, 17, 23
 - Education of Respondent 10, 12, 17, 23, 28
 - Enumeration Unit 31
 - Error 1
 - Component 1, 6, 36, 37, 40, 41
 - Inoculation 30, 31
 - Model 36, 37, 39
 - Parameter 37
 - Standard 30
 - Exact Match 32
 - Face-to-Face Interviews 2, 28
 - False Negatives 30
 - False Positives 30
 - Fatigue 10
 - Federal Register 25
 - Feedback 15, 17, 19, 20
 - Feldshers 11
 - Friends Estimates 31, 32
 - Funds Committed to Research 41
 - Hansen, Hurwitz, Bershad Model 36, 39
 - Household Interview Survey 27
 - Household Panel 32
 - Household Surveys 27
 - Illegitimacy 29
 - Illitimate Births 29
 - Income Tax 18
 - Cheating On 27, 28
 - Federal 8, 27
 - Income Questions 20
 - Information Matrix 2
 - Inject Error 30
 - Inoculating False Positive 30
 - Inocuous Questions 29
 - Interaction 13, 15, 18-20, 22, 36
 - Interval Measurement 27

Interview		
Cost	22	
Length (Telephone)	15	
Panel	13, 15	
Interviewer	1, 3, 10	
Behavior	2, 17, 19	
Characteristics	2, 18	
Cues	14, 15, 17, 18	
Pace	18	
Race	2, 3, 17, 18	
Role of	23	
Training and Supervision	2, 3, 5, 15, 17, 19, 20	
Variance	15, 19, 10	
Inventory of Data Systems	25	
Johns Hopkins Medical Economic Study	8	
Johns Hopkins University	26	
Large Data Banks	25	
Length of Interview	3, 7, 10, 12	
Linkage		
File	25, 29	
of Record	24, 32	
Problem	27	
Studies	26	
Mail Surveys	2, 10, 29	
Marijuana	11, 20, 26	
Matching	17, 18	
Problems	27	
Records	24, 26	
Maximum Protectional	28	
Mean of Distribution	24	
Mean Square Error	24, 31	
Measurement of Central Tendency	24	
Medical Economy Study	26	
Medical Program in Tennessee	32	
Memory		
Aid	2, 8	
Decay	8	
Loss	7, 9	
Michigan Office of Drug Abuse	31	
Michigan Survey of Substance Use	31	
Minimum Variance	28	
Mismatches	26	
Model Neighborhood Health Centers	10	
Monotonic Function	8	
Multichotomous Scale	24	
Multiplicity Estimators	31	
National Center for Health Statistics	5-8, 32	
National Center for Health Statistics		
Medical Economic Study	32	
National Health Survey	8	
National Opinion Research Center	8, 10	
National Research Council	25	
Native Language Questionnaire	11	
Net Bias	36	
Net Systematic Error	36	
Network		
Estimators	31	
Surveys	2, 31	
Nonmatches	26	
Nonresponses	24, 39	
Nonresponse Bias	39	
Nonsampling Error	1, 5, 37-39	
Nonsensitive	30	
Nonsensitive Questions	27, 28, 30	
One Shot Surveys	23, 38	
Optimum Sampling Design	37	
Overreporting	17	
Outcome Statistics	26	
Panel Studies	7	
Perception	16, 39	
Perception of the Respondent	9, 17, 28	
Pilot Link Study	32	
Positive Matches	26	
Privacy	2, 20, 28	
Act	2, 25, 26	
Individual	26	
Invasion	5, 24-27, 29	
Problem, Invasion of	24, 26	
Violation	5, 15, 28	
Probes	18, 20	
Program & Regulatory Purposes	26	
Protection of Civil Liberties	25	
Proxy Respondents	2, 7, 9, 12	
Quality	2, 3-7, 10, 12, 16, 17, 20, 22	
Question		
Difficulty	15, 18	
Form	17, 18	
Pace	18	
Manipulation	19	
Race	28	
Interviewer	2, 3, 18	
Matching	2, 18	
Question, Topics of	18	
Respondent	2, 18	
Rand Health Insurance Study	8, 9	
Randomized Devices	27	
Randomized Responses	2, 27-30	
Rapport	19	
Recall Period	3, 7-12, 40	
Recall Techniques	8, 9	
Record Checks	8, 32	
Record Checks Study	32	
Records	3, 12, 27	
Record System	25	
Record System Used for Statistical Purposes	26	
Re-interview (Phone)	13	
Reliability	1, 2, 7, 10, 12	
Reporting Error	39	
Reporting Periods	7, 8, 37	
Response		
Bias	3, 19, 31	
Error	18	
Rates	2, 5, 7, 20, 39	
Rate & Race	17	
Rate and Compensation	16	
Set	16-18	

Respondent	2	Length	15
Behavior	10, 11, 17, 20	Limitations	3, 14
Burden	7, 9-12, 17	Panels	13, 15
Commitment	19, 20	Re-interview	13
Cultural Differences	10, 18	Response Rates	14, 15
Preparation	14, 15	Sample	14
Respondents	2, 7, 9	Ownership	14
Elite	15	Random Digit Dialing	14
Retreivable Records	25	Telescoping	3, 7, 8, 11
Right to Know	25	Threat	15, 18
Saliency	9; 12	Income Questions	20
Sample	2	Total Mean Square Error	36, 37, 39
Sampling		Total Survey Design	1-3, 6, 36-38, 40, 41
Devices	10, 28, 29	Total Survey Error	36
Error	2, 31, 36, 39	Total Survey Error Models	36
Variance	24, 28, 31, 39	Transfer of Data in a Record System	25
Variates	27	Transfer of Information	25
Sampling Error Variance	36	Transmission over Telephone Lines	27, 29
Self Administered Questionnaires	28	TSD Information Matrix	38-40
Self Reports	12	Underreporting	10, 17, 27, 28, 31
Sensitive	5, 18, 20, 28	Universal Identifier	25
Conditions	31	University of Chicago National Study	25
Data	2, 28	Untruthful Reporting	2, 24, 29
Groups	28	Validity; Validation	1-3, 5, 10, 12, 27, 32
Questions	9, 14, 15, 18, 20, 24, 25, 27, 28	Comparison	29
Sex	28	Records	27
Differences in Reporting	17	Responses	2, 27
Interviewer	2, 17, 18	Studies	10, 20, 27
Simple Random Sampling	36	Variable Error Components	37
Simple Response Variance	36, 37, 39	Variance	15, 19, 20, 30
Simultaneous Record Checks	24	Varying Lengths of Recall	7-10
Social Approval	11	Verifying Information	7
Social Class Differentials in Mortality	25	Volumetric Flask	28
Social Desirability	17, 18	Weighing Designs	29, 30
Social Security Administration	32		
Social Security Numbers	26		
Social Security Records	26, 32		
Socially Desirable Activities	18		
Socially Undesirable	18		
Socioeconomic Status	18		
Standard Error	30		
Statistical Records	26		
Status	10, 17, 18, 23		
Stimuli	11, 16		
Storage on Computers	27, 29		
Storage of Data in Computer Files	30		
Stratification	37		
Structured Interview Schedule	20, 29		
Survey			
Costs	31		
Instruments	7		
Survey Research Center	8		
Telephone	6, 13-16		
Costs	13, 22		
Differences Compared to Personal	3, 14, 15		
Directories	14		
Interview	2, 3, 28		
Advantages	3, 15		
Costs	13, 14, 22		

GLOSSARY

Aquiescence—A tendency of the respondent to base his reply on some stimulus other than the question content. It may be stimulated by the desire to please the interviewer, the agency collecting the data or some other cue such as the unbalanced question.

Agree-disagree—Form of question in which the respondent responds by stating his or her concurrence or non-concurrence with a statement.

Anonymous Replies—Survey information gathered in which the respondent's identity cannot in any way be linked to the information provided.

Bias (or Net Systematic Error) of a Survey Estimate—The difference between the expected value (taken over the sampling design and the distribution of measurement errors) of the estimator and the "true" value of the parameter being estimated. This is particularly acute in surveys concerning sensitive or confidential matters, and in which it might be expected that the estimates are consistently below or above the true population parameter. A consistent pattern of under or overreporting will result in bias.

Binomial Variable—In the context of survey research, a response to a question to which only two choices are possible. The respondent is instructed to pick the one which better describes his condition, behavior, or experience. The question is usually referred to as being on a dichotomous nominal scale.

Bounded Recall—An interview where the respondent is reminded of what he reported in an earlier interview and is then asked only to report on any new events that occurred subsequent to the bounding interview.

Contamination Procedures—A procedure in which the respondent is instructed to perform a simple probability exercise, and for a given outcome, to answer the survey question correctly or incorrectly. Given the distribution of the probability exercise and the reported rate of the behavior, the true rate can be estimated. In some cases, the probability exercise may instruct the respondent to add or multiply a random variable to the true response in order to protect the privacy of the reply. These procedures are designed to protect the survey respondent's privacy.

Coding Procedures—Techniques for providing unique numerical designations to data such that quantitative analysis of the data can be performed. These techniques may be used for assigning labels to survey respondents which, while allowing identification of data as coming from a single source, protects the identity of the person who is that source. The method can also be used to conceal the true value of data, especially that stored in computers, so that interpretation of the coded data is impossible and meaningless until the data are decoded.

Cost Model—A mathematical formulation of the costs which would be incurred through use of a given sample design. The preferred survey design is that which minimizes the total mean-square error of estimate for a given survey cost or which yields minimum cost given a specified level of precision.

Cryptic Device—A code or system of codes which conceals the true state of affairs. It may be used to conceal the identity of survey respondents.

Cue—Some characteristics of the interview or the interviewer, the question wording or the interviewer's behavior, including feedback, which influences the direction of answers to one or more questions.

Diary—A written record kept concurrently by an individual respondent or household about events that would usually otherwise be difficult to remember.

Dichotomous—A random variable is said to be dichotomous if it assumes only one of two responses or values.

Distribution Function—A (cumulative) distribution (cdf) is the total frequency of members of a variate with value less than or equal to some point, x . Probabilities with values less than or equal to some point, x . Probability (pdf) provides the probability of a value of x as a

function of x . The pdf, or frequency function, can be regarded as the derivative of the distribution function.

Error Model—A mathematical relationship which postulates the manner in which both sampling and nonsampling errors arise in the conduct and analysis of a sample survey.

The measurement error model developed at the Bureau of the Census postulates that each survey response is a realization of a random variable possessing finite second moments. Under this model the total variance of a survey estimate may be divided into several components:

Variance

- a. The *simple response variance* contribution to the total variance arises from the variability of each survey response about its own expected value. In terms of a simple random sampling design, the simple response variance is the population mean of the variances of each population unit.
- b. The *correlated response variance* is the contribution to the total variance arising from non-zero correlations (in the sense of the distribution of measurement errors) between the responses of sample units.
- c. The *response variance* of a survey estimator is the sum of the simple response variance and the correlated response variance.
- d. The *sampling variance* is that contribution to the total variance arising from the random selection of a sample, rather than a complete enumeration, from the population.
- e. The *interaction* contribution to the total variance of estimate is that component arising from a non-zero covariance between measurement error and sampling error.

External Validity Criterion—When an independent source of information exists regarding a population being surveyed, then the individual survey responses may be checked (vis-a-vis the independent source) for accuracy.

Follow-up—A procedure whereby those members of a selected sample for whom a response is not obtained by one data collection strategy (e.g., telephone or mail) are contacted by the same or another data collection strategy in order to increase response rate. It can also be used to designate repeated surveys among a panel of respondents.

Ingratiating Behavior—Behavior on the part of either the respondent and/or the interviewer designed primarily to please the other person.

Interviewer Feedback—Some verbal or non-verbal communication by the interviewer in response to respondent behavior.

Linkage—The process or technique for joining data describing a single sampling unit, usually a person, from one or more primary data sources. If institutional records were being used to check information provided by a survey respondent, such as a record

check, linkage would concern whether the institutional records apply to the survey respondent, whether the respondent and the records refer to identical points in time and spatial locus, and whether the respondent and institutional source define relevant matters in congruent ways. In other cases, the linkage might serve as supplemental information about a respondent in order to facilitate correlations and other analyses of association.

Matrix Sampling—A procedure to reduce the length of complex questionnaires by asking any respondent only a subset of all the questions of interest, rotating the subset among the respondents.

Memory Failure or Decay—The universally observed phenomenon that the longer ago the event occurred in the past, the more likely the respondent is to have difficulty recalling the event. This rule may not hold true where the event is associated with some dramatic period of time in the life of the respondent.

Monotonic—Referring to data that always move in the same direction or are constant with reference to time or another variable. The data never move in the opposite direction.

Multiplicity Estimators—An unbiased network estimator that weights the sample elements by the inverses of the number of enumeration units at which they are eligible to be enumerated. The information needed to determine the weight is collected in the survey from the enumeration units that report the elements.

Network Estimators—Estimators which adjust for the varying probabilities of enumerating elements in network surveys by appropriately weighting the sample elements.

Network Survey—In this context a technique of estimating the incidence of behaviors which are both rare and sensitive. The respondent is instructed to report the number of his friends who have committed the behavior under consideration, and then to estimate the number of friends of his friends who have committed the behavior.

Non-Response Rate—The complement of response rate. The numerator is those eligible respondents selected in a sample for whom information is not obtained because of refusals, not found at home, unavailable by reason of illness, incompetence, language difficulty, etc. The denominator is the total number of eligible respondents initially selected for the sample.

Overreporting—Survey responses which produce a higher estimate of the incidence of some event or characteristic than is accurate.

Panel—A study design involving re-interview or a series of questionnaires with the same sample or respondents (or household units) at two or more different times. Usually used to study changes over time, giving rise to longitudinal data.

Proxy Respondents—Respondents who provide information about other persons, generally within the same household, in addition to or instead of providing information about themselves.

Random Digit Dialing—A procedure for obtaining a probability sample of households with telephones. Numbers are selected at random from exchanges without prior knowledge of whether they are working numbers, business numbers, or residential household numbers. The strength of the procedure is the inclusion of those households with unlisted numbers. Caution must be taken to assure that the digits used, whether terminal or otherwise, are uniformly distributed.

Randomized Response—An answer to a question or set of questions randomly selected from a defined larger universe of questions in a survey. The technique is particularly well suited for obtaining sensitive information and assuring the respondent of anonymity. The respondent is instructed to respond to the sensitive question after a simple probability exercise in selecting the question. Since only the respondent knows the outcome of the probability exercise, there is no possibility that his or her response can be linked with certainty to his or her relationship with the sensitive issue. Since the distribution of outcomes to the probability exercise is known in advance, it is possible to estimate, for a population, the rate of the behavior or experience under consideration according to the given probability distribution of the sampling exercise, and the reported response rate.

Rapport—A broadly defined term used to refer to the quality of the relationship of interaction between the interviewer and respondent. Usually this refers to characteristics of warmth and friendliness and open communication in interpersonal relationship.

Recall Period—The time period over which a respondent is required to remember what events have occurred. This period is characterized by the total length of the time period by the elapsed time from the time of inquiry.

$A \qquad \qquad B \qquad \qquad C$
 $A \rightarrow B = \text{REFERENCE PERIOD}$
 $A \rightarrow C = \text{RECALL PERIOD}$
 $B \rightarrow C = \text{LAG PERIOD; MAY VARY IN TIME}$
 $\qquad \qquad \qquad \text{FROM 0 TO ?}$
 $A < B \leq C$

Record Checks—The comparison of information provided by a respondent in a survey, with information obtained from other sources, especially governmental or institutional records including census, Social Security, vital records, dispensaries, hospitals, mental health agencies, pharmacies, and municipal activities such as police and fire department functions.

Reliability—Correspondence, repeatability or consistency between identical survey questions, at two different times.

Respondent Burden—The level of demand placed upon the respondent necessary to answer the questions in the survey instrument. This includes the total time demands on the respondent, the demands on his memory, difficulty in understanding the question and possible embarrassment.

Response Rate—The percentage of an eligible sample for whom information is obtained. For an interview survey the numerator of the formula is the number of interviews. The denominator is the total sample size minus non-eligible respondents; that is, minus those not meeting the criteria for a potential respondent as defined for that particular study.

Response Set—A tendency to respond in a particular way based on a stimulus other than the content of the question.

Social Desirability Bias—Answers which reflect an attempt to enhance some socially desirable characteristics or minimize the presence of some socially undesirable characteristics. Source of the expectations or values influencing answers can be the person himself (ego-threatening), the perception of the interviewer, or society as a whole; may give rise to an acquiescent response.

Standardized Modules or Measures—Tested and validated measures of major variables, such as those dealing with illness and demographic characteristics. The use of standard measures and measuring techniques provide a basis for comparability of information from investigator to investigator.

Stratification—A design technique employed in sample surveys whereby the finite population is classified into several parts (or strata) and a random sample is independently selected from each stratum. The purpose of stratification is to reduce the sampling variance.

Telescoping—A reporting error in which the time an event occurred is remembered as having been more recent than it actually was. Events may also be placed backward in time.

Total Survey Design (TSD)—A concept that implies an efficient allocation of survey resources among the different error components in order to minimize the total error of estimates.

Total Survey Error—The aggregate of all components of error occurring in the conduct or analysis of a sample survey. Included in the total survey error are all sampling and nonsampling errors.

Total Mean Square Error—In a survey estimate this is the expected value of the squared difference between the estimator and the population parameter being estimated, where the expectation is taken over the sampling design and the distribution of measurement errors.

Unbalanced Format—Form of a question in which only one alternative or choice is stated in the ques-

tion. A balanced format includes both alternatives or all choices.

Underreporting—Survey responses that produce lower estimates of the incidence of some event of characteristics than is accurate.

Validity—A valid measure is one that measures what it claims to and not something else. Validity is a continuous concept so most measures fall between total validity and total nonvalidity. A totally valid measure is one without bias.

Variance—See Error Model

CONSOLIDATED REFERENCES

1. Abernathy, James R., Bernard G. Greenberg and Daniel G. Horvitz
1970 "Estimates of Induced Abortion in Urban North Carolina," *Demography*, Vol. 7, No. 1 (February), pp. 19-29.
2. Andersen, Ronald, Joanna Lion and Odin W. Anderson
1976 *Two Decades of Health Service: Social Survey Trends in Use and Expenditures*, Cambridge, Massachusetts: Balinger Publishing Co.
3. Anderson, Henry
1969 "Statistical Surveillance of a Title XIX Program," *Journal of the American Public Health Association*, Vol. 59 No. 2 (February), pp. 275-289.
4. Bailar, Barbara
1975 "Some Sources of Error and Their Effects on Census Statistics," Paper Presented at the Annual Meeting of the Population Association of America, April 18, Seattle.
5. Bailey, M. B., Paul W. Haberman and J. Sheinberg
1966 "Identifying Alcoholics in Population Surveys: A Report on Reliability," *Quarterly Journal of Studies on Alcohol*, Vol. 27, pp. 300-315.
6. Banerjee, Kali S.
1975 *Weighing Designs*, New York: Marcel Dekker, Inc.
7. Barksdale, William B.
1971 "New Randomized Response Techniques for Control of Non-Sampling Errors in Survey," Department of Biostatistics, Chapel Hill, North Carolina, June.
8. Berelson, Bernard and Gary A. Steiner
1964 *Human Behavior, An Inventory of Scientific Findings*, Harcourt, Brace and World, Inc.: New York.
9. Berkanovic, Emil
1974 "An Appraisal of Medical Records as a Data Source," *Medical Care*, Vol. XII, No. 3 (July), pp. 590-595.

10. Birnbaum, Z. W. and Monroe G. Sirken
1950 "Bias Due to Nonavailability in Sampling Surveys," *Journal of the American Statistical Association*, Vol. 45, No. 249 (March), pp. 98-111.
11. Birnbaum, Z. W. and Monroe G. Sirken
1968 "Designs of Sample Surveys to Estimate the Prevalence of Rare Diseases: Three Unbiased Estimates," *Vital and Health Statistics*, National Center for Health Statistics, Publication No. 1000, Series 2, No. 11, Washington, D.C.: U.S. Government Printing Office.
12. Bradburn, Norman M. and William M. Mason
1964 "The Order Effect of Question Order on Responses," *Journal of Marketing Research*, (November), pp. 57-61.
13. Brown, George H.
1974 "Drug Usage Rates as Related to Method of Data Acquisition," Technical Report 74-20, Human Resources Research Organization, August.
14. Brown, George H.
1975 "Randomized Inquiry Vs. Conventional Questionnaire Method in Estimating Drug Usage Rates Through Mail Surveys," Technical Report 75-14, Human Resources Research Organization, June.
15. Brown, George H. and Francis D. Harding
1973 "A Comparison of Methods of Studying Illicit Drug Usage" Technical Report 73-9, Human Resources Research Organization, April.
16. Bryant, Edward C. and Morris H. Hanson
1975 "Invasion of Privacy in Surveys: A Growing Dilemma," Paper Presented at the Smithsonian-Navy Conference on Survey Alternatives in Sante Fe, New Mexico, April 22-24.
17. Cahalan, Don, Ira Cisin and Helen Crossley
1969 *American Drinking Practices: A National Study of Drinking Behavior and Attitudes*, New Brunswick, N.J.: Rutgers University Press.
18. Campbell, Donald T., Robert F. Boruch, Richard D. Schwartz, and Joseph Steinberg
1974 "Confidentiality-Preserving Modes of Access to Files and to Interfile Exchange for Useful Statistical Analysis," Appendix A, *Report of the National Research Council Committee on Federal Agency Evaluation Research*.
19. Cannell, Charles F. and Floyd J. Fowler
1963 "A Study of the Reporting of Visits to Doctors in the National Health Survey," Ann Arbor, Michigan, Survey Research Center.
20. Cannell, Charles F. and Floyd J. Fowler
1965 "Comparison of Hospitalization Reporting in Three Survey Procedures," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 8, Washington, D.C.: U.S. Government Printing Office.
21. Cannell, Charles F., Gordon Fisher and Thomas Bakker
1965 "Reporting of Hospitalization in the Health Interview Survey," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 6, Washington, D.C.: U.S. Government Printing Office.
22. Cannell, Charles F., Floyd J. Fowler and Kent H. Marquis
1968 "The Influence of Interviewer and Respondent Psychological and Behavioral Variables on the Reporting in Household Interviews," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 26, Washington, D.C.: U.S. Government Printing Office.
23. Carr, Leslie G.
1971 "The Role Items and Acquiescence," *American Sociological Review*, Vol. 36, No. 2 (April), pp. 287-293.
24. Carr, Leslie G.
1975 "Class, Racial and Ethnic Factors in Agreeing Behavior: A Multi-Method Study of Acquiescence," (Unpublished Paper).
25. Chromy, J. and Daniel Horvitz
1974 "The Use of Monetary Incentives in National Assessment Household Surveys," *Proceedings of the American Statistical Association, Social Statistics Section*.
26. Ciocco, Antonio and Martha D. Ring
1951 "Methods in Public Health Research," *American Journal of Public Health*, Vol. 41, No. 8 (August), Part II, pp. 1-17.
and
Ciocco, Antonio and Martha D. Ring
1951 "Methods in Public Health Research," *Public Health Reports*, Vol. 66, No. 28 (July), Part II, pp. 883-897.
Proceedings of a Conference Held Under the Auspices of the Public Health Study Section, NIH, in Conjunction with the Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, May 10-11, 1950, Lowell J. Reed, Conference Chairman.
27. Cobb, Sidney
"Higher Illness Reported Among Unemployed Persons," Personal Communication to Bernard G. Greenberg.
28. Colombotos, John
1969 "Personal Versus Telephone Interviews:

- Effect on Responses," *Public Health Reports*, Vol. 84, No. 9 (September), pp. 773-781.
29. Colombotos, John, Jack Elinson and Regina Loewenstein
1968 "Effect of Interviewers' Sex on Interview Responses," *Public Health Reports*, Vol. 83, No. 8 (August), pp. 685-689.
 30. Coombs, Lolagene and Ronald Freedman
1964 "Use of Telephone Interviews in a Longitudinal Fertility Study," *The Public Opinion Quarterly*, Vol. 28, No. 1 (Spring), pp. 112-117.
 31. Dalenius, Tore
1974 "The Invasion of Privacy Problem and Statistics Production: An Overview," *Statistisk Tidskrift*, Vol. 3, pp. 213-225.
 32. Densen, Paul, Sam Shapiro and Eve Balamuth
1963 "Health Interview Responses Compared with Medical Records," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 7, Washington, D.C.: U.S. Government Printing Office.
 33. Dillman, Don A., Edwin Carpenter, James Christensen and Ralph Brooks
1974 "Increasing Mail Questionnaire Response: A Four State Comparison," *American Sociological Review*, Vol. 39, No. 5 (October), pp. 744-756.
 34. Dillman, Don A. and James H. Frey
1974 "Coming of Age: Interviews by Telephone," Unpublished Paper Presented at the Pacific Sociological Association Meetings, San Jose, California, March.
 35. Dillman, Don A. and James H. Frey
"Mail and Telephone Data Collection Methods," Forthcoming.
 36. Dillman, Don A., Jean Gallegos and James H. Frey
"Reducing Refusal Rates to Telephone Interviews," Forthcoming in *The Public Opinion Quarterly*.
 37. Dohrenwend, Barbara S., John Colombotos and Bruce P. Dohrenwend
1968 "Social Distance and Interview Effects," *The Public Opinion Quarterly*, Vol. 32, No. 3 (Fall), pp. 410-422.
 38. Dohrenwend, Barbara S., J. A. Williams and C. P. Weiss
1969 "Interviewer Bias Effect: Toward a Reconciliation of Findings," *The Public Opinion Quarterly*, Vol. 33, No. 1 (Spring), pp. 121-129.
 39. Dyer, Alan R.
1974 "Hypothesis Testing Procedures for Separate Families of Hypotheses," *Journal of the American Statistical Association*, Vol. 69, No. 345 (March), pp. 140-150.
 40. Elinson, Jack, E. Padilla and M. E. Perkins
1967 *Public Image of Mental Health Services*, New York City: Mental Health Materials Center, Inc., For N.Y.C. Community Mental Health Board.
 41. Fabrega, Horatio and Robert Roberts
1969 "Social Psychological Correlates of Physician Use by Economically Disadvantaged Negro Urban Residents," *Medical Care*, Vol. 10, No. 3 (May-June), pp. 215-223.
 42. Fowler, Floyd J.
1965 "Education, Interaction and Interview Performance," Unpublished Doctoral Dissertation, the University of Michigan.
 43. Fowler, Floyd J.
1973 "Final Report: Vermont Health Utilization Survey," Survey Research Program. Boston, Massachusetts, May.
 44. Garey, M. R. and F. K. Hwang
1974 "Isolating a Single Defective Using Group Testing," *Journal of the American Statistical Association*, Vol. 69, No. 345 (March), pp. 151-155.
 45. Gleason, Geraldine A.
1963 "Interviewing Methods in the Health Interview Study," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 48, Washington, D.C.: U.S. Government Printing Office.
 46. Glenn, William A.
1975 "Recent Surveys of Nonmedical Drug Use: A compendium of Abstract II," Final Report, Contract No. HSM 42-72-169, Prepared by Research Triangle Institute, Research Triangle Park, North Carolina, for the National Institute of Drug Abuse, Rockville, Maryland, March.
 47. Greenberg, Bernard G., Abdel-Latif, A. Abul-Ela, Walt R. Simmons and Daniel Horvitz
1969 "The Unrelated Question Randomized Response Model: Theoretical Framework," *Journal of the American Statistical Association*, Vol. 64, No. 326 (June), pp. 520-539.
 48. Greenberg, Bernard G. and John C. Cassel
1974 Letter to Editor on "Desegregating Health Statistics," *American Journal of Public Health*, Vol. 64, No. 2 (February), pp. 98, 170-171.
 49. Greenberg, Bernard G., Daniel G. Horvitz and Joseph R. Abernathy
1974 "A Comparison of Randomized Response Designs," *Reliability and Biometry* by Frank R. Proschan and R. J. Serfling, SIAM, pp. 787-815.

50. Gurin, Patricia, Gerald Gurin, Rosina C. Lao and Beattie Muriel
1969 "Internal-External Control in the Motivational Dynamics of Negro Youth," *Journal of Social Issues*, Vol. 25, No. 3 (Summer), pp. 29-53.
51. Haberman, Paul W.
1963 "Differences Between Families Admitting and Denying Existing Drinking Problem," *Journal of Health and Human Behavior*, Vol. 4 pp. 141-145.
52. Haberman, Paul W.
1971 "Psychiatric Symptoms Among Puerto Rico and New York City," *Sociological Abstracts*, Vol. 19 Suppl. 17 (August), p. 1166.
53. Haberman, Paul W. et al.
1972 "High School Drug Behavior: A Methodological Report on Pilot Studies," S. Einstein and S. Allen, Editors, *The Proceedings of the First International Symposium on Student Drug Use*, Farmingdale, N.Y.: Baywood Publishing Co., pp. 103-121.
54. Haberman, Paul W.
1973 *Study on Sanitary Criteria For Salt Water Bathing Beaches: Pretest of Illness Inquiry System and Site Selection, Part 1, 1972*, November 1973, and *Part 2, 1973*, January 1974, Office of Research and Development, U.S. Environmental Protection Agency.
55. Hamblin, Robert L.
1971 "Mathematical Experimentation and Sociology Theory: A Critical Analysis," *Sociometry*, Vol. 34, No. 4 (December), pp. 423-452.
56. Hansen, M. H., W. N. Hurwitz and M. A. Bershad
1961 "Measurement Errors in Censuses and Surveys," *Bulletin of the International Statistics Institute*, Vol. 38 (Part II), pp. 359-374.
57. Hansen, M. H., W. N. Hurwitz and Leon Pritzker
1964 "The Estimation and Interpretation of Gross Differences and the Simple Response Variance," *Contributions to Statistics Presented to Professor P. C. Mahalanobis on the Occasion of his 70th Birthday*, C. R. Rao, Editor, Calcutta: Pergamon Press, Ltd., pp. 111-136.
58. Henson, Ramon, Aleda Roth and Charles F. Cannell
1974 "Personal Vs. Telephone Interviews and the Effects of Telephone Reinterviews on Reporting of Psychiatric Symptomatology," Survey Research Center, Institute for Social Research, The University of Michigan, October—Also Submitted to the *Journal of Applied Psychiatry*.
59. Hochstim, Joseph R.
1967 "A Critical Comparison of Three Strategies of Collecting Data from Households," *Journal of the American Statistical Association*, Vol. 62, No. 319 (September), pp. 976-989.
60. Horvitz, D. G., D. T. Searls and I. Drutman
1966 "Computer Simulation of Hospital Discharges," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 13, Washington, D.C.: U.S. Government Printing Office.
61. Hsieh, N. C.
1970 "Some Estimator Techniques for Utilizing Information from Elements Not in the Sample," Survey Research Center, UCLA.
62. Hyman, Herbert H. et al.
1954 *Interviewing in Social Research*, Chicago: University of Chicago Press.
63. Hyman, Herbert H.
1972 *Secondary Analysis of Sample Surveys: Principles, Procedures and Potentialities*, New York: John Wiley and Sons, Inc.
64. I-Cheng, Chi, L. P. Chow and Rowland V. Rider
1972 "The Randomized Response Techniques as Used in the Taiwan Outcome of Pregnancy Study," *Studies in Family Planning*, A Publication of the Population Council, Vol. 3, No. 11, pp. 265-269.
65. Inderfurth, Gail
1975 "Interviewer-Respondent Interactions in a Telephone Survey of Physician Offices," Paper Presented at the American Association for Public Opinion Research, June.
66. Institute for Survey Research
1975 "Final Report of a Study of Supplementary Security Income Redetermination," Submitted to the Social Security Administration by the Institute for Survey Research, Temple University.
67. Jabine, Thomas and Benjamin Tepping
1973 "Controlling the Quality of Occupation and Industry Data," Presented at the Meeting of the International Statistical Institute in Vienna, August.
68. Johns Hopkins Medical Institution
1975 "Analytic Plan for the National Center for Health Statistics Medical Economic Study," The Health Services Research and Development Center, Unpublished.
69. Kalimo, Esko, Thomas Bice and Marija Novosel

- 1970 "Cross-Cultural Analysis of Selected Emotional Questions from the Cornell Medical Index," *British Journal of Preventive Medicine*, Vol. 24, No. 4 (November), pp. 229-240.
70. Kinder, Donald R. and Leo G. Reeder
1975 "Ethnic Differences in Belief About Control," *Sociometry*, Vol. 38, No. 2 (June), pp. 261-272.
71. Klecka, W. R. and A. Tuchfarber, Jr.
1974 "The Efficacy, Biases, and Problems of Random Digit Dialing," Prepared for Delivery at the Annual Conference of the American Association for Public Opinion Research, Bolton Landing, N.Y., May 31-June 2, 1974.
72. Koch, Gary G.
1973 "An Alternative Approach to Multivariate Response Error Models for Sample Survey Data with Applications to Estimation Involving Subclass Means," *Journal of the American Statistical Association*, Vol. 68, No. 344 (December), pp. 906-913.
73. Krotki, Karol J. and Bonnie Fox
1974 "Randomized Response Technique, the Interview, and the Self-Administered Questionnaire: An Empirical Comparison of Fertility Reports," *Proceedings of the Social Statistics Section, American Statistical Association*, pp. 367-371.
74. Kulley, Andrew
"Evaluation of Experimental Health Program Surveys," Unpublished Paper, Purdue University.
75. Leamer, Edward E.
1974 "False Models and Post-Data Model Construction," *Journal of the American Statistical Association*, Vol. 69, No. 345 (March), pp. 122-131.
76. Lessler, Judith T.
1974 "A Double Sampling Model for Eliminating Measurement Process Bias and Estimating Measurement Error in Surveys," Institute of Statistics, *Mimeo Series No. 949*, Chapel Hill, N.C., University of North Carolina.
77. Locander, William, Seymour Sudman and Norman Bradburn
1974 "An Investigation of Interview Methods, Threat and the Response Distortion," *Proceedings of the Social Statistics Section, American Statistical Association*, Paper to be Published in the June 1976 Edition of the *Journal of the American Statistical Association*.
78. LoSciuto, Leonard A.
1970 "Measures and Correlates of Teenage Fertility—Pretest and Methodological Study," July and December, Unpublished.
79. Loewenstein, Regina, John Colombotos and Jack Elinson
1962 "Interviews Hardest to Obtain in an Urban Health Survey," *Proceedings of the Social Statistics Section, American Statistical Association*, Columbia University School of Public Health, pp. 160-166.
80. Madow, William C.
1963 "Interview Data on Chronic Conditions Compared with Information Derived from Medical Records," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 23, Washington, D.C.: U.S. Government Printing Office.
81. Marks, Eli S., William Seltzer and Karol J. Krotki
1974 *Population Growth Estimation: A Handbook on Vital Statistics Measurement*, New York: Population Council.
82. Marquis, Kent H. and Charles F. Cannell
1963 "Effect of Some Experimental Interviewing Techniques on Reporting in the Health Interview Study," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 41, Washington, D.C.: U.S. Government Printing Office.
83. Marquis, Kent H., Charles F. Cannell and André Laurent
1972 "Reporting for Health Events in Household Interviews: Effects of Reinforcement, Question Length, and Reinterviews," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 45, Washington, D.C.: U.S. Government Printing Office.
84. Matthews, V. L., J. Feather and J. Crawford
1972 World Health Organization International Collaborative Study of Medical Care Utilization: "Response-Record Discrepancy Study."
85. Mooney, H. William
1962 *Methodology in Two California Health Surveys*, Public Health Monograph No. 70, Washington, D.C.: U.S. Government Printing Office.
86. Mooney, H. William, Beatrice R. Pollack and Leslie Corsa
1964 "Uses of Telephone Interviewing to Study Human Reproduction," *Public Health Reports*, Vol. 83, No. 12 (September), pp. 1049-1060.
87. Moulford, Harold
1966 "Identifying Problem Drinkers in a Household Survey," *Vital and Health*

Statistics, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 16, Washington, D.C.: U.S. Government Printing Office.

- 52
88. Nathan, Gad
1972 "Imputation for Partially Missing Cells in Contingency Tables With Known Marginals," Report Prepared for the Office of Education by Research Triangle Institute.
89. Neter, John and Joseph Waksberg
1964 "A Study of Response Errors in Expenditures Data from Household Interviews," *Journal of the American Statistical Association*, Vol. 59, No. 305 (March), pp. 18-55.
90. Nisselson, Harold and Theodore V. Woolsey
1959 "Some Problems of the Household Interview Design for the National Health Survey," *Journal of the American Statistical Association*, Vol. 54, No. 280 (March), pp. 69-87.
91. Page, David and Bernice Radovich
1974 "Fibrocystic Disease of the Female Breast and Its Relationship to Mammary Carcinoma," A Study Started in January 1974 and is Still Ongoing, Contract No. NO1-CB-43871.
92. Phelps, Charles E. and Joseph R. Newhouse
1972 "Effect of Co-Insurance: A Multivariate Analysis," *Social Security Bulletin*, Vol. 35, No. 6 (June), pp. 3-19.
93. Poole, Kenneth W.
1974 "Estimation on the Distribution Function of a Continuous Type Random Variable Through Randomized Response," *Journal of the American Statistical Association*, Vol. 69, No. 348 (December), pp. 1002-1005.
94. Ragan, Paulene
1975 "Response Bias and 'Yea Saying' Among Mexican Americans," Paper Delivered at the Annual Meeting of the Pacific Chapter, American Association for Public Opinion Research, Los Angeles, California.
95. Raghavaro, D. and W. T. Federer
1973 "Application of BIB Designs as an Alternative to the Randomized Response Methods in Survey Sampling," BU-490-M in the Mimeo Series of the Biometrics Unit, Cornell University, Ithaca, New York.
96. Reaser, Joel M., John A. Richards and Steven L. Hartsock
1975 "The Prevalence of Drug Abuse in the Army: A Comparison of Urinalysis and Survey Rates," Technical Report 75-17, Human Resources Research Organization, June.
97. Reaser, Joel M., Steven L. Hartsock and Arthur J. Hoehn
1975 "A Test of the Forced-Alternative Random Response Questionnaire Technique," Technical Report 75-9, Human Resources Research Organization, June.
98. Reeder, Leo G.
Unpublished Paper on Internal-External Control.
99. Riskey, H. W., R. F. Boruch, D. T. Campbell, N. Caplan, T. K. Glennan, J. Pratt, A. Rees and W. Williams
1974 *Social Experimenters: A Method for Planning and Evaluating Social Intervention*, New York: Academic Press.
100. Robbins, Hubert and David O. Siegmund
1974 "Sequential Tests Involving Two Populations," *Journal of the American Statistical Association*, Vol. 69, No. 345 (March), pp. 132-139
101. Roberts, Robert, Ronald Forthöver and Horatio Fabrega
1974 "The Langer Items and Acquiescence," Forthcoming in *Social Science and Medicine*, Presented at the American Sociology Association Meeting in Montreal.
102. Rogers, Theresa F.
1975 "Reinterviewing Urban Residents by Telephone and in Person: Effects on Responses," Paper Presented at the American Association for Public Opinion Research, June.
103. Rotter, J. B., S. Liverant and M. Seeman
1962 "Internal-External Control of Reinforcement: A Major Variable in Behavior Theory," N. Washburne, Editor, *Decisions, Values and Groups*, Vol. 2, London: Pergamon Press.
104. Rotter, J. B.
1966 "Generalized Expectancies for Internal Control Vs. External Control of Reinforcement," *Psychological Monographs*, Vol. 80, No. 690, pp. 1-28.
105. Scach, E., T. Bice, D. Haythorne, L. Kovacis, V. L. Matthews, J. M. Paganini and D. Rabin
1972 World Health Organization International Collaborative Study of Medical Care Utilization: "International Comparisons of Medical Care III: Methodological Results," *Milbank Memorial Fund Quarterly*, September.
106. Schanck, R. L.
1973 "A Study of a Community and Its Groups and Institutions Conceived of as Behaviors of Individuals," *Psychological Monographs*, Vol. 43, No. 2.
107. Scheuren, Frederick, Benjamin Bridges and Beth Kills

- 1973 "Report No. 1: Subsampling the Current Population Survey: 1963 Pilot Link Study," DHEW Publication No. (SSA) 74-11750 (August).
108. Scheuren, Frederick, Beth Kills and H. Locke Oh
1973 "Report No. 2: Coverage Differences, Noninterview and Nonresponse, and the 1960 Census Undercount: 1963 Pilot Link Study," DHEW Publication No. (SSA) 75-11750 (December).
109. Schuman, Howard and Jean M. Converse
1971 "Effects of Black and White Interviewers on Black Responses in 1968," *The Public Opinion Quarterly*, Vol. 35, No. 1 (Spring), pp. 44-68.
110. Scitovsky, Anne A. and Nelda M. Snyder
1972 "Effects of Co-Insurance on Use of Physician Services," *Social Security Bulletin*, Vol. 35, No. 6 (June), pp. 20-38.
111. Sears, David O. and Jonathan L. Freedman
1967 "Selective Exposure to Information: A Critical Review," *The Public Opinion Quarterly*, Vol. XXXI, No. 2 (Summer), pp. 194-213.
112. Sirken, Monroe G.
1970 "Survey Strategies for Estimating Rare Health Conditions," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, pp. 135-144.
113. Sirken, Monroe G.
1975 "Evaluation and Critique of Household Surveys of Substance Use," Chapter IV, *Alcohol and Other Drug Use and Abuse in the State of Michigan*, Office of Substance Abuse Services, Michigan Department of Public Health.
114. Smith, L. L., W. T. Federer and D. Raghavarao
1974 "A Comparison of Three Techniques for Eliciting Answers to Sensitive Questions," *BU-525-M, Mimeo Series, Biometrics Unit*, Cornell University, Ithaca, N.Y., and 1974 *Proceedings of the Social Statistics Section*, American Statistical Association.
115. Steinberg, Joseph
1973 "Some Observations of Linkage of Survey and Administrative Record Data," DHEW Publication No. (SSA) 74-11750.
116. Sudman, Seymour and Robert Ferber
1970 *Experiments in Obtaining Consumer Expenditures in Durable Goods by Recall Procedures*, Urbana, Illinois: Survey Research Laboratory, University of Illinois.
117. Sudman, Seymour and Norman Bradburn
1974 *Responsive Effects in Surveys: A Review and Synthesis*, Chicago, Illinois: Adline Publishing Co.
118. Sudman, Seymour, Wallace Wilson and Robert Ferber
1974 *The Cost-Effectiveness of Using the Diary as an Instrument for Collecting Health Data in Household Surveys*, Urbana, Illinois: Survey Research Laboratory, University of Illinois.
119. Tanur, Judith, Frederick Mosteller, William H. Kruskal, Richard F. Link, Richard S. Pieters and Gerald Rising, Editors
1972 *Statistics: A Guide to the Unknown*, San Francisco: Holden-Day.
120. Terris, Milton
1973 "Desegregating Health Statistics," *American Journal of Public Health*, Vol. 63, No. 6 (June), pp. 477-480.
121. Tisdale, Elizabeth
1975 "Improved Collection of Data on Personal and Confidential Issues by Using Interviewers of Opposite Ethnicity from Respondents in Charlotte, N.C.," Personal Communication to Bernard G. Greenberg.
122. Thornberry, Owen and H. D. Scott
1973 "Methodology of a Health Interview Survey for a Population of One Million," Rhode Island Health Services Research, Inc., Paper Presented at the 101st Annual Meeting of the APHA, San Francisco, California, November.
123. U.S. Bureau of the Census, Division of Health Interview Statistics and Statistical Research Division
1963 "Interview Response on Health Insurance Compared with Insurance Records, U.S. 1960," *Vital and Health Statistics*, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 18, Washington, D.C.: U.S. Government Printing Office.
124. U.S. Bureau of the Census
1963 *The Current Population Survey—A Report on Methodology*, Technical Paper No. 7, Washington, D.C.: U.S. Government Printing Office.
125. U.S. Bureau of the Census
1974 *Index to Survey Methodology Literature*, Technical Paper No. 34, Washington, D.C.: U.S. Government Printing Office.
126. U.S. Government
1974 Privacy Act of 1974, P.L. 93-579.
127. Wallis, W. Allen and Harry V. Roberts
1956 *Statistics, a New Approach*, Glencoe, Illinois, the Free Press, pp. 477-478.
128. Warner, Stanley L.
1965 "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American*

- Statistical Association*, Vol. 60, No. 309 (March), pp. 63-69.
129. Warner, Stanley L.
1971 "The Linear Randomized Response Model," *Journal of the American Statistical Association*, Vol. 66, No. 336 (December), pp. 884-888.
130. Weiss, Carol H.
1968 "Validity of Welfare Mothers' Interview Responses," *The Public Opinion Quarterly*, Vol. XXXII, No. 4 (Winter), pp. 622-633.
131. Wilson, Thurrow R. and Theodore H. Rosen
1975 "Self-Disclosure on Army Surveys: Survey Procedures and Respondent Beliefs Related to Candidness," Technical Report 75-2, Human Resources Research Organization, April.
132. Yngve, Victor H.
1970 "On Getting a Word in Edgewise," Papers from the 6th Regional Meeting of the Chicago Linguistics Society, Editors: M. S. Campbell et al, Chicago: Department of Linguistics, University of Chicago, pp. 567-578.
133. Zill, Nicholas and Roxanne Van Dlusen, Editors
1975 *Basic Background Items for Household Surveys*, New York: Social Science Research Council, October.
134. Papers Presented at the 1975 International Association of Survey Statisticians Session in Warsaw:
- a. A. J. Scott, University of Auckland, Auckland, New Zealand, "Some Comments on the Problem of Randomization."
 - b. R. M. Royal, The Johns Hopkins University, Baltimore, Maryland, U.S.A., "Randomization Balance and Inference."
 - c. M. E. Muller, International Bank for Reconstruction and Development, Washington, D.C., U.S.A., "Challenges in Using Computers in Statistical Applications in Developing Countries."
 - d. M. W. Dickinson, E. K. Foreman, and D. C. Leaver, Australian Bureau of Statistics, Canberra A.C.T., "Why General Survey Systems—Australian GSS. Progress and Prospects."
 - e. E. Outrata and W. M. Podehl, Statistics Canada, Ottawa, Canada, "The Computer as a Control Tool in Survey Data Processing."
 - f. O. Frank, University of Lund, Lund Sweden, "Sampling Population Graphs."
 - g. M. R. Sampford, University of Liverpool, England, "Review Paper on Unequal-Probability Sampling."
 - h. J. N. K. Roa, Carleton University, Ottawa, Canada, "PPS-Sampling."
 - i. G. K. G. Forbrig, University of Rostock, Rostock, G.D.R., "Measurement in Industrial and Labor Statistics."
 - j. Singh, I.A.R.S., New Delhi, India, "Measurement in Surveys."
 - k. J. Lorigny, I.N.S.E.E., O.N.I., Orleans Oedex, France, "Application of Information Theory to Questionnaire Design."
 - l. D. G. Horvitz, Research Triangle Institute, Research Triangle Park, North Carolina and B. G. Greenberg, University of North Carolina, Chapel Hill, North Carolina, U.S.A., "Randomized Response: A Data Gathering Device for Sensitive Questions."
 - m. S. L. Warner, York University, Downsview, Ontario, Canada, "Optimal Randomized Response Models."
 - n. J. Lanke, University of Lund, Lund Sweden, "On the Degree of Protection in Randomized Interviews."
 - o. G. M. K. Kpedekpo, Makerere University, Kampala, Uganda, "Survey Errors, in Demographic Enquiries."
 - p. T. A. Mijares, National Census and Statistics Office, Manila, Philippines, "Survey Errors in Agricultural Statistics."
 - q. H. S. Konijn, Tel Aviv University, Tel Aviv and B. Landman, Central Bureau of Statistics, Jerusalem, Israel, "Generalized Systematic and Simple Random Sampling, with Several Interviews and With or Without Complete or Partial Interviewing."
 - r. J. Sefransk, State University of New York, Buffalo, New York, U.S.A., "Survey Error: Decomposition and Adjustment."
 - s. S. H. Khamis and D. C. Alonzo, FOA, Rome, Italy, "Changes in Methods and Concepts in the 1980 World of Census of Agriculture Programme."
 - t. D. Singh, Institute of Agricultural Research Statistics, New Delhi, India, "Establishment of Sampling Frames to Increase Efficiency and Reliability of Agricultural Censuses."
 - u. M. Lenco, Haute Comite de l'Environnement, Paris, France, "Tabulation and Collection Methods for Special Studies on Structure of Agriculture."
 - v. H. F. Huddleston and W. H. Wigton, Statistical Reporting Service, Department of Agriculture, Washington, D.C., U.S.A., "Uses of Remote Sensing in Sampling for Agricultural Data."
 - w. D. Singh and M. P. Jha, Institute of Agricultural Research Statistics, New Delhi, India, "Statistical Problems in

- Pre-harvest Forecasting of Crop Production."
- x. D. Casley, FAO, Rome, Italy, "Yield and Area Data Under Conditions of Continuous Harvest."
 - y. A. Sattar, UNDP, Khartoum, Sudan, "Counting Livestock of Nomadic Tribes."
 - z. M. C. Fessey and R. D. Lewis, Business Statistics Office, London, England, "Registers for Economic Surveys."
 - aa. D. A. Binder, C. A. Patrick, and A. B. Sunter, Statistics Canada, "On the Editing of Survey Data."
 - bb. S. Kallek, G. Minton, H. Nisselson, and T. Sturdevant, U.S. Bureau of the Census, Washington, D.C., U.S.A., "Evaluation Studies of the 1972 U.S. Economic Censuses."
 - cc. S. K. Gupta and V. R. Rao, Central Statistical Organization, New Delhi, India, "Design Problems of Surveys of Economic Operating Units in the Unorganized Sector."
 - dd. J. P. M. R. Desabie, INSEE, Paris, France, "Dix an d'Experience Francaise d'Enquete Permanente Sur Les Conditions de Vie et Depenses des Menages."
 - ee. E. K. Foreman, Australian Bureau of Statistics, Canberra, A.C.T., "A Survey Design Strategy for a National Statistical Agency."
 - ff. V. R. Rao, Central Statistical Organization, New Delhi, India, "Evolution of a Total Survey Design—The Indian Experience."
 - gg. A. B. Sunter, Statistics Canada, Ottawa, Canada, "A Survey Design Strategy for a National Statistical Agency."
 - hh. E. Rapaport and B. Sundgren, National Central Bureau of Statistics, Stockholm Sweden, "Output Protection in Statistical Data Bases."
 - ii. D. Newman, Office of Population Census and Surveys, London, England, "Rounding and Error Injection in Preserving Confidentiality of Census Data."
 - jj. V. P. Barabba and D. L. Kaplan, Bureau of Census, Washington, D.C., U.S.A., "U.S. Census Bureau Statistical Techniques to Prevent Disclosure—the Right to Privacy Vs. the Need to Know."
 - kk. R. F. Boruch, Northwestern University, Evanston, Illinois, U.S.A., "Reaction to Three Papers on Techniques for Ensuring Confidentiality."

APPENDICES CONFERENCE ON HEALTH SURVEY RESEARCH METHODS

Arrival, Afternoon or Evening, Wednesday
April 30, 1975

Wednesday, April 30
8:30—9:00 p.m.

Remarks by Chairman of the Conference (introduction to the conference, procedures, plans for publication of proceedings), Dr. Leo G. Reeder

55

Thursday, May 1
9:00 a.m.—12:15 p.m.

Health Survey Research Instruments
Session Chairman: Dr. Seymour Sudman
Rapporteur: Dr. Ronald Andersen

1. Respondent burden
2. Length of time required to complete
3. Standardized modules or standardized measures
4. Level of difficulty to use with different groups—
Mr. Robert Fuchsberg

12:15—1:45 p.m.

LUNCH

1:45—5:15 p.m.

Interviewers and Interviewing Techniques
Session Chairman: Dr. Charles Cannell
Rapporteur: Dr. Floyd Fowler

1. Characteristics of interviewers
2. Special issues in telephone surveys
3. Training and supervision of interviewers
4. Research in interviewing techniques: motivation, use of incentives, etc.
5. Indirect techniques (ECHO)—
Dr. Cannell and Dr. Reeder

5:15—6:30 p.m.

BREAK

6:30—8:00 p.m.

DINNER

8:00—9:30 p.m.

Policy Issues and Communication of Research Results
Conference Chairman: Dr. Leo G. Reeder

Friday, May 2
9:00 a.m.—12:15 p.m.

Validity Problems

Session Chairman: Dr. Bernard Greenberg

Rapporteur: Dr. Monroe Sirken

1. Confidentiality and protection of privacy
2. Linkage
 - a. Desirability and uses
 - b. Methodology
 - c. Problems of confidentiality
3. Sensitive questions
 - a. Randomized response
 - b. Contamination methods
 - c. Coding techniques (cryptographic ideas)
 - d. Weighing designs
4. Record checks

12:15—1:30 p.m.

LUNCH

1:30—3:30 p.m.

Total Survey Design

Session Chairman: Dr. Daniel Horvitz

Rapporteur: Dr. Kirk Wolter

1. Costs and estimates
2. Survey research and alternative procedures
3. Panels and longitudinal surveys

**INVITATIONAL CONFERENCE:
HEALTH SERVICES
SURVEY RESEARCH**

April 30,
May 1 & 2, 1975

Airlie House
Airlie, Virginia

- Aday, Lu Ann, Ph.D.
Research Associate
Study Director, Johnson Project
University of Chicago
Center for Health Administration Studies
5720 South Woodlawn
Chicago, Illinois 60637
- Andersen, Ronald M., Ph.D.
Associate Professor & Research Associate
University of Chicago
Center for Health Administration Studies
5720 South Woodlawn
Chicago, Illinois 60637
- Anderson, James G., Ph.D.
Professor of Sociology
Department of Sociology
Purdue University
West Lafayette, Indiana 47907

- Barnes, Mrs. Margaret
Supervisory Statistician
Division of Planning Methods Technology
Bureau of Health Planning & Resource Development
12-14 Parklawn Building
5600 Fishers Lane
Rockville, Maryland 20852

- Berkanovic, Emil, Ph.D.
School of Public Health
University of California
405 Hilgard Avenue
Los Angeles, California 90024

- Boisen, Mr. Morton
Chief, Statistical Methods Division
Bureau of the Census, FOB #3
Washington, D.C. 20233

- Boruch, Robert F., Ph.D.
Associate Professor of Psychology
Northwestern University
Evanston, Illinois 60201

- Bradburn, Mr. Norman M.
Professor & Chairman
Department of Behavioral Science
University of Chicago, Psy. G. 102
Chicago, Illinois 60637

- Cannell, Charles F., Ph.D.
Program Director
Survey Research Center
Institute for Social Research
University of Michigan
Ann Arbor, Michigan 48106

- Carpenter, E. H., Ph.D.
Assistant Professor for Rural Sociology
Department of Agricultural Economics
University of Arizona
Tucson, Arizona 85219

- Carr, Leslie G., Ph.D.
Department of Sociology
University of Akron
Akron, Ohio 44325

- Colombotos, John L., Ph.D.
Associate Professor of Socio-Medical Sciences
Columbia University
Faculty Medicine Room 19 E
School of Public Health
100 Haven Avenue
New York, New York 10032

- Copley, Mr. Archer
Acting Director, Division of Health Services Research
and Analysis
National Center for Health Services Research
15-16 Parklawn Building
5600 Fishers Lane
Rockville, Maryland 20852

- Dalenius, Tore, Ph.D.
Visiting Professor of Statistics
Division for Applied Mathematics
Brown University
Providence, Rhode Island 02912
- Daniel, Wayne W., Ph.D.
Associate Professor
Department of Quantitative Methods
Georgia State University
33 Gilmer Street, S.E.
Atlanta, Georgia 30303
- de la Puente, Mr. Joseph L., Chief
Health Services Research Methods Branch, DHSRA
National Center for Health Services Research
15A-16 Parklawn Building
5600 Fishers Lane
Rockville, Maryland 20852
- Dillard, C. Dianne
Statistician (Health)
Health Services Research Methods Branch, DHSRA
National Center for Health Services Research
15A-16 Parklawn Building
5600 Fishers Lane
Rockville, Maryland 20852
- Dillman, Donald, Ph.D.
Chairman, Department of Rural Sociology
Department of Sociology
Washington State University
Pullman, Washington 99163
- Eckerman, William, Ph.D.
Vice President for Social Research
Research Triangle Institute
Research Triangle Park, North Carolina 27709
- Edmonston, Barry, Ph.D.
Acting Assistant Professor
Food Research Institute
Stanford University
Stanford, California 94305
- Federspiel, Charles F., Ph.D.
Associate Professor & Director
Division of Biostatistics
Vanderbilt University
School of Medicine
Station #17
Nashville, Tennessee 37232
- Fowler, Floyd J., Ph.D.
Director, Survey Research Program
100 Arlington Street
University of Massachusetts
Boston, Massachusetts 02116
- Fuchsberg, Mr. Robert, Director
Division of Health Interview Statistics
National Center for Health Statistics
9-35 Parklawn Building
5600 Fishers Lane
Rockville, Maryland 20852
- Gerson, Mr. Earle J., Chief
Demographic Surveys Division
Bureau of the Census
Washington, D.C. 20233
- Greenberg, Bernard G., Ph.D.
Dean, School of Public Health
Rosenau Hall
University of North Carolina
Chapel Hill, North Carolina 27514
- Haberman, Paul W., M.B.A.
Senior Research Associate
School of Public Health
Columbia University
60 Haven Avenue, Suite B-4
New York, New York 10032
- Hensler, Debbie, Ph.D.
Senior Researcher
Rand Corporation
1700 Main Street
Santa Monica, California 90406
- Horvitz, Daniel G., Ph.D.
Vice President, Research Triangle Institute
P.O. Box 12194
Research Triangle Park, North Carolina 27709
- Hyman, Mrs. Marian
Health Services Research Methods Branch, DHSRA
National Center for Health Services Research
15A-16 Parklawn Building
5600 Fishers Lane
Rockville, Maryland 20852
- Jabine, Mr. Thomas
Chief Mathematical Statistician
Office of Research & Statistics, Room 931-C
Social Security Administration
1875 Connecticut Avenue, N.W.
Washington, D.C. 20009
- Kelly, Mr. Thomas
Alcohol Drug Abuse, & Mental Health Administration
12C-23 Parklawn Building
5600 Fishers Lane
Rockville, Maryland 20852
- Kitching, Mr. William M.
Public Health Advisor
Health Services Research Methods Branch, DHSRA
National Center for Health Services Research
15A-16 Parklawn Building
5600 Fishers Lane
Rockville, Maryland 20852
- Koons, Mr. David A.
U.S. Department of Commerce, SESA
Bureau of the Census
Washington, D.C.
- Kulley, Andrew M., Ph.D.
Department of Sociology
Purdue University
West Lafayette, Indiana 47907

- Lohr, Mr. William
Associate Chief
Health Services Research Methods Branch, DHSRA
National Center for Health Services Research
15A-16 Parklawn Building
5600 Fishers Lane
Rockville, Maryland 20852
- LoSciuto, Leonard A., Ph.D.
Director, Institute for Survey Research
Room 502
1601 North Broad Street
Temple University
Philadelphia, Pennsylvania 19122
- Marquis, Kent, Ph.D.
Senior Researcher
Rand Corporation
1700 Main Street
Santa Monica, California 90406
- Meyers, Mr. Samuel M.
Research Association
Bureau of Social Science Research Incorporated
Suite 700
1990 M. Street N.W.
Washington, D.C. 20036
- Monteiro, Lois, Ph.D.
Associate Professor of Sociology & Community Health
Box 1916
Brown University
Providence, Rhode Island 02912
- Reeder, Leo G., Ph.D.
University of California
School of Public Health
405 Hilgard Avenue
Los Angeles, California 90024
- Roberts, Robert E., Ph.D.
Associate Professor of Sociology & Epidemiology
Department of Psychiatry
University of Texas Medical School
Houston, Texas 77025
- Rosenthal, Gerald, Ph.D.
Director, National Center for Health Services Research
15-05 Parklawn Building
5600 Fishers Lane
Rockville, Maryland 20852
- Scharff, Mr. Jack
Social Security Administration, ORS
DHIS Room 5251
Dickinson Towers
1500 Woodlawn Drive
Baltimore, Maryland 21241
- Schuman, Howard, Ph.D.
Institute for Social Research
Survey Research Center, Room 405
University of Michigan
Ann Arbor, Michigan 48104
- Shapiro, Mr. Sam
Director, Health Services Research & Development
Center
Hampton House Room 182
624 North Broadway
Baltimore, Maryland 21205
- Sirken, Monroe G., Ph.D.
Chief Mathematical Statistician & Statistical Advisor
National Center for Health Statistics
8-30 Parklawn Building
5600 Fishers Lane
Rockville, Maryland 20852
- Sudman, Seymour, Ph.D.
Professor of Business Administration and Sociology
Survey Research Laboratory
University of Illinois
414 David Kinley Hall
Urbana, Illinois 61801
- Waksberg, Mr. Joseph
Vice President
Westat, Incorporated
11600 Nebel Street
Rockville, Maryland 20852
- Walden, Daniel, Ph.D.
Intramural Research Unit
National Center for Health Services Research
15-86 Parklawn Building
5600 Fishers Lane
Rockville, Maryland 20852
- Ware, John E., Jr., Ph.D.
Researcher
Rand Corporation
1700 Main Street
Santa Monica, California 90406
- White, Mr. Elijah
Acting Associate Director for Data Systems
National Center for Health Statistics
9-15 Parklawn Building
5600 Fishers Lane
Rockville, Maryland 20852
- Williams, Mr. Sherman R.
Johns Hopkins University
Health Services R & D Center
Hampton House
624 North Broadway
Baltimore, Maryland 21205
- Wolter, Kirk M., Ph.D.
Research Center for Measurement Methods
U.S. Bureau of the Census
Room 3555, FOB #3
Washington, D.C. 20233
- Woolsey, Mr. Theodore
Westat Incorporated
11600 Nebel Street
Rockville, Maryland 20852