DOCUMENT RESUME

07675 - [C3108198]

Audits & Social Experiments: A Report Prepared for the U.S. General Accounting Office. PAD-79-1; B-193022. October 1978. 75 pp. + 5 appendices (38 pp.).

Report by Robert F. Baruch, et al., Committee on Evaluation Research.

Issue Area: Program Evaluation Systems (2600).
Contact: Program Analysis Div.
Budget Function: General Government (800).
Organization Concerned: Executive Office of the President; Social Science Research Council: Committee on Evaluation Research.
Authority: Congressional Budget Act of 1974, title VII.

Under contract with GAO, the Social Science Research Council conducted a study on methods and techniques for auditing social experiments. Its report focused on the use of reinterviews and other alternatives for monitoring research quality, the general problem of quality assurance, and the roles which GAO might take in meeting its oversight responsibility. Findings/Conclusions: There were not enough instances of the use of the reinterview to predict its impact on research participants, but there seems to be a significant risk of its disrupting research. Possible alternatives to reinterviews are the use of: information on research procedures to determine quality, parallel sampling by auditors, surrogate auditors, subsample for reinterviews by auditors, and record-linking and other statistical techniques. GAO's possible contributions in social experiments at various stages of the research and evaluation process were identified. The potential costs and benefits of audits were considered as applied to researchers, sponsors, and participants in social experiments. The roles of sponsors were discussed with regard to developing and implementing solicitations, the monitoring of projects, and the dissemination of the project's results. Responsibilities of the researcher were considered in terms of assuring the quality of research. Recommendations: GAO should recognize the following: unplanned research, such as reinterview, can disrupt research; a reinterview is often unnecessary, and alternatives can generate statistical information on the quality of data; some projects will require that GAO become involved early in the research process; and a small scale effort is more justified than a major in-house testing program to assay the effects of reinterview. GAO should prepare a document which clarifies its likely roles in evaluation of social experiments, and its oversight role should be coordinated with other groups responsible for assuring quality. GAO's efforts to report both deficiencies and proficiencies of social experiments were endorsed. GAO should take an active role in helping to develop guidelines on the management and budgeting of social experiments and of program

evaluation. (ETW)

8198

Audits & Social Experiments:

A Report Prepared for The U.S. General Accounting Office

by the Committee on Evaluation Research
Social Science Research Council

PAD-79-1

October 1978

COMPTROLLER GENERAL OF THE UNITED STATES
WASHINGTON, D.C. 20548

B-193022

The review and evaluation of programs authorized by Federal legislation is a primary mission of the General Accounting Office. This role is emphasized in recent legislation, particularly Title VII of the Congressional Budget Act of 1974. It is also included in several of the more recent "sunset" proposals, which have been under consideration in the Congress. It is further emphasized in many of the specific mandates and requests from the Congress or its committees, which require GAO to evaluate particular programs and activities, including a number of social programs.

In addition to its own evaluations, GAO is often involved in assessing evaluations performed by executive agencies or their contractors. Some of these assessments have involved GAO in auditing and reanalyzing the results of social experiments where the results were expected to affect policy regarding major proposed social programs. While few appear to question the need for such audit and reanalysis, there has been considerable concern among social researchers about the possible impact of audit on their research. There has also been concern about the possible violation of pledges of confidentiality made to individual participants if GAO exercises its right of access to research data.

GAO has performed extensive studies of the issues involved in balancing privacy and accountability, including a thorough review of its own practices and policies. To assist us in this effort we requested the Social Science Research Council (SSRC) to undertake a study and to recommend appropriate methods and techniques for the audit of social experiments.

The enclosed report by the SSRC is, we believe, an excellent discussion of the concerns of researchers, and contains useful recommendations to GAO. The recommendations are being studied in the context of GAO's experience, and the degree to which a recommendation is implemented will be based upon a number of considerations.

We believe this report may also be of use to you. Over the coming months, GAO will be publishing various guideline documents based upon its own experience and studies, which will expand on the guidance contained in the SSRC report. These will include a checklist for assessing social program impact evaluations and guidelines for audit and reanalysis of social research for policy.

Sincerely yours,

Comptroller General
of the United States

AUDITS AND SOCIAL EXPERIMENTS

A REPORT PREPARED FOR

THE U.S. GENERAL ACCOUNTING OFFICE

Committee on Evaluation Research

Social Science Research Council

July 31 1978

Robert F. Boruch
(Chairman)
Howard E. Freeman
Edward Gramlich
Richard Light
Peter H. Rossi
Joseph Steinberg
Harold W. Watts

Ronald P. Abeles
(Staff)

TABLE OF CONTENTS

AUDITS AND SOCIAL EXPERIMENTS

Committee on Evaluation Research
Social Science Research Council

SUMMARY

The purpose of this report is to assist the U.S. General Accounting Office (GAO) in the "development of methods and techniques for auditing social experiments." It is submitted in accordance with the contract between the Social Science Research Council and the GAO (Contract No. 7130078, March 1977, as amended on March 10, 1977 and January 16, 1978). The report is divided into six sections. The first two sections focus narrowly on the use of reinterviews and other alternatives for monitoring research quality. The next four consider the problem of assuring quality more generally, and the roles which GAO might take in meeting its oversight responsibility.

Evidence on the Impact of Reinterview

The committee concludes that there have been insufficient instances of audits by the GAO or other governmental agencies involving reinterviews to permit confident prediction of the impact of reinterview on research participants. The available evidence suggests that there is a notable risk of disrupting the research; but the evidence is ambiguous as to the magnitude of the risk and as to consequences of any disruption.

Alternatives to Reinterviews

The need for and benefits of reinterviews by GAO should be considered as marginal. Well-designed social experiments will have built-in checks on the validity and reliability of measures, and the quality of their data can be assayed, up to a point, without reinterviews. Reinterviews by an outside agency may be limited in value since they yield data that are hard to interpret. That is, discrepancies between the researcher's and the auditor's results cannot be easily ascribed to deficiencies in the researcher's design or procedures.

Possible alternatives to reinterviews are the use of: information on research procedures to determine quality; parallel sampling by auditors; surrogate auditors; subsample for reinterviews by auditors; and record-linking and other statistical techniques. The costs, benefits, and appropriateness of alternatives are discussed.

GAO Role in Social Experiments

Various opportunities which the GAO has to improve or monitor social experiments and their evaluations are considered at each stage of the research and evaluation process. While believing in the benefits of periodic early involvement by the GAO in review of social experiments, the committee does not propose that the GAO participate actively in every stage of the process. Rather, it attempts to lay out a field of inquiry and to identify where the GAO can conceivably make distinctive contributions.

In addition, the general style of GAO audits and the perception of that style by sponsors and researchers are discussed. The committee endorses the GAO's policy and efforts towards the identification of (1) solutions to problems raised in audits and of (2) the "successes" as well

as the deficiencies of particular projects, and the committee emphasizes the need for even greater efforts in these directions. The GAO's efforts to develop a more diversified and qualified professional staff (in regard to auditing of social experiments) are recognized and encouraged. The need to coordinate any role with other monitors of research, in order to avoid delaying research or increasing research costs unnecessarily, is recognized and coordination encouraged.

Some Costs and Benefits of Audits

Potential costs and benefits of GAO audits are considered as they may accrue to researchers, sponsors, and participants in social experiments. For the researcher, the potential costs consist mainly of increased demands on his time and manpower resources, while the possible benefits are better documentation of the experiment and suggestions for improving either the current or future experiments. For the sponsor, potential costs consist of the expenditure of financial as well as manpower resources, and, again, the potential benefits are improving future endeavors. Finally, for the participant the possible costs are another intrusion on his time and privacy, in the case of reinterviews. The long term benefits apply to the research participants as citizen and taxpayer.

Costs and benefits to Congress or other groups are not considered because they lie beyond the scope of this report.

Role and Responsibilities of Sponsors and Researchers

Whether the GAO will need to reinterview respondents and what roles the GAO could most effectively take in overseeing evaluations of social

experiments depend partly on the normal roles and responsibilities of researchers and their sponsors. The roles of sponsors are discussed in regard to developing and implementing "Requests for Proposals," the monitoring of projects, and the dissemination of the project's results. The responsibilities of the researcher are considered in terms of assuring the quality of the research: the specification of objectives and development of measures; quality control during field operations; pilot tests; and checking the quality of data processing. The section closes with a call for the wider dissemination of reports and the availability of public-use data tapes for secondary analyses.

Recommendations: Brief Summary

The first set of recommendations, designated with the prefix A, covers GAO reinterviews as a device for assessing quality of data in social experiments. The second set (B), more important in the long term, concerns GAO roles in monitoring quality generally.

A.1. Unplanned interventions in research, such as reinterview, can disrupt the research and introduce biases which are difficult or impossible to estimate. The evidence available is sufficient only to make the risk clearly plausible; it is not sufficient to gauge magnitude of risk or severity of consequences.

A.2. Reinterview of respondents is often unnecessary. Other evidence is generally sufficient to identify quality of data and of research, and these do not engender a serious risk of affecting research participants. There are exceptional cases in which direct reinterview by GAO staff are warranted. Reinterviews will at times yield results which are difficult or impossible to interpret.

A.3. GAO should recognize in its oversight policy and pertinent manuals and guidelines that alternatives to direct reinterview of research participants by GAO generate statistical information on the quality of data. Aside from using information about research procedures, the alternatives include parallel sampling, subsampling from an augmented main sample in an experiment, using surrogate auditors, and other tactics, such as mutually insulated file linkages.

A.4. Some projects will require that GAO become involved early in the research process. This means that an executive agency must be prepared to seek GAO advice where it is clear that GAO has pertinent experience and expertise. This is especially pertinent for the occasional project in which reinterview is necessary, for GAO reinterview may require that the sample necessary for a social experiment be augmented.

A.5. GAO should not mount a major in-house testing program to assay the effects of reinterview. A small scale, well designed effort, which capitalizes on the assistance of agencies with substantial experience and expertise in methodological studies is much better justified at this point. This may include, for example, case studies, archives of audit experiences, and field experiments on the effects of reinterviews and of audits in general or social experiments.

B.6. The GAO should prepare a document which clarifies its possible and likely roles in the evaluation of social experiments, and which specifies GAO intentions. This document should be distributed widely to researchers and research sponsors.

B.7. Any role which GAO chooses to take in oversight of social experiments, at any stage of the research process, should be coordinated with other groups with responsibility for assuring quality. This includes, for example, the Office of Management and Budget, the Statistical Policy Division of the U.S. Commerce Department, agency monitors of projects and their advisory boards and clearance boards. Coordination is crucial to avoid unnecessary redundancy and delays in research.

B.8. The committee endorses the GAO's policy and efforts to report both deficiencies and proficiencies of social experiments. Because formal scientific tests of social programs are a relatively new undertaking, even greater emphasis should be put on establishing and documenting solutions to problems in GAO reports, manuals, and policy.

B.9. The committee recognizes and endorses the GAO's efforts to diversify and develop staff with experience and expertise in the review of social experiments. Accelerated development of staff and augmentation of staff in this area is essential for effective review, and the committee encourages acceleration.

In order to sustain this increased expertise and to assure that it is recognized by the scientific community, the committee believes that those GAO staff with the direct operational responsibility for the review of experiments should participate even more

vigorously in scientific forums and professional discussions on social experimentation. The committee recognizes that some staff are already active in this regard and it endorses and encourages that activity.

B.10. The GAO should take an active role in helping to develop guidelines on the management and budgeting of social experiments and of program evaluation. This area of inquiry has not been well explored, and the GAO is in a remarkably good position to exercise a leadership role. Specific suggestions about the ways in which these recommendations can be implemented are described in the text.

The purpose of this report is to assist the U.S. General Accounting Office (GAO) in the "development of methods and techniques for auditing social experiments." It has been prepared in accordance with the contract between the Social Science Research Council and the GAO (Contract No. 7130078, March 1977, as amended March 10, 1977 and January 16, 1978). Appendix I relates sections of the report to specific contractual obligations.

The first two sections of the report are concerned with a relatively narrow issue: GAO interest in reinterview of research participants as a device for gauging the quality of a social experiment. GAO's interest is consistent with its mandate to oversee Federal agency evaluations. However, both GAO staff and researchers have been concerned that reinterviewing may engender a breach of assurances of confidentiality made to research participants and, more generally, may disrupt research; the two sections examine that concern. The remaining four sections of the report are more general and, in the committee's view, likely to be more important in the long run. They concern some ways in which GAO can meet its responsibilities for oversight of social experiments and they cover the interest shared by GAO and the larger research community in assuring quality of policy related social research in general.

In Section I, the committee offers its judgments on the evidence concerning both the benign and disruptive effects of reinterviews on social experiments. In Section II, alternatives to direct GAO reinterviews of respondents are suggested as devices for evaluating the quality of data emerging from social experiments. Section III considers the role of the GAO at various stages in the life cycle of social experiments,

primarily to help understand how GAO can contribute to monitoring quality of evaluations in general. The committee weighs some costs and benefits of audits for the sponsors, researchers, and participants in social experiments in Section IV. Some major role responsibilities of the sponsors and researchers are specified in relation to good research practice and to audits in Section V. The committee's report concludes, in Section VI, with recommendations to the GAO regarding its role in social experiments.

SECTION I

EVIDENCE ON THE IMPACT OF GAO REINTERVIEWS ON RESEARCH PARTICIPANTS

To avoid needless ambiguities, we begin with a few working definitions, and then examine the evidence on the effects of reinterviews.

Preliminary Definitions. A social experiment is defined here as a planned effort to introduce a new social program and to assess its effects scientifically. This includes, for example, formal randomized experiments: individuals are assigned randomly to one of two or more program variations, for the sake of obtaining a fair comparison of costs and benefits of each variation. The Demand Experiment component of the Experimental Housing Allowance Program (EHAP) illustrates the type. The definition also includes so-called quasi-experiments in which an innovation is introduced and estimates of its effects are developed in planned comparison against historical standards, as in time series analyses, or against some comparison group. The terminology here is consistent with contemporary texts on assessing the effects of innovative social programs or program components, and with GAO reports on the topic. See U.S. General Accounting Office (1975), Riecken, et al (1974), and Appendix II of this report for other references.

A survey is defined as a scientifically planned effort to elicit information from a sample of a population. The individuals who provide the information are, in the following remarks, referred to as research participants or respondents.

The distinction between social experiments and surveys is crucial. Experimentation invariably demands the use of control conditions or comparison groups against which the value of program variations can be judged. Moreover, the group's composition must be such that judgments about program effect are as unequivocal as possible, i.e., not subject to a variety of competing explanations. The rationale underlies development, over the past 50 years, of design of experiments.

Surveys, on the other hand, have traditionally been used as a descriptive device. The Current Population Surveys run by the U.S. Census Bureau, for example, are justified for the information they furnish about character and change in the population, rather than on the grounds that they yield unambiguous information about the effects of new social programs. Some social experiments do rely heavily, of course, on sample surveys. The Vera Institute's experimental tests of programs for ex-addicts, for instance, involves interviewing members of both the participants in the novel program and members of a control group. But, many experiments can rely solely on existing archives, reporting systems, rather than on surveys to furnish the necessary data. In principle, the statistical criteria used to review a sample survey also pertain to reviewing a survey component of a social experiment. Review of a social experiment implies other criteria as well, since the experiment includes features such as control groups and a program which themselves are subject to review.

The phrase program evaluation is also used in this report. The committee understands this to mean efforts to appraise one or more of the major features of a new or existing program: inputs, including fiscal and manpower resources; processes, including management and other operational features of the program; outcomes; and operational settings and constraints on program performance (U.S. General Accounting Office, 1975). An innovative program may be evaluated -- via a social experiment -- as to its effects on participants. But contemporary use of the term program evaluation does not necessarily imply a social experiment. All the data collection issues considered in this report, however, have a bearing on program evaluations more generally. The stress on social experimentation stems from the origins of this report, notably discussion of the Experimental Housing Allowance Program.

The main topics discussed here concern reinterview of research participants and audit of social experiments. By reinterview, the Committee means that a research participant, having furnished information at least once to the researcher, is asked by a GAO interviewer to supply information. One purpose of the reinterview, according to GAO, is to check on the quality of data obtained in the experiment. GAO audit or review is defined as an examination, by GAO, of one or more of the following: the plans, procedures, conduct and management, and/or results of an experiment. Reinterview, if undertaken at all, then forms a small part of a larger enterprise. The broad definition of audit is consistent with GAO's current manuals and documents on the topic (e.g., U.S. General Accounting Office, 1974, 1975, 1977; Havens, 1977).

GAO Reinterviews in the Experimental Housing Allowance Program

One of the GAO's specific charges to the Council was to survey what is known from past experience about the effects of audits on data quality and conduct of field research. It is the committee's understanding that the GAO is particularly interested in the effect of reinterviews on research participants. On this matter, the committee has two principal observations:

(1) There has been an insufficient number of audits by the GAO or other government agencies involving the reinterview of research participants to enable the committee to reach a firm judgment on the impact of reinterview on respondents.

(2) The available indirect evidence suggests that there is a clear risk of disruption of research. However, the available evidence is ambiguous as to the magnitude of the risk and as to the consequences of a disruption should one occur.

Each of these observations is discussed fully below. The idea that auditor reinterviews may be unnecessary, regardless of their effects, is critical and is discussed in Section II.

GAO and SSRC staff reviewed and discussed recent GAO audits of social programs and experiments and corresponded with state audit agencies. The staff found that in only one instance had participants in social research been interviewed first by a research team and then reinterviewed by the GAO. This single example, the Experimental Housing Allowance Program (EHAP), does not lend itself well to analysis or interpretation of the potential impact of reinterviews primarily because it was designed for other purposes (U.S. General Accounting Office, 1968).

Only a small number of respondents were reinterviewed, for example, and the activity was undertaken after the researchers had completed collection of the experimental data. The timing of the reinterviews meant that no meaningful subsequent data are available that could be used to assess the impact of the reinterviews on the experiment. The relatively small sample size of the reinterviews makes it difficult to establish or anticipate the value of any statistical analyses based on those data. Further, the self-selective nature of the reinterview sample and the level of the participation rate (from 60 to 80 percent) make it difficult to generalize beyond the data obtained from the sample. One implication of this participation rate, for example, could be that similar cooperation rates might be obtained by the GAO in the future. However, only limited information is available on the respondents and nonrespondents. Their attitudes towards and knowledge about the reinterview process and its auspices are for the most part unknown. Consequently, it is difficult to reach any conclusions about the implications of these reinterviews for the participation rates in future GAO audits, let alone for anticipating the impact or lack of impact of the reinterview on EHAP.

This GAO follow-up study was, on the other hand, informative for the limited purpose for which it was designed. Before the study was undertaken, virtually nothing was known about the rate at which research participants would consent to a follow-up by a certified public accounting firm acting on GAO's behalf. The rates found at each site ranged between 58 and 80%, far higher than a few experts thought possible. That a rate of 60% is likely to be inadequate for some GAO purposes is also clear. That 80% rates can be improved is not an unreasonable view since other common devices for

improving response rate could be exploited. Finally, the process of eliciting consent for reinterview appears to have involved no remarkable problems in execution.

Indirect Evidence on the Impact of Auditor's Reinterview

Some indirect evidence on the potential impact of audits is available from studies of the influence of confidentiality pledges on respondents' willingness to cooperate in research. In these case studies, experiments, and surveys, willingness is typically defined in terms of either refusals to respond (i.e., refusals to answer any questions at all or some particular questions) or refusals to respond truthfully. The evidence is only indirectly relevant to GAO audits, since none of the cases, surveys, or experiments involved the GAO.

However, the evidence is pertinent in the sense that it concerns willingness to respond when confidentiality of research records is put into conflict with administrative interest in accessing those records. By "administrative interest" is meant that an individual's record, which contains both identification and substantive information collected by the researcher, may be used to make personal administrative decisions about the individual. For example, data collected by researchers might be appropriated and used by an investigatory agency in deciding whether welfare payments should or should not be made to particular individuals. In this case, the appropriation of individually identified research records for non-research purposes will run counter to the researcher's assurances to the respondents and may have harmful consequences for the experiment.

References to case studies, experiments, and surveys described here are given in Appendix III.

Illustrative Case Studies. At times, research has been disrupted because of administrative threats to the confidentiality of research records. In the American Council on Education's studies of Campus Unrest, for example, confidentiality was promised by the original researchers. However, threats of subpoena by the McClellan Commitee and other groups (e.g., local police) provoked distrust of the researchers' assurances and resulted in three or four study sites (i.e., colleges) dropping out of the study and considerable difficulty in negotiating with student groups and funding agencies over collection of data. Similarly, Spergel's evaluation of the Woodlawn manpower training program was disrupted because trainees thought Spergel's assurances were breached by Congressional investigators. The disruptive effects of unplanned administrative intervention in research projects is more dramatic in drug research. Testimony by Congressmen as well as researchers emphasized the need for confidentiality assurances to research participants, and cited specific disruptions attributable to governmental attempts to appropriate research records. That just the fear of investigative agencies appropriating research data, regardless of whether the concern is warranted, can be damaging is evident in Norwegian and Swedish research on adolescent development; Norway, for example, bowed out of a multinational study partly on those grounds. On the other hand, the impact of Congressional investigations of the Negative Income Tax experiments and of disclosures of respondents' records despite confidentiality assurances appears to have been minimal.

Illustrative Experiments. A variety of field experiments have been conducted in order to determine if strong confidentiality assurances elicit more cooperation than weak or no confidentiality assurances. For example,

so-called randomized response methods have been developed to assure that there is no way to link individuals to their answers to particular questions, even in interview situations. Some large experimental tests of this method suggest that, in comparison with normal interview procedures, respondents are more likely to report accurately on such sensitive topics as abortion, racial attitudes, drug abuse, etc. However, the method has not been shown to be remarkably superior in other studies dealing with convictions for drunk driving, bankruptcy, and other topics.

Recent large-scale experiments conducted by the U.S. Bureau of the Census (under the auspices of the National Academy of Sciences) and by Eleanor Singer at the National Opinion Research Center have investigated the consequences of different types of pledges of confidentiality. In one component of the Census study, five different pledges were given to different samples: (a) data would be kept confidential forever, (b) 75 years, (c) 25 years, (d) no promise of confidentiality at all, and (e) an explicit statement that confidentiality could not be guaranteed. The overall pattern of refusal rates increased monotonically across each condition, and the pattern is statistically significant: the weaker the confidentiality assurance, the lower the cooperation rate. Respondents in this experiment also indicated that they believed the national government could be trusted to keep information confidential more than universities, state governments, or private companies. Interestingly, a sizeable minority could not recollect whether promises of confidentiality had been made, even just after being interviewed.

Singer's work demonstrates, among other things, that requiring a signature of the respondent as a device for securing informed consent has a notable decrease in the cooperation rate. Such a signature might

be required by some agencies to meet criteria of Human Subjects Review Committees for ethical experimentation, e.g., disclosure of research records to an agency.

Illustrative Surveys. Survey research on this topic has focused on (a) cooperation rates as a function of respondent sensitivity to privacy issues and on (b) public attitudes towards surveys. Regarding the former, some experts have argued that nonresponse rates in large-scale surveys have increased over the past ten years and that the increase is the result of respondents' fears about invasion of "privacy." However, the evidence for a decline is weak and that for the cause of the decline is ambiguous.

Regarding the latter, public attitudes generally indicate a strong concern for confidentiality in surveys. Yet, there is still little evidence that these attitudes influence behavior. Some studies suggest that respondents do take into account who the sponsors of the survey are, which presumably reflects their trust in the confidentiality of the data. For example, servicemen in at least one major study were less inclined to report drug abuse to military surveyors than to civilian researchers.

Conclusion. These illustrations indicate that predicting the consequences of a breach in confidentiality assurances is not always possible on the basis of available data. Some research has been disrupted by breaches, while other studies have not. However, the evidence is sufficient to lead the committee to expect problems such as lower response rates and less truthful responses, should confidentiality assurances be absent or be breached and should the breach become generally known or even generally believed. The existing data are not sufficient to predict the severity of the problems.

SECTION II

ALTERNATIVES TO REINTERVIEWS[1]

This chapter of the Committee's report deals generally with problems of assessing the quality of data generated in a social experiment, and with specific problems of and alternatives to GAO reinterviews as a device for assessing quality. GAO's particular interests, described in the GAO (1977) Background Paper for this Committee, are considered in each subsection. Those interests include: reinterview for the sake of verifying "...that the subject selection procedures have been carried out..." and "...to verify that...variables are correctly recorded...".

The Committee first considers the purposes and worth of reinterviews, and their potential problems. Succeeding sections discuss the types of information which are most likely to be of interest in reinterviews, and alternatives to reinterviews which can be used to gauge the quality of data.

Reinterviews: Purposes and Problems

For the auditor with an interest in establishing the quality of interview data, wholesale reinterviews may appear to be a natural option. The tactic has some scientific merit in that independent observations of the same phenomena are generally desirable. It has some

---

[1] This section is adapted from Boruch and Cecil (1978), with some modification by the committee.

institutional merit, too, since an uninformed public may believe that reinterviews are the only acceptable device for assessing the quality of the original (researcher's) interviews. However, the committee believes that establishing quality is basically a statistical problem, i.e., obtaining summary measures of quality, and not producing anecdotal information about particular identifiable individuals.

The benefit of wholesale reinterviews will, for a variety of reasons, be marginal. First, the well-designed study will have included a side study on the validity and reliability of its measures. For example, the veracity of responses to factual questions may be compared to already existing archival data. In such cases, reinterviews as validity or reliability checks add little, if anything, to preexisting checks. The needs of an audit agency could just as well be satisfied by participating in the design of the original side-studies, which would avoid direct contact between the auditors and respondents. This does require that contractor or grantee, research sponsor, and GAO collaborate in design of such studies.

Second, any major study will generate a file of information whose quality can be assayed -- up to a point -- without reinterviews. The researcher's procedure for editing records, internal checks on consistency of information provided by respondents, and comparisons of the researcher's statistical data with similar data from prior studies can be reviewed by auditors without reinterviews. Indeed, "verifying that subject selection procedures have been carried out according to the experimental design" (GAO (1977) Background Paper) clearly depends in a fundamental way on the project documentation which prescribes the procedure and quality control devices, and on the on-site observation of project staff.

Third, and most importantly, reinterviews by outside auditors can yield ambiguous results. Ordinary measurement error will, for example, typically produce differences between data yielded by the interview and the reinterview. Differences in their results may also reflect actual changes in respondents that occur in the time between the two interviews. Respondents, for example, will have had time to think about a topic at issue, even a factual matter such as income, and may change their judgement about it or their interpretation of the question by the time the reinterview is conducted. This is not an uncommon finding in panel studies such as the Current Population Survey, for "factual" matters such as employment experience, income, and so on, as well as for opinion surveys (see for example, Robert Ferber's The Reliability of Consumer Reports of Financial Assets and Debts, and other monographs cited in References.) Furthermore, discrepancies may appear as the result of differences in the interview procedures and skills employed by the researchers and the auditors. Finally, reinterviews may produce discrepant results merely because they are carried out under different auspices than the original interviews. In particular, a knowledgeable group of research participants will be able to distinguish between GAO and other agencies conducting research. To the extent that they do, and to the extent that they regard GAO as an investigative agency, differences in response candor and cooperation are likely to appear.

In short, reinterviews by auditors will often be unnecessary or will produce ambiguous or irrelevant data.

Despite this, the Committee believes assessment by audit agencies is warranted. Not all evaluation research projects are good, and the over-

sight by an audit agency may help to avoid or to detect gross incompetence. Accepting the idea that auditor scrutiny is justified despite its limited scientific benefits does not imply that GAO reinterviews are warranted. In fact, it is not unreasonable to expect an auditor's reinterview of research respondents to have negative effects. For, with a few exceptions, commercial and governmental auditors are not trained to handle surveys nor do they often have the manpower at hand to do the job well. Moreover, even if the skills are available, the respondent may view the auditor-sponsored survey as threatening or at least as less innocuous than the original research interview. As a consequence of the reinterview, the respondent may then refuse to cooperate further in the research or may change the style of response (e.g., become less candid) in subsequent contacts with the researcher. This is the worst possible outcome. The committee expects problems to be much less crucial to the extent that the interview skills of the audit agency are well-developed, that the audit agency is viewed as a professional research group rather than as an administrative agency (with proper regard for statistical uses of data rather than punitive administrative uses), and that the rewards for participation in research override the costs associated with being reinterviewed by an audit agency.

The evidence on whether auditor contact with research participants poses some risk to the research project is indirect and certainly not uniform (see earlier remarks). None of the studies that the committee has examined, other than EHAP, involved reinterviews by a governmental auditor. However, the evidence is sufficient to indicate that research

participants can be sensitive to third party interrogation of research records and, in some instances, become less cooperative in further research. Consequently, it should be recognized that unplanned third party intervention (e.g., by auditors) can be a hazard and that strategies minimizing that risk should be developed.

Classifying Information and Prior Agreements

It is not possible to reconcile audit objectives with social research aims without specifying the kinds of information which might be of interest to an audit agency. Research and audit agencies should develop methods for identifying and collecting information of potential interest to auditors as early as possible. Delaying this process until the research is in progress can disrupt the project's functioning, as was the case in the Negative Income Tax Experiment.

Whether an element of information can be or should be regarded as disclosable to an audit agency depends partly on the character of the research. For example, if an experimental program is viewed as a strict prototype for a real social program, then it can be argued that the rules governing access to records must also be prototypical. Any information that would be normally collected by an auditor as an auditor and not as a professional researcher in the anticipated regular program would then be disclosable in the experiment. Any information collected solely for purposes of statistical (i.e., experimental) analysis would not be disclosable. Following this line of argument (and ignoring the question of prior pledges of confidentiality), the identification of participants in EHAP should be made available, since identification of housing allowance recipients would be a matter of public record, if and when the program was

adopted. Data on an identifiable individual's eligibility for services or subsidy would be similarly regarded as accessible to an auditor. This position implies that GAO reinterview is permissible, so long as research participants are informed of the possibility, and so long as the reinterview topics are limited. That is, the reinterview would be dedicated to verifying that in fact: the individual, who was supposed to have been interviewed, was interviewed; the individual has general characteristics which he is supposed to have as a participant in the program. However, other information which would not be ordinarily available or accessible in a fully implemented, real program, such as respondents' attitudes, would not be accessible to auditors of the experimental program.

Whenever the research project or experimental program cannot be regarded as a strict prototype of a program, then a second basis for making decisions about accessibility of records for audits should be employed. For a variety of reasons, the research project may be viewed as a partial prototype -- a fragile pilot effort that is necessarily under extraordinary control by the researcher. The control may be essential for careful surveillance of the phenomena under study or for the detection of subtle effects of the experimental program. Any interference which would jeopardize the integrity of control or examination could destroy the total effect under study. The level of control required here certainly detracts from the realism of the prototype, but it does make the program's performance simpler to track and evaluate. It provides useful information to be employed in the development of more "realistic" programs.

In such instances, then, the research may require that even the identity of program participants be kept confidential with respect to an audit agency. In the case of EHAP, for example, the identification of a housing allowance recipient may technically be a public matter or, at least, legally accessible to an auditor. Yet, as a practical matter, publicity about the identification or disclosure of participants to an auditor may present a severe shock to the research. If the research is important enough to justify considerable expenditures in its support and if the probability of a shock is high, then disclosures would be unjustified.

The priority attached to the research goals here does not mean that audit goals are to be ignored or abrogated. It does mean that auditors and researchers should reach agreement as early as possible in the research process in order to minimize dangers to the research, while maximizing the efficiency and effectiveness of the audit.

Where it is reasonable to reach prior decisions, the negotiation should be deliberately planned so as to be as efficient as possible. The current institutional barriers to mounting quickly policy research (e.g., clearance processes) are sufficiently time consuming and costly to justify considerable attention to reducing any additional activity which may increase those costs or time delays.

Alternatives to Direct GAO Reinterviews

While the early identification of the information needs of auditors may help avoid conflicts between auditors and researchers, it will not always assure smooth collaboration between them. Researchers may legitimately continue to object to reinterviews as endangering the

viability of their experiments. In some instances, it may be impossible to establish an empirical basis for this objection, since the respondents' behaviors or attitudes toward reinterviews cannot be well predicted. In other research, such as of drug and alcohol abuse, the target population will be understood well enough to predict degraded cooperation as a function of an audit agency's reinterviews. Consequently, auditors and researchers need to consider alternatives to reinterviewing respondents. The strategies suggested by the committee are: Review of research procedures, parallel sampling, use of surrogate auditors, subsampling for reinterviews, and other options.

Review of Research Procedures. It is clear that poor research data will be generated by poor research practices, in social experiments and surveys. It is also clear that an examination of the research procedures is a sufficient basis for determining whether data will be poor for the large majority of such research projects. Further, critical examination of the research methods and procedures generally involves no serious privacy problems and can be undertaken without serious risk of disrupting research in the field. It is for these reasons that the committee recommends that the GAO exploit the opportunity to examine research procedures before reinterviews are considered. Such an examination, in any event, is likely to be necessary to fulfill oversight responsibility.

The pertinent research procedures are described in Section V on roles and responsibilities of the researcher. Briefly, GAO's review should focus on the researcher's sample design and sampling procedures, on the survey process including interviewing, and on data processing

procedures, including quality control. The activities involved in each of these determine, in large measure, the quality of resulting data. Sampling frame and procedures for sampling can be verified and adherence to plans monitored by GAO. Novel interview forms or methods must often be pilot tested beforehand, and both plans and practice for them can be monitored or checked by GAO. Interviewer training methods and supervisory checks on their performance should be explicit and both plans and practice can be reviewed by GAO without contact with actual research participants. Data processing procedures should be explicit and both plans and activity can also be reviewed for their quality. Side studies run in the field by the researcher are necessary at times, and their plans, conduct, and results can be reviewed to provide direct statistical evidence on quality of data. In fact, it is reasonable to expect that side studies can be designed to conform to GAO guidelines or experience in assuring that evidence on quality is available.

That the GAO can identify weaknesses in the quality of research and research data by examining these procedures is clear from GAO practice, e.g., GAO's review of the Federal Aviation Administration's surveys of the impact of the Concorde Airplane on certain communities (U.S. GAO, 1977). In principle, GAO can assist in researcher and agency efforts to encourage good practice in designing, implementing, and controlling quality of these research procedures.

Parallel Sampling by Auditors. When a main objective of the audit agency is to establish the accuracy of sampling and responses, then a reasonable strategy is to obtain an independent sample of the same target population used in the original survey. For example, GAO might

have adopted the same sampling design and target population used by HUD in its original EHAP interview surveys in order to generate an independent, equally valid, and non-overlapping sample of respondents. Comparing estimates of population parameters from the GAO survey with those from the HUD survey would supply an indication of the quality of estimates generated in the original research. The differences in auspices of the interview and other factors described earlier may, of course, introduce differences in the research results. To reduce the risks of "contamination" of one agency's survey by the other agency's survey, each might be undertaken in a different geographic areas (e.g., census tracts or larger regions).

The main product of this strategy is a legitimate statistical basis for judging the quality of the initial survey. In particular, this device will help to understand if a sample, advertised to have been selected using certain methods by the researcher, is similar in characteristics to a sample selected by GAO from the same population and using the same selection methods That understanding is implicit in GAO's (1977) Background Paper statement of interests. The main benefit is that the data are obtained without disrupting the original sample. Indeed, the parallel data may help considerably to strengthen the research.

An objection to this strategy might be that it appears to be considerably more expensive than simply reinterviewing the original sample of respondents. But, in fact, most of the problems (and costs) required by parallel sampling would have already been worked out for the original survey. Any parallel effort could capitalize on the original

survey design, target population listing, and procedures developed for sampling, callbacks, etc. The additional cost of manpower for the parallel sample is likely to be marginal in comparison to the investment in the original design. If there is a high risk of disruption of an expensive experiment through investigatory reinterviews, then the costs of parallel interviews are likely to be marginal and the benefits great.

Surrogate Auditors. Direct reinterviews of respondents in an ongoing social experiment can disrupt the research if the reinterviews are conducted by people whom respondents view with strong suspicion. In particular, some respondents who are able to distinguish between auditors and researchers will be less cooperative with the former perhaps on privacy grounds. It may be possible to accommodate this problem by using a surrogate interviewing agency to reinterview a subsample of the original respondents. The results of the surrogate's survey would be provided in statistical form, so that no identification of individual responses is possible; this includes screening statistical results to prevent deductive disclosure of identified responses.

The surrogate might be the one already under contract to the experimenting agency (e.g., the original contractor to HUD may service both HUD and GAO needs for information). If this approach is acceptable to the audit agency, perhaps all that is required is straightforward, but more intensive, reinterview on questions of interest to the audit agency. The approach is likely to be burdensome to the respondent insofar as the demands on his ability and willingness to supply information are increased. But with good pilot testing of questions and interviewer training, perhaps even that burden can be minimized. The

main benefit is that the interviewing agency already under contract in the field has established some rapport with respondents and that respondents are less likely to feel uncomfortable with a familiar agency and familiar interviewers. The scientific benefit is a measure of the stability of respondent reporting, if carried out under the same conditions as the original interview. If different inverviewers are used in reinterview, the resultant data serve as a measure of interinterviewer variability and temporal variability in response. For research, these statistical measures are considerably more important than identifying a few individually identified cases in which the initial interview was poor. This emphasis on the statistical rather than individual result is crucial, must be recognized as such by GAO staff, and is implicit in GAO's Background Paper and other GAO documents on quality in research.

The Committee recognizes, however, that the point of many audits is to verify the integrity of the original contractor's performance. So, for example, GAO may need to establish that a contractor did indeed engage in interviews with particular individuals and that the responses of those persons were of a certain kind. Under this circumstance, a heretofore uninvolved third party may be an acceptable surrogate for the audit agency, insofar as the third party is more neutral or less suspect than the original contractor. The potential benefits of the strategy are that the process of verification can be removed a step from direct government investigation and so may attenuate the possible problems generated by direct government contact with the respondents.

A second benefit of this approach is tied to more general secondary analysis. That is, a research group whose primary mission is secondary analysis could, in some cases, also serve as a surrogate for the audit agency. This outside research group would take primary responsibility for reanalysis of statistical data and for verifying its internal consistency and conclusions based on the data. And it would also take responsibility for, say, reinterviews with a sample of original respondents to verify the credibility of the original research records on those respondents. In this latter activity, it fulfills an auditor's functions, except that it would not provide information on individually identifiable respondents to the audit agency. It can serve primarily as a neutral intermediary to establish statistical reliability of original reports produced by the researchers.

The product of using this approach is information of interest to GAO, notably verifying in part that indeed "selection procedures were carried out" and "variables were recorded correctly ..." The surrogate approach may be particularly useful when audit agencies other than GAO have the skills necessary to reinterview peculiar and suspicious target groups, and where audit agencies other than GAO are less likely to be regarded as a threatening investigatory arm of government.

Subsample Reinterview by Auditors and Sample Augmentation. Direct reinterviews may on occasion be essential to accomplish audit agency goals. For example, GAO could choose to verify that interviews had indeed been conducted by a data collection agency and that certain responses were given in order to check the integrity of the original interviewers. Parallel sampling is normally insufficient for accomplishing these goals, and the use of surrogate audit agency may be unacceptable to GAO.

An obvious approach to minimizing disruption of the ongoing research is to minimize the number of individuals who must be reinterviewed. The GAO, for instance, might select a probability sample from the existing experimental sample. Members of the subsample would then be reinterviewed by GAO staff to elicit the information of interest to them.

At best, the audit agency taking this approach will obtain the basic statistics necessary to establish integrity of data collection (i.e., that the respondents were indeed interviewed by the researchers). Again, at best, it does so without major disruption of the research effort. In addition, the audit agency's records can be linked to earlier information collected by the researchers on the same respondents by using some variation on procedural strategies for assuring confidentiality of sensitive records obtained from independent archives e.g., insulated file linkage (see Campbell, et al, 1977). The normal purpose of such a linkage is to compare average levels of agreement between original interviews and audit-based interviews.

The subsample involved in the auditor's reinterview may subsequently be of little or no use to the researchers. That is, reinterviews may provoke individuals to drop out of the study or to respond differently in the future. If these conditions prevailed, the research agency could remove the GAO sample from its sample and analyze the remaining probability sample with results properly weighted to reflect the reduced sample size. Especially in research characterized by small samples, however, the possible reduction in sample size may be intolerable, for it will reduce notably the precision in estimating the effects of the experimental treatments. The loss can be avoided if the need for such (possibly) destructive testing is anticipated in the design of a large

social experiment and if the sample size is increased accordingly. This implies that if audit reinterviews are anticipated, and the involvement may have negative effects, then samples must be increased beyond those normally required at the research design stage, in order to accommodate subsequent attrition or distortion in responses.

One major risk of this subsampling strategy is that the reinterview process may have an adverse effect on the members of the original sample who are not reinterviewed. If the reinterviews have the character of an administrative investigation and if they are widely publicized, then cooperation rates in the remaining sample are likely to decline. It is not clear how such publicity can be avoided, especially in controversial social experiments, unless different geographic areas are sampled. It is reasonable to suppose that if participants are told beforehand that reports might be selected at random for verification by an outside auditor, then the effects of this might be reduced, if not eliminated. However, there are no good empirical data to support that supposition. Side studies of the issue are warranted.

Other Options. When the auditor's primary concern is verifying the factual accuracy of respondent records, then record linkage may be more appropriate than reinterviews. For instance, research participants' reports of income to the researcher may be better assessed by linking those reports with institutional records on income for the same respondents. The institutional sources may, for example, include employer records, hospital archives, school records, and so on. If they are private rather than public records, then special methods must be used to accomplish linkage without breaching privacy (see the next paragraph). The assessment based on linkage can be better than reinterviews in the

sense that lapses in memory and other factors may degrade reporting to interviewers, and in the sense that institutional records represent one standard against which accuracy can be judged. (Of course, institutional records themselves are imperfect.) Also, if verification of response based on an outside standard is the only objective, and the record linkage is adequate, then the potential problems of reinterview can be avoided.

There are a variety of procedural tactics for linking records from different sources without breaching confidentiality rules governing their disclosure. So-called "insulated data file linkages" are relevant here (Campbell, et al., 1977). In the simplest of such strategies, the researcher supplies a file on respondents including respondents' reports of income to the institutional archive. The archive links its data with the file reports, strips respondent identifiers off the file, and returns the statistical file to the researcher. Variations on this tactic can be designed to enhance efficiency and the privacy protection afforded to the respondent. The researcher may, for example, cryptographically encode the data in his own records to assure that the respondent's status cannot be determined by the institutional archive from the data provided by the researcher. These strategies have been used by researchers to verify the statistical quality of response relative to contents of proprietary records on the same record system (see Locander, Sudman, Bradburn (1975), for example). They are directly pertinent to GAO's objective of verifying the quality of reports obtained from respondents in social experiments.

Some recently developed statistical strategies may also be relevant to the problem of assessing the accuracy of respondent reports

(e.g., Warner 1971). These methods, however, are usually only applicable when very large samples are employed, the interview information is sensitive, and when the planned data analysis is simple. These tactics permit the researcher or auditor to elicit sensitive information from identified respondents -- even in face-to-face interview situa ions -- but without linking particular persons to particular responses. The simplest of such tactics (so-called "contamination method") requires that the respondent inject his answers with some random error (e.g., by rolling a die and lying if a "one" turns up). The rules for contamination are such that the researcher knows the general parameters for errors across the group of respondents, but not the specific degree of contamination for any given person. This method makes the verification of a particular respondent's records impossible, although it makes possible the verification of the credibility of group-level (i.e., aggregate) statistics. The random contamination of response method might be used with a parallel sample in order to compare the aggregate statistics it yields with those obtained from the main study under review.

Testing for the Effects of Reinterview

The committee believes that reinterviews (i.e., direct contact between respondents and auditors) can and should be minimized. There are many alternatives for gauging the quality of social experiments and of their data. Consequently, reinterviews should not be viewed as a routinely necessary task.

This being the case, it is not clear that an immediate major program of investigation to estimate the effects of reinterview could be justified.

Yet, a modest effort to assess the impact of audits in general and of reinterview in particular may be warranted. A few advantages and disadvantages of various tactics are considered in the following paragraphs.

Case studies. A case study might be designed in order to better consider possible consequences of reinterview. In general, case studies are a rich source of process information and ideas.

An anthropologist's exploration of a small group of families involved in poverty research may, for example, conclude that the families fail to discriminate ably between the GAO and any other private or public agency which collects records for research or any other purpose. On the other hand, he or she may find in parallel studies of affluent families participating in energy research, that the families not only distinguish ably among data collection agencies but have strong opinions on the propriety of each agency's interest in private facts. Other case studies may involve examining a small set of programs which GAO has audited, and in which GAO contact with pr gram participants has been substantial. The objective would be to summarize some GAO experience and opinion on the nature of contacts, the problems encountered, and the way they were resolved.

The major shortcomings of the case study approach are (1) that the information depends heavily on the experience and expertise of the individual doing the case study and (2) that it is often difficult, if not impossible, to rule out alternative explanations for the results obtained. In addition, case studies undertaken long after the intervention would not seem to be useful, since memory lapses, loss or destruction of records, mobility of staff, etc. can prevent a reasonable assessment of the intervention. Regardless of these shortcomings, they can be useful in generating ideas about the way more sophisticated research can be designed.

Surveys. Another device is the surveying of the reactions and intentions of the target group involved in the program. In order to gauge the respondents' opinions about reinterviews conducted by the GAO, a few questions might be added to routinely administered questionnaires, for example. Of course, opinions do not necessarily predict behavior. By themselves, they might not be particularly informative about the consequences of reinterviews. But they can help to understand the expectations, understanding, and concerns of the target group regarding interview or reinterview.

Field experiments. The use of field tests employing randomized experimental designs could be developed to investigate the impact of auditor reinterviews. These need not be elaborate and could be designed to focus on whether informing people that they will be reinterviewed affects cooperation rates and response validity. Such an experiment could be a side-study adjoined to a contemporary program evaluation. For example, the sample in the evaluation study could be increased, and the additional respondents used by GAO to appraise effects. This is basically the subsample reinterview mode described on pages 23-25. Or the GAO might choose to mount an independent study divorced from any experimental program's operations to assess alternative ways in which reinterviews by GAO may influence social experiments.

Cost and benefits. Whether any of these simple options should be exploited depends on the benefits and costs of each. Case studies are cheap, often rich in ideas, but often ambiguous in conclusions. Surveys are more expensive and can be designed to appraise opinions well. But they may not be sufficient in themselves. Experiments are

most expensive, put greater demands on staff, and can yield more information on effects. A series of small experiments adjoined to ongoing projects is still more demanding, but more consistent in building an understanding of the potential costs and benefits of GAO interventions.

The level of investment in these options depends, then, on GAO's interests and the potential payoffs. Interest has already been registered through a variety of current activities, including the contract with the Social Science Research Council. In any event, if the GAO considers a program of testing, helpful advice must be obtained from other groups (e.g., the National Academy of Sciences Committee on Federal Statistics) and other agencies (e.g., the U.S. Bureau of the Census, the National Center for Health Statistics) that have notable experience in conducting similar methodological studies.

The committee believes that a modest effort to monitor the effects of future GAO interviews and reinterviews is advisable, provided the GAO recognizes that this should not be a one shot effort. At a minimum, for example, it is reasonable to maintain -- at a central location -- logs and correspondence on reinterviews or interview activities for each project. In this manner, a small documentary resource could be built, which could be analyzed periodically to identify reoccurring problems and issues (or lack of them). So, for example, each research project in which GAO has had direct contact with research participants might be listed, the peculiar problems of contact described, and the consequences of that contact (on cooperation in research, for instance) might be briefly described. These could serve ultimately as the subject of papers presented at the appropriate professional forums for discussion and peer review. For a given project, the documentary file might be supplemented with brief telephone interviews with researchers and GAO staff in order to determine the extent to which

opinion about effects (or lack of effects) are verifiable, to assay roughly the size and importance of effects, and to identify competing explanations for the effects. This sort of activity could be undertaken alone or in conjunction with more elaborate GAO efforts to estimate effects of its involvement in social experiments and social program evaluations. As in any such effort, GAO ought to maintain its general interest in minimizing the bureaucratization of the process. The product of the effort is a consolidated archive on the nature and effects of GAO involvement in interview or reinterview of participants in research. If the GAO's investment in monitoring quality of interview data is substantial, such an archive serves as an institutional memory bank and is likely to be essential in keeping GAO performance at a high level.

A more elaborate investigation -- for example, of a specific project or class of projects -- would emphasize active estimation of the effects of GAO reinterviews. The subsampling and augmented sample method described earlier is perhaps most feasible. Provided such small experiments are well-designed, they would permit the verification that there is indeed a problem resulting from GAO reinterviews or contrariwise, that the benefits of reinterviewing outweigh the costs: For example, one might discover that the subsequent participation rate in the GAO's reinterviewed sample is higher, rather than lower, then in the larger non-reinterviewed sample. If neither group differs with respect to cooperation in future experimental interviews, this could suggest that in similar projects GAO involvement will not have large negative effects. In any event, the results of such studies should be replicated through a series of small experiments.

Obviously, if these small experiments are adjoined to large tests of social programs, they require the cooperation of principal investigators (i.e., project directors) and the early involvement of GAO in the design of limited features of the research. The experimental approach also requires that GAO staff design the side-study and implement it. These demands are high, and the committee does not believe that such testing should be undertaken unless resources are available to do the job very well.

SECTION III

GAO ROLE IN SOCIAL EXPERIMENTS

GAO and the Phases of the Experiment

In the following remarks, the committee considers the opportunities which the GAO has to improve or monitor evaluations of social experiments at each stage of the evaluation process. The committee has examined alternative roles for two reasons. First, a GAO decision about whether to reinterview, or more generally, decisions about how GAO should monitor quality of research data, depends in part on the other roles which GAO takes in monitoring research. Second, the focus on reinterview as a device for assessing quality is very narrow. Other GAO activities are likely to be at least as important, and in some cases far more important, in helping to improve the quality of information made available to the Congress by field researchers.

Under the 1974 Budget Control Act,* the GAO has been assigned responsibility to oversee all federally supported program evaluations. The committee understands from the GAO Background Paper (April 8, 1977) that

*The Congressional Budget and Impoundment Control Act of 1974.

there is a special interest in assessing the design and conduct of social experiments. Consequently, the committee has focused on social experiments, but it recognizes and emphasizes that our observations will hold true for a large variety of other evaluation-related activities.

The topics to be described briefly in this section include activities in which the GAO has the "right and option" to participate. By recognizing that the GAO has both the right and the option to participate at any given stage, the committee intends to indicate that it is GAO's decision whether it should participate or not. This decision will depend, in part, on the GAO's current interests, manpower, and budget. The committee does express some opinions, however, on what it believes to be the best GAO roles.

The topics described below are organized in terms of the typical stages of the process of conducting an experiment: (1) formulation of research problem, (2) program formulation, (3) procurement, (4) evaluation budget, (5) design of the experiment, (6) implementation of the experimental design, (7) reporting schedule, (8) close of field operations including data checking, and (9) analysis. To the degree that GAO's participation at any stage is timely and of high quality, and does not impose notable added burdens on the research process, the greatest will be the benefits.

The committee does not propose that the GAO participate actively in every stage of the social experimentation process. Rather, the idea is to lay out the field of inquiry and to identify where GAO might conceivably make distinctive contributions. In addition, it is the committee's belief that ultimately any resolutions of arguments about GAO access to data and

privacy assurances depend in part on the nature of the research process and on the typical roles that GAO plays in that process.

The only major qualification on the committee's remarks is tied to the ideas of pluralism and quality in scientific, including social scientific research. The committee cautions the GAO against intentionally or unintentionally setting itself up as a primary arbiter of conflicts in experiments and as a primary definer of research standards.

Formulation of the Research Problem. Whether an evaluation research question is framed adequately is certainly a legitimate concern for the GAO. But, there is no clear standard on which one can rely in judging the adequacy of evaluation research questions. Furthermore, the reformulation of the questions by the GAO after an experiment is underway may be counterproductive. That is, the critique of the evaluation study may become focussed predominantly on the "formulation issue" with less attention being given to other important aspects of the social experiment. The committee believes that when the GAO regards the formulation to be inadequate or incorrect, it has the responsibility to state its opinion. However, it should also address itself to how adequately the social experiment will help to answer the questions as formulated by the researchers. That is, care should be taken to separate arguments about the formulation of the questions from arguments and criticism about the design and implementation of experiments addressed to particular questions.

Program Formulation. GAO may also choose to participate in the process of program formulation. However, its role here may be limited severely by constraints on staff time and expertise. It is difficult

for the career researcher, let alone an individual operating outside a strong social research tradition, to keep track of all the ideas and plans available for program development. Moreover, that program formulation has been a mission of other agencies by law and tradition. If there is a definite mandate (e.g., from Congress) to review the researcher's or sponsor's formulation, then GAO participation may be useful. In particular, when competent staff are given enough time to understand the issues involved, they could make valuable contributions to the undertaking and be in a better position to evaluate the evidence from the social experiment when it is subsequently obtained. For example, many GAO reports on social programs of the 1960s stress shortcomings in program formulation. To the degree that those same problems appear in future tests of innovative programs, it is reasonable for executive agencies and program developers to consult GAO reports in order to anticipate problems in program formulation and to help devise alternative methods for their solution.

The process of program development is time consuming and complicated. Routine GAO involvement is likely to be unwarranted. Rather, where it is clear that GAO experience and expertise is relevant to the program at hand, its advice ought to be sought by the agency developing the program. Further, any such involvement must be coordinated with other principals so that research is not delayed any more than it currently is, nor costs increased significantly, for both long delays and significant costs are likely to degrade the timeliness and worth of the research.

Procurement. There does not appear to be a clear role at this stage for GAO beyond making sure that lawful and sound contracting procedures are used. The procurement stage of the research process is a critical link in the chain, however, and is not always carried out competently. GAO could conceivably expand its role by, for example, requiring that contracts include provisions for a representative subsample of respondents to be available for possible GAO audit and review. If the GAO is willing and able to specify its own audit needs before the experiment is put into the field, then this may help improve the quality of the subsequent relationship between the GAO and the researchers. More generally, GAO might choose to review periodically the procurement procedures of agencies to assay the extent to which the procedures degrade or enhance the quality, efficiency, and timeliness of research.

Evaluation Budget. Sometimes the budget for a social experiment is inappropriate -- either too little or too much for the job at hand. The committee believes that GAO could exercise influence at this stage by assessing the realism of budgets in an experiment -- provided that GAO's experience in making such judgments is sufficient. To be useful, this judgment must be made before the experiment is put into the field rather than after its execution. While understanding the costs of evaluation research, including social experiments, independent of program costs is complicated and difficult, the GAO may be in a good position to encourage better budgeting procedures and the documentation of costs as well as in a position to propose and analyze alternative ways of accounting for costs. In particular, not enough is currently known about the costs of high quality social experiments or about program evaluation more generally.

It appears to the committee that GAO is in a good position to take a major role in consolidating what is known about costs, in identifying areas where by virtue of the innovativeness of a program that evaluation costs cannot be anticipated well, and in developing guidelines on description of costs.

Design of the Experiment. The GAO has the opportunity to review the adequacy of the statistical, managerial, and other features of the experimental design. The review may involve assuring that ordinary good practice in statistical design of experiments has been recognized by the contractor or grantee. This might include, for example, verifying that: an explicit design has been constructed; that statistical power of the design has been examined; that the design is pertinent to questions or hypotheses at issue; that ordinary problems in implementing designs, such as attrition, have been anticipated, and so on. Any more sophisticated review depends considerably on familiarity with a rapidly developing state of the art, notably the way designs must be tailored to accommodate institutional, legal, and ethical constraints on research, the way needs for internal validity (unbiased estimation) can be balanced against external validity (roughly speaking, generalizability of results), and other issues.

However, the issue of the adequacy of the design should be separated from issues of the adequacy in formulating the research question. Judgements about adequacy of design at the local (project) level can rely on existing guidelines for good practice, though there will always be some design issues on which reasonable experts will disagree. In addition, it should be recognized that the state of the art in experimental design is

not always well known by expe. s in related substantive areas. For example, an expert in conductir , manpower surveys is not necessarily an expert in designing or asse sing the design of an experiment on manpower training. There is some overlap in skills, but the discrepancy in skills required by the two tasks ought to be recognized.

From the GAO's (1977) Background paper, the committee understands that the GAO takes some responsibility for review at this stage and the committee recommends that the activity be coordinated with other agencies with similar responsibility. The Office of Management and Budget has, for example, routinely made reviews of some classes of survey designs. Any GAO involvement should avoid unnecessary redundancy and significant increases in the time and costs to the public or to the grantee or contractor in review.

Implementation of the Experimental Design. The GAO may choose to assess the researcher's implementation of the design in order to determine whether randomized assignment, sampling, measurement, and so forth are carried out according to the design plan. Since the standard for judgment here is an explicit design plan, evaluative judgments are somewhat easier to reach. The committee believes that the GAO can be helpful in assuring that the design plan is well implemented. This role is, indeed, one that the GAO has played in the past and is explicit in GAO's (1977) Background paper for this committee.

However, in making its evaluations, the GAO should realize that no research design is ever implemented perfectly and that experience and expert judgment is essential in gauging adequacy of implementation. Given this caveat, the GAO still can play a valuable part in assuring the cor-

rect implementation of the design and in encouraging the adequate documentation of any deviations from the original design. To be most useful in this phase of the social experiment, the GAO should participate during the process of design implementation rather than after the fact.

Close of Field Operations and Data Checks. If the GAO wishes to assume it, there is an important role to be played at the close of field operations. This role involves encouraging and verifying the prompt and complete documentation of the field procedures and operations, before the program and/or contractor staff have dispersed. Closely related to this is encouraging the preparation of public use data files, which can be made available for reanalysis after the final report (or first analyses of the data) is completed. The GAO has the opportunity, and as we understand it the resources necessary, to check such data files in order to assure that the contents are well documented and internally consistent. This last activity is very much in the spirit of more traditional accounting roles; it is better tied to clear standards of adequacy than many of the other tasks discussed here, and is perhaps the most feasible and important task so far considered.

The GAO role here may take a variety of forms. At a minimum, it may involve merely emphasizing the need for good documentation in GAO standards, guidelines, or manuals for review of policy-related research. It may involve GAO's developing more explicit guidelines than are now available on what to document and how to document, so that quality of an experiment is easier to understand and so that GAO's standards for quality in documentation are satisfied. It may, through publications, catalogue other professional group's guidelines on documentation of social science

data files, to aid both GAO staff and agency staff in their adoption of those guidelines which are most suitable for the particular experiment.

Analysis. The GAO will at times have the opportunity to analyze or reanalyze raw data (e.g., anonymous records on program recipients) from the experiment. This role can be useful in the simplest sense of detecting conspicuous errors made by a grantee or contractor. It is clearly most important in the sense of avoiding an unwarranted reliance on a single data analysis, and in contributing an independent analysis. GAO's analysis, like the researcher's and others' can span the entire range of research problems: understanding the nature, accessibility, and character of the target group for the program; understanding the degree to which the program has been implemented in the field tests; estimating the relative effects and efficiency of the program; making carefully justified generalizations beyond the immediate setting of the experiment; and integrating the findings on an experiment with information from other sources. But as the GAO recognizes, it should not use its office to override or dictate particular analyses. In this respect, the GAO must be regarded by others and must regard itself as only one of several independent analysts.

To be more specific, consider that some evaluations result in outcome data which can be analyzed in a variety of ways. Estimates of program effect may differ appreciably depending on the analyses performed and on the assumptions which underly analysis. When the GAO performs analyses of its own, it should recognize this and be prepared to defend its choice in professional forums. This is especially true for new methods of analysis and advanced techniques that are not yet part of routine practice. The Committee recognizes that GAO already makes its analyses

of a project public and does invite counterargument and comment from staff of the agency which sponsors the project. The Committee believes that the practice ought to be broadened to include comment and exchange with the wider research community. This may include, for example, publication of GAO analyses in scholarly journals, reports on analyses at professional society meetings (e.g., the American Statistical Association, Council for Applied Social Research, Evaluation Research Society, and so on) so that scholarly discussion and criticism can be fostered.

GAO can and should capitalize on analyses of a social experiment which are undertaken independently of GAO and of the original experiment's staff. Secondary analysis of data from program evaluations has been undertaken more frequently in recent years by nongovernment researchers. And the reanalysis of data from experimental tests of "Sesame Street" (Cook, et al, 1974) and others are valuable in several senses: they are run by researchers outside the original staff for the experiment and will have different views about the value of the information and the way it should be analyzed. The secondary analyst may also dedicate more time and resources than either the original researcher or GAO have to expend on a reanalysis. (Analyses by consultants commissioned by the GAO are considered below.)

An Observation on Early GAO Involvement

The committee believes that earlier rather than later involvement by the GAO in some program evaluations is desirable. In particular, it is less productive than might otherwise be the case for the GAO to criticize the objectives and design of a social experiment long after the experiment's implementation. At such a late stage, it is often impossible to correct

program deficiencies and/or to implement recommendations for improving the experiment. If the GAO were to review a program before it was put into the field, then its criticisms and recommendations would, at least, have the potential of being acted upon.

The committee does not believe that the GAO's rendering an early opinion will compromise the GAO's ability to later review and criticize the implementation of the program's objectives and design. The earlier criticism, if not met or responded to by a sponsor or researcher, could always be reiterated in subsequent GAO reports on the project. This view does not deviate substantially from the current occasional practice of early criticism, as in the case of EHAP.

Early involvement does not appear to be feasible or desirable on a routine basis. It is most likely to be useful where GAO's experience and expertise are clearly relevant to an experiment, and where the policy-relevance of the study justify the increased complexity and costs engendered by earlier involvement. It is in the interest of the agency sponsoring the program and its evaluation to seek expert advice and reaction wherever it is available, including from the GAO, and GAO has some responsibility to consider such requests. Even occasional early involvement should be scheduled so that the process of mounting the research is not delayed unnecessarily and it should be coordinated with other principals with review responsibilities (such as OMB and non-Federal review groups). Any review which delays an already slow process notably and is redundant with, or less expert than, existing reviews will not enhance the quality of program evaluations or social experiments.

General Style of Audits

The GAO's style of operating will affect its performance in this area and it will influence outsiders' reactions to that performance. Three specific issues which may influence GAO efforts to monitor quality of research have come to the committee's attention. They include research community's view of GAO's role, the enumeration of criteria by which research quality can be judged, and the experience and expertise available to GAO for assessing a social experiment or program evaluation. They are discussed here briefly.

The research community's view of GAO. Some major sponsors and researchers hold the view that audits have been traditionally mounted to identify and dwell on problems rather than to assist in the solution to problems or to recognize good solutions. This is not, as we understand current GAO policy and products, a uniformly accurate view. Nonetheless, it is important since an inaccurate and uncomplimentary view affects the outside research community's opinion of GAO work and interferes unnecessarily with scientific discussion between the GAO and the research community. The view that the GAO emphasis is deficiency reporting creates a climate of confrontation rather than one of considered professional exchange. In the worst cases, their concern is that the GAO's need to respond to Congressional demands for audits of a program may lead to unnecessary distraction and may do nothing to improve the quality of their program. The committee recognizes and encourages the current GAO emphasis on balanced reporting and professional examination of topics. But, we believe that in carrying out this role, in emphasizing its mission of recognizing solutions as well as problems, the GAO needs to make

its efforts better known. That is, inaccuracies in the views held by the research community must be dealt with, in part, by producing the best possible quality reports and by disseminating those reports widely to the relevant research community. In order to reduce the incidence of inaccurate and uncomplimentary views, it is essential that GAO staff -- who are involved in the oversight of social experiments -- be most able and be willing to develop their own skills and the GAO's reputation for fair and capable reviews. As noted elsewhere, this can be accomplished by the GAO staff's participation in scientific forums on social experimentation and research.

In this respect, the GAO's position is analogous to that of any secondary analyst. That is, any outside researcher who performs competing analyses of someone else's data may provoke that person's anxiety and suspicion. (For example, see Cook's reanalysis of Sesame Street and the various attempts to reanalyze the Negative Income Tax data.) But, the nature of the reactions to GAO as opposed to outside researchers differs considerably.

Enumeration of criteria for judging research. The difference in reaction stems partly from the fact that the GAO's criteria for the judging of the adequacy of an experiment have not always been clear to the researchers and from the fact that the GAO's criteria may differ notably from those of researchers. If an academic researcher undertakes a secondary analysis, the standards are a bit clearer to the original researcher, if only because there is an obvious commonality of training and experience between the two of them. The lack of obvious commonality among researchers and auditors leads to suspicions on behalf of the researchers that they will not be "fairly" or "correctly" judged. The ambiguity in standards

means that -- from the researcher's viewpoint -- virtually any deviation from a design plan can be regarded by an audit agency as an error. And this, the committee believes, increases the researcher's perception that the GAO is more interested in fault finding than in evenhanded assessment of the quality of research.

A related issue generating some strain between researchers and auditors is that of the researcher's proprietary rights to his data. The social scientist who dedicates considerable time and energy to a social experiment must be permitted to capitalize on that experience. The opportunity to discover and to report on unique discovery is clearly one of the strong influences on the growth of of the sciences. Eliminating the opportunity or making it considerably more difficult to emerge is likely to degrade the calibre of research. Consequently, the committee believes that outside secondary analysts, including an audit agency, should recognize that the analysis of data and its prior publication by the secondary analysts may preempt the original investigator. Thus, some effort should be devoted to assuring that the original investigator's rights to the first exploitation of the data are not violated by secondary analysts. Of course, this does <u>not</u> mean that the original investigator has the right to delay analyses indefinitely. Assurances which facilitate public interest in sustaining both rewards for research and timely competing analysis by an audit agency are varied. They may take the form of contract provisions under which the researcher is required to release analyses, <u>and</u> the statistical data on which they are based, in a specified schedule. The assurance may take the form of GAO guidelines and policy which recognize proprietary rights explicitly and which enumerate conditions under which those rights must play a primary and a secondary role. Where conflicts

occur, as they must in complex and controversial research, some mechanism for adjudicating them, such as an ombudsman, might be developed.

Expertise available to GAO. An important issue in the style of GAO operations is the professional qualifications of GAO staff. The researcher's concern is that the GAO staff will not be scientifically responsible in the evaluation of his or her project. This concern is real, regardless of its accuracy. In order for the GAO to function effectively in the evaluation of social experiments -- especially with the degree of scope as the GAO's position paper to the committee implies -- it will have to enlarge its investment in the development of a highly skilled staff. Even then, it will take time for the GAO to build its reputation among researchers in this area; for it is through peer review, participation in professional forms, publication of reports, and the like that a reputation for high quality research is developed. The GAO's fine reputation in orthodox audit operations may be sufficient for its traditional role, but not for assuring high quality assessment of social experiments.

The committee agrees with the GAO position that no single profession is normally equipped to handle all aspects of evaluating social experiments. But, the technical aspects of social experiments and a fast developing state of the art require high caliber statisticians and methodologists who are knowledgeable about experimental and quasi-experimental designs. The committee supports and encourages the GAO's efforts to develop a cadre of internal experts and outside consultants with these skills.

The committee recognizes and applauds GAO efforts to augment staff efforts, where necessary, by using consultants. The committee recommends, however, that outside consultants be acknowledged by name and institutional affiliation when they contribute to audits of program evaluations. This

will, the committee believes, help to enhance the quality of the advice they offer, and help to inform the research community about the range and caliber of advisors available to GAO. The committee did not discuss the matter of acknowledging authorship of GAO reports by specific GAO staff members. However, the idea has merit in helping to enhance quality of audits for similar reasons, just as not acknowledging individuals has some merit. (For a discussion of this issue, see Kruskal, 1978).

A Closing Comment

The committee believes that there are potential problems as well as benefits associated with GAO participation at any or all stages of the research process. In particular,

> It would require a substantial staff trained in statistics and the social sciences or of social scientists trained in auditing in order for the GAO to participate in all or most levels.
>
> The demands on sponsors and researchers will be greatly increased should GAO become more involved, since any GAO audit will parallel monitoring by other agencies and external review groups.

The committee will not comment on the first problem, but the second is considered in some detail in the following section on the "Costs and Benefits of GAO Audits." The theme of the next section is that because the demands on sponsors and researchers are great during an audit (in terms of time and effort) and because they can delay and increase the complexity of the research, the GAO's role should be coordinated as much as possible with the monitoring roles of others. The premise is that reviews should be simultaneous (rather than sequential), coordinated, and not unnecessarily redundant with other monitoring agencies or with internal monitoring activity of the executive agency. For example, for several

months both the GAO and the OMB were reviewing EHAP, and -- to the committee's knowledge -- these reviews went on in mutual ignorance and isolation.

SECTION IV

SOME COSTS AND BENEFITS OF AUDITS

In this section, the committee offers its opinion on potential costs and benefits of GAO audits for the researcher, the sponsor, and for the participant in social experiments. We do not consider costs and benefits to Congress and other agencies, nor to the general public, since they lie well beyond the scope of this report.

For the Researcher

The potential benefits of an audit to the researcher are indirect and delayed. One possible benefit is that GAO involvement in an experiment can result in better documentation of problems and of solutions in design, implementation, analysis, etc., and this could help improve subsequent research. That is, as the GAO contributes to what we know about persistent problems and solutions in a large and complicated arena of research, the research community can rely more heavily on GAO reports as a competent source of information and independent analysis. Another potential benefit, and perhaps a more immediate one, might be the GAO's explicit recognition that the experiment is consistent with good research practice. This is likely to enhance the reputation of the researcher and credibility of the research in the short run, and it could help encourage good research practice in the long run.

For the researcher, the potential costs of an audit consist mainly of increased demands on time and manpower resources in order to prepare

documents and to meet with GAO auditors. In some instances, these costs are passed on to the sponsor, who may reimburse the researcher for the financial costs generated by the audit. Whether or not they are, the process of the audit may engender delays in the progress of the research and divert attention away from research operations. The delays may be sufficient to prevent reports from being prepared on schedule, and the diversion from day-to-day activities may affect the quality of operations.

In considering costs and benefits to the researchers, the committee is also aware of ripple effects on subcontractors. For example, whenever a GAO audit distracts attention from a fragile research activity, influences a research schedule, or deflects efforts to manage an unstable network of collaborating institutions, then subcontractors may suffer from delays and lack of attention from the principal investigators.

A ripple effect on benefits is equally plausible: subcontractors may improve performance in anticipation of GAO examination of their performance.

For the Sponsor

A possible direct benefit of an audit to a sponsor is that the GAO's independent and expert opinion helps improve the quality of the current program and future programs of the sponsoring agency. To the extent that the auditor's opinions and recommendations are competent, fair, and well articulated, the greater this benefit can be.

The potential costs of an audit to the sponsor may be both direct and indirect. The researcher's cost in time and resources may be charged directly to the sponsor in terms of increased dollar costs for the experiment, as was the case in EHAP. When all costs cannot be charged in this

manner, the researchers may in future bids on contracts or grants increase the size of their bid in order to cover the anticipated financial burden of an audit.

For the Participant

The committee believes that the benefits of GAO reviews, though difficult to measure and normally indirect, can be notable for the research participant. The benefit is greater to the extent that the independent review of research operations is timely, economical, and does not impede the progress of research. The benefits of a GAO reinterview (in contrast to other audit activities) for the individual research participant are not so clear. There may be a direct benefit in the same sense that any interview carries a direct benefit: many people enjoy being interviewed (see, for example, the National Academy of Sciences, 1978). However, we do not have sufficient information on the matter to reach an opinion for interviews conducted by GAO.

The conceivable costs to the research participant are likely to be direct and notable only if reinterviews are undertaken. If agreements on whether the GAO should reinterview are reached beforehand and if the participant is informed of the agreement, then the cost to the participant may include:

> An increased demand on the participant's time, due to the reinterviews.
>
> An additional intrusion upon the participant's privacy.

Both can be entirely nominal demands, and the participant may agree to meet them. However, good research practice dictates that the burdens placed on respondents be minimized. These demands should not be made if they are unnecessary -- that is, if it can be shown that their benefits do not offset their costs.

In the case where agreements about reinterviews are not made in advance and if reinterviews by an audit agency breach previous assurances of privacy and confidentiality, then the costs to the participant may be higher. The social cost of not fulfilling an agreement is less direct and less measurable, but it warrants serious concern. To the extent that the controversy erupts and persists into the public domain, the greater the likelihood of costs in terms of lost privacy to the participants. In addition, the loss of goodwill among participants may make others less willing to participate in future social experiments.

SECTION V

ROLE AND RESPONSIBILITIES OF SPONSORS AND RESEARCHERS

Whether the GAO will need to reinterview respondents and what roles the GAO could most effectively take in overseeing evaluations of social experiments depends partly on the normal roles of sponsors and researchers. Consequently, the committee reviews these roles briefly here.

This section serves as general background to the earlier discussion of alternatives to reinterview. It is more pertinent as an outline of institutional factors which affect the quality of research. In particular, the committee views some of the institutional processes involved in developing a social experiment as essential. However, the poor implementation of processes is a problem currently being explored by researchers, by staff within executive agencies, by the GAO, and by others.

The subsection entitled "The Researcher and Quality Control" furnishes the committee's view on researchers' responsibilities in collecting and maintaining data from a research project. It constitutes a basis for determining the quality of data in GAO's examination of a project (pages 18-19).

Roles of Sponsors

This discussion considers several ways in which the sponsors of social experiments and of their evaluations contribute to the experiment's success or failure. It is based heavily on work by Ilene Bernstein and Howard Freeman (1975) and on the current assessment of evaluation projects that Freeman is undertaking with Linda Bourque. This body of work has made it clear that successful evaluations of social experiments are as much dependent on the activities of the sponsors as they are on the working styles and outlook of evaluation researchers themselves. There are three key ways in which sponsors influence the quality of studies.

Developing and Implementing RFPs. Almost without exception, social experiments and their evaluations are now undertaken in response to "Requests for Proposals" (RFPs) issued by sponsoring agencies. Moreover, it is very likely that this competitive bidding system will continue to be used, since the Secretary of Health, Education, and Welfare has taken a strong stand against sole source contracts.

There are three aspects of RFPs that should be considered in regard to their impact on the substantive and methodological quality of social research. First, many RFPs inadequately describe the research problem. In some cases, the description is so inadequate that the prospective researchers must create their own study questions. In so doing, a methodologically astute investigator may fail to obtain a grant or contract because he has failed to second-guess the intent of the study correctly. Further, many RFPs fail to describe fully the policy considerations that led to the need for the social experiment or program evaluation. And they often fail to specify the policy alternatives that are practical and politically feasible in the event of different types of findings resulting from the study. Without such information, it is impossible to develop appropriate study designs.

Second, the time available for the preparation of good proposals based

upon the RFPs can be notably insufficient. That is, the length of time between the advertising of the RFPs and when proposals have to be submitted is sometimes so short that no organization can be expected to develop a high quality design. The problem is compounded when RFPs are published with many amendments and confusing budgetary instructions, as some are. That too makes it difficult for competent researchers to respond in a timely manner.

Finally, the quality of the award process can be very uneven. The judgment processes range from peer-review procedures (similar to those at NIH) to in-house administrators, some of whom do not have sufficient training to make competent decisions. In addition, the awards are often made hurriedly, e.g., a sizeable proportion of RFPs end up processed in the last months of the federal fiscal year. See, for example, the relative frequency of RFPs issued over the course of the year and those issued just prior to the close of the fiscal year in Business Commerce Daily. This minimizes the opportunities for serious and deliberate pre-decision discussions with the various bidders.

Monitoring Projects. The project monitor is literally the only person who has a sustained opportunity to make sure the work is done in reasonable conformity to the design, to modify the design and funding because of unexpected contingencies, and to assure the proper conduct of the project in terms of the protection of human subjects, privacy, and the like.

Ideally, the monitors should be persons with considerable research experience and management ability. However, monitors vary greatly in their research background, interest in their role, and their availability to the project. That makes their use of advisory boards and consultants crucial, and both should be regarded as sharing responsibility in monitoring projects. Further, project monitors are often hindered by federal and state regulations that prevent them from being flexible, even when they and the project see the

advantages of modifications in design, budget, or time schedule. The clearance process required for review of educational surveys is an often cited example of an activity with good objectives, which is unnecessarily inflexible and for that reason, sometimes inefficient (Carter, 1977; Duncan and Haber, 1978). Because of these problems, monitoring often becomes a pro forma activity, and the social control opportunities as well as consultative inputs are foregone.

Besides overseeing the project's general operations, project monitors and their formal advisors can play an important role in the validation of the project's results. If they perform their role correctly, both monitors and advisors can testify to the adequacy of the design, its implementation, and the data analyses.

Dissemination. The sponsors of research need to provide maximum opportunities for the dissemination of the results of their studies. At the present time, for example, final reports from projects are often extremely difficult for the research community to obtain. Despite their citation and use in Congressional testimony, for example, some reports are not printed in sufficient numbers or circulated widely outside the executive agency providing testimony (Hedrick, et al 1978). That it is annoyingly difficult to locate source documents cited by government reports concerning topics other than program evaluation is also clear (Kruskal, 1978). The publication of compendiums, such as GAO's (1976) Federal Program Evaluations, can make identification of pertinent reports much simpler. The dissemination of information, including reports, is important for two reasons. First if studies are worth doing and worth supporting, then their findings should be made known to the widest possible audience. This is true whether

or not the projects are "successful." It is equally important that studies revealing a lack of success and efficacy of social interventions be known as well as those that indicate the value of particular programs.

Second, dissemination serves as another control mechanism for improving the quality of research. If researchers and sponsors realize that their work will be subject to public and peer scrutiny and that this scrutiny may influence their personal careers, then they are likely to perform better. Under ideal circumstances, sponsors of research not only should disseminate studies widely and provide notices of completed studies to a broad audience, they should encourage the presentation of professional papers and the publication of results in professional journals. Such efforts would not only work toward improving the quality of individual investigations, but would cumulatively result in higher quality work in general.

The Researcher and Quality Control of Data

One important responsibility of the sponsors and researchers is insuring the quality of the data collected and processed. However, there is no one set of standards for quality that apply equally well to data collection activities for all social experiments or social program evaluations. Although any set of standards would have to be tailored to particular settings, the committee can identify general guidelines to reasonable practice, which can be used to help assure the quality of the resultant data. The committee outlines them here primarily because the researcher's quality control procedures are an important factor to be weighed by an auditor in deciding upon the scope of an audit. In particular, the better the quality control procedures used by the researcher, the less the need for auditors to re-interview respondents.

Basic considerations in assaying quality of research data are described in Riecken et al (1974), Nunnally and Durham (1975), and other contemporary texts. Advances in the area are reported in proceedings of specialized meetings, e.g., on health statistics, by the National Center for Health Services Research (1977) and an attitude measurement by Sinaiko and Broedling (1976), for example. Recent monographs which summarize research on factors influencing quality include for instance Ferber's (1966) text on financial reporting, and Sudman & Bradburn's (1974) reports on over 200 such projects. Bibliographies, such as the ones produced by the U.S. Census Bureau (1974), Dalenius and Lyberg (1976), Kulley (1974), and others are also pertinent.

The following remarks concern quality beginning at the stage at which research participants are sampled.

Objectives and Indices. The basic objectives of quality control systems in data collection and processing are to assure (a) the validity of sampling and (b) the reliability and validity of the measures. Roughly speaking, the validity of sampling means that the actual sample surveyed (e.g., individuals, schools, firms, etc.) should match the target sample as closely as possible. Reliability and validity of measures refer to the interpretability of the participants' responses under some explicit criteria (i.e., the measures correctly assess the behaviors, qualities, etc. that they are supposed to and that they do this in a consistent -- reliable -- fashion).

This implies that the researcher has a clear responsibility to justify and specify the population and the sampling frame from which the sample will be drawn, to justify and specify the sampling plan in detail, and to provide

a complete accounting of the sample units (e.g., persons, housing units). These responsibilities are usually met by supplying quantitative indices of sampling validity after the research is underway, such as response rates. For long term experiments and longitudinal surveys, these indices become more elaborate, since response rates vary over time and over experimental/control groups. Indices of reliability should also be provided in the form of either short-term test-retest (stability) measures or of internal consistency measures.

Field Operations. In the field, quality control mechanisms generally include the selection, training, and monitoring of interviewers and/or test administrators on a sampling basis. It is the researcher's responsibility to make explicit this process.

One aspect of the monitoring of data collection is generally built around a centralized internal editing procedure. Usually, this involves some form of supervisorial verification of a fraction of the sample: that is, supervisors may reinterview respondents on the sampling list by telephone or in person in order to verify that they were indeed contacted by the interviewers, that particular questions were asked and (in some cases) that they provided particular responses to key items. In very large surveys this practice is common, but it is less so in smaller scale studies. Another method of monitoring for verification purposes, especially in large, complex experiments, is the review of individual (interview) protocols for internal inconsistencies. Less common and usually experimental devices for monitoring include tape recordings of actual interviews (especially of telephone interviews) and interviewing in the presence of supervisors.

One important responsibility in the monitoring of data collection is complete documentation, i.e., maintaining of a management log, record, or archive that includes descriptions of special problems and their solutions. Such a management log documents, therefore, any deviations from the original design.

Pilot Tests. Aside from supervisory checks on the process of interviewing or testing, it is generally advisable that pilot tests and side studies of reliability and validity be carried out. For instance, pilot tests of an achievement test used in compensatory education programs will be essential whenever the test itself is new or is a standardized test being applied to a novel target group. Similarly, pilot tests of attitude measures, personality inventories, rating scales, and the like are customary in large studies to assure that the responses are interpretable (i.e., meaningful). It is the responsibility of the researcher to make explicit plans for such pilot tests or later side studies, to justify their presence or absence in the context of the particular project, and to provide the results of these studies.

Data Processing. Quality control over data processing is generally designed to assure that the translation of responses from hard copy (i.e., the raw data of the questionnaires, etc.) to magnetic tape or IBM punch cards is as accurate as possible. Usually, the hard copy is sent to a central processing station, where the quality control is exercised. Part of the quality control system involves the training of coders who translate the hard copy into machine readable codes. In the case of massive studies, the coder's work is often checked at the 100 percent level for a brief period during or after training, and later at the 10-20 percent level. For some

major studies, there is an additional checking of the keypunching in order to catch typographical errors. Regardless of the specific nature of the checks, the researcher should make explicit the statistical quality control standards used and the procedures employed, and should maintain records on results of the control process.

The next step in the process is often the machine editing of magnetic tape or cards. The typical criteria here include range checks to assure that all transcribed observations fall within prescribed bounds for the measures, and internal consistency checks to assure that responses are consistent or reasonable. For instance, a punch card showing that an individual has an income of $60,000 per year and lives in a house worth $3,000 would imply an error in transcription or response. It would "fail" a test of reasonableness or consistency of responses and be flagged for further scrutiny.

Because no standards of control are uniformly applicable, and because control at this stage is typically necessary, the committee believes that it is the responsibility of the researcher to describe explicitly the control procedures. The information should be available to the GAO and the community of researchers.

A number of professional organizations have undertaken workshops to develop better documentation standards and guidelines for social science research, e.g. the International Association for Social Science Information Service and Technology (IASSIST), the National Bureau of Economic Research, and others. As those guides become available and as they are tested in the field, they will represent a useful source of guidance for both the research community and the GAO.

A Joint Responsibility: Public Access to Data

An additional responsibility of researchers and sponsors should be the dissemination of reports and data. Currently, it is normal practice for government sponsors to require a final report on a project evaluation. However, many such reports are not disseminated widely, if at all, beyond the researcher and the sponsor. In particular, low quality evaluations and social experiments are rarely reported in public forums and often are not even stored. Consequently, they are not accessible to competent critics or other interested parties.

The committee believes that final reports should be accessible and disseminated as widely as possible and that this is the responsibility of researchers and sponsors alike.

There are several ways in which the GAO may be able to assure the GAO's coverage of "evaluations" in reports such as Federal Program Evaluations (U.S. General Accounting Office, 1976) can help to assure that the public is informed about the projects. The development of GAO guidelines on oversight can and should stress the need for dissemination, to assure public scrutiny of reports, and to facilitate reanlaysis by interested members of the research community. In cases where GAO oversight is intensive, as in an audit, verification of the existence and adequacy of reports should be routine. More generally, GAO can assist by encouraging the dissemination and reanalysis of reports by the wider community, in its discussions with agencies, in its staff presentations at professional meetings, and the like.

In addition, it is also reasonable to assure the availability of statistical data generated by the experiment along with the pertinent documentation. In particular, the data ought to be available for reanalysis.

Reanalysis can be justified on scientific grounds, since criticism and examination can help to better judge the quality of the project and to make informed recommendations for future research.

While the responsibility for assuring the availability of such "public access data files" should fall to the researchers and the sponsors, there is no uniform and well articulated policy as yet. Nonetheless, the committee believes that this is an issue which must be considered by the researchers and sponsors and by the GAO, in those cases where the contract, grant, or a law requires it. In the long run, the availability of such data files would be beneficial to the GAO, since GAO could then capitalize on a variety of competing analyses of the data in its own review of projects.

In most projects no special privacy problems are engendered by the release of statistical records. In particular, the deletion of identifiers of individual research participants will often be sufficient to protect privacy. In those cases where deductive disclosure is possible, the strategies developed to minimize or eliminate the problem can be used. They are catalogued in publications issued by the U.S. Census Bureau, and other agencies and new developments are reported often in the Proceedings of the American Statistical Association and similar sources.

The committee believes that GAO can assist in encouraging that statistical data be made available for reanalysis. It may do so by recognizing the value of competing analysis in its policy statements on oversight and in its guidelines and manuals on evaluating evaluations. It may do so, in the projects subjected to intensive GAO review, by verifying that statistical data which are supposed to be available to the sponsor or research community are indeed available and well documented. GAO may do so by

documenting the availability of data in its compendiums on Federal program evaluations. And it may do so by capitalizing on competing analyses done outside GAO when a project is subjected to major review.

SECTION VI

RECOMMENDATIONS

These recommendations were developed in order to facilitate the GAO's monitoring of the quality of social experimental research. The committee offers them with the understanding that any such monitoring must be designed so as to minimize the likelihood of disrupting the research or of making research a more costly and time consuming process. This theme is explicit in the Council's contract with the GAO. The following recommendations bear on specific actions as well as on the general role of the GAO, since -- in the committee's view -- the two are inseparable.

Alternatives to Reinterviews

A.1. The GAO should recognize the research tradition that underlies the social scientists' assumptions about the impact of unplanned events on experimental results (see references in Appendix II). Social scientists assume that it is difficult to anticipate the consequences that may result from an unplanned or unanticipated intervention -- including audits -- which are outside of the (experimental) study design. Furthermore, experience has taught social scientists that the resulting biases in the data are difficult to estimate and may be impossible to eliminate. This should be recognized in GAO manuals which guide the oversight process and in GAO policy governing oversight.

The current evidence on disruptions of research by unanticipated or unplanned interventions is sufficient to support the plausibility of this basic assumption. However, it is not sufficient to allow prediction of the direction and magnitude of biases in particular instances, such as those that might be introduced by GAO reinterviews.

A.2. The GAO should recognize that as a matter of policy and practice reinterviews of respondents in social experiments are often unnecessary.

The first and more crucial reason for this view is that quality, especially poor quality in research, will generally be obvious well in advance of any reinterviews. The existence and quality of the researcher's sample design and sampling procedures, the use of quality control procedures for surveys and data processing, the conduct and results of pilot tests are usually a sufficient basis for judging quality. They must be sufficient where responses in research are provided anonymously.

The second reason for the view is that results of reinterviews may be difficult to interpret or may be uninterpretable. For example, differences in the style of GAO and researcher's interviews, in the timing of the original and the reinterviews, in the auspices of the two interviews, etc. can be expected to produce differences in the answers participants give in the original interview and in the GAO reinterview. In addition, the respondents' behaviors may change over time. When discrepancies are found between the researcher's and the GAO's interviews then, they cannot be easily and directly ascribed to problems or deficiencies in the research interviews.

In those few cases where reinterviews are deemed essential, the conditions of the reinterviews should then match as closely as possible the conditions of the original interviews so that the results can be more interpretable.

A.3. The CAO should recognize that alternatives to reinterviews of research participants can be used to gauge the quality of data, while at the same time reducing the risks of disrupting the research. The alternatives considered in detail in the earlier remarks are:

(a) Restricting the range of topics for reinterview to those in which the GAO has a critical interest and trying to reach agreement with the researchers on those topics as early as possible.

(b) Engaging in parallel interviews by auditors of a sample which is independent (i.e., non-overlapping) of the program sample, but drawn from the identical target population and using the same interview methods as those of the researcher.

(c) Using surrogates for GAO staff for reinterviewing, whenever GAO's auspices per se may have a disruptive effect or when GAO staff do not have the appropriate experience, qualifications, or time for reinterviewing.

(d) Reinterviewing of a subsample of the researcher's sample. Subsample reinterviewing may result in a decrease in the subsample's cooperation in the research or otherwise influence its future responses. Since such a "contamination" of the subsample is possible, the overall research sample's size should be augmented so that contamination can be

assayed and so that the reinterviewed subsample can be discarded without damaging the main study if contamination is serious.

A.4. The GAO should initiate its review of some projects as early as feasible in the project's life cycle in order to maximize the utility and effectiveness of its oversight.

This recommendation is made for the sake of anticipating privacy-related problems as early as possible, and for the sake of enhancing the quality of GAO reviews more generally. In a sense, this recommendation extends and reiterates current GAO practice. It is intended to emphasize the importance of that practice. In making the recommendation, the committee realizes that early involvement in all projects is not feasible or desirable.

In some cases, early involvement may mean beginning at the RFP stage of the research process. In any event, the GAO should develop mechanisms for early involvement in consultation with sponsoring agencies and researchers. In this spirit, the GAO should be open to receiving questions and requests for advice tendered by researchers and sponsoring agencies.

This recommendation also stems in part from the recognition that in order to capitalize on some alternatives to reinterview, it is necessary for the GAO to make its interests and intentions known to the sponsors and researchers early in the research process. For example, if the GAO intends to reinterview a subsample, then the researchers need to know this so that at the design stage the size of the main sample can be augmented to accommodate the loss of the

reinterviewed subsample. Similarly, identifying the types of information that are most likely to be of interest to the GAO should occur at the design stage so that participants in the study can be made aware of the possibility of being audited, as a condition of their informed consent in agreeing to participate.

Testing for the Effects of Audits

A.5. The GAO should not mount a major testing program in-house. However, small-scale special purpose research projects may be warranted.

Although in the long term the understanding of the effects, problems, and solutions resulting from CAO oversight (e.g., reinterviews) need to be investigated and understood, the committee believes that a formal program of testing would demand a notable investment of time and manpower. Given that reinterviews by an audit agency will rarely be necessary and that reasonable alternatives exist to reinterviews, it is questionable whether the expenditure of resources to studying the consequences of reinterviews would be commensurate with the knowledge obtained.

(a) A small-scale program should include, at the very least, a systematic documentation of the GAO's experiences in both initial interviews and reinterviews. This would be helpful to researchers in their anticipating GAO involvement in the future and it would be useful in developing hypotheses about the consequences of GAO involvement.

(b) A small-scale program must capitalize on the advice of other agencies which have considerable experience in such methodological investigations (e.g., the Bureau of the Census, National Center for Health Statistics). These and other federal agencies have mounted strong research programs to test the effects of interviewing methods and style and to identify methods of reducing costs to and burden on respondents.

(c) A small-scale program might be targeted towards a particular research project or cl s of projects in order to develop an understanding of how the cooperation and behavior of respondents is affected by the GAO's intervention.

In this case, a sensible strategy would be to make plans in cooperation with the agency undertaking the social experiment so that tests of the effects of reinterview in particular, or of audits more generally, could be built into the social experiment. For example, such tests may involve augmenting the research sample used in the experiment so that a subsample could be assigned to the GAO for testing. Pilot testing of a variety of methods of reinter..ewing and/or informing respondents of reinterviews might be undertaken, for instance.

Role and Style of Operation

B. 6. The GAO should prepare a document clarifying its possible and likely roles in the evaluation of social experiments, and this document should be distributed widely among sponsors and researchers.

Such a statement would enhance communication with the research community and increase its understanding of the ways in which GAO involvement might help to improve research. The preparation of the document may serve to identify areas of unnecessary or unfeasible involvement and to sharpen the GAO's ability to allocate resources in this area. The statement would better enable sponsors and researchers to anticipate the occurrence of audits and the types of information that may be required by the GAO. This may lead to fewer misunderstandings and greater cooperation.

B.7. The GAO should coordinate its oversight activities as much as possible with those of other agencies.

Such coordination would help lessen the difficulty of doing field research in the sense that it would avoid repetitious and redundant demands upon sponsors, researchers, and monitors of social experiments. The current sporadic, and often simultaneous -- but uncoordinated -- reviews by the GAO and other agencies (e.g., OMB) place great demands upon the resources of sponsors and researchers in an inefficient manner. That is, sponsors and researchers may be forced to go over exactly the same ground with different review agencies.

B.8. Because formal scientific tests of social programs are a relatively new undertaking, even greater emphasis than usual should be placed on establishing and documenting solutions to problems in GAO reports, manuals, and policy.

The GAO's traditional emphasis on deficiency reporting is decreasing, and the committee wishes to encourage this trend towards a more evenhanded treatment of research projects. A better balance in the reporting of the "good" along with the "bad" will be more helpful to the Congress, the sponsors, and the researchers, as well as to the GAO in developing an understanding of the nature of social research. Furthermore, an emphasis on deficiency reporting is counterproductive in a research context where many problems cannot be anticipated, since such an emphasis leads to unnecessary disagreements between auditors and researchers. Finally, an equally vigorous interest in solutions to identified problems could enhance the quality of future research or the future implementation of programs.

B. 9. The committee recognizes and endorses the GAO's efforts to diversify and develop staff with experience and expertise in the review of social experiments. Accelerated development of staff and augmentation of staff in this area is essential for effective review, and the committee encourages acceleration.

This emphasis needs to be strengthened by mechanisms such as the following:

(a) More training of GAO staff in theory, methodology, and practice of social experiments is warranted.

Short courses are likely to be helpful, but insufficient. Intensive training at the graduate level would be more useful, especially if a leave policy permitting the training were instituted.

(b) The GAO should encourage its staff to participate actively in professional associations concerned with evaluation research, in the publication of its staff's work in professional journals, and to otherwise become an active member of the wider evaluation research community.

This would be beneficial to the development of the GAO staff's professional competency (in the area of evaluation research and social experimentation) and to the wider dissemination of the products of GAO reviews.

3.10. The GAO should take an active role in helping to develop guidelines on the management and budgeting of social experiments and their evaluation.

The GAO has considerable experience in examining the budgeting and management of research and that experience could be used to stimulate the development of better guidelines. In the committee's view, this is an area in which social researchers could benefit greatly from expert advice and guidance.

For example, the GAO might construct a document on budgeting similar to its paper on Evaluation and Analysis to Support Decision Making. The GAO might encourage other agencies with such expertise to pool resources with the GAO in joint efforts to improve current practices in social experiments and evaluations. In any event, the focus of the effort should be on guidelines (as opposed to hard and fast rules) for budgeting, procurement, staffing, etc. The guidelines should recognize the major differences between the problems of budgeting for social experiments and those for other types of research (e.g., notably hardware development, such as aircraft).

REFERENCES

Bernstein, I. and Freeman, H. Academic and entrepreneurial research. New York: Russell Sage, 1975.

Boruch, R. F. and Cecil, J. S. Assuring confidentiality and privacy in social research. Philadelphia, Pa.: University of Pennsylvania Press, 1978 (in press).

Campbell, D. T., et al. Confidentiality preserving modes of access to files and to interfile exchange for useful statistical analysis. In A. Rivlin (Ed.) Report of the National Academy of Sciences, Committee on Federal Agency Evaluation Policy: Protecting individual privacy in evaluation research, Washington, D.C.: NAS, 1975, Appendix A. Reprinted in: Evaluation Quarterly, 1977, 1(2), 269-299.

Carter, L. F. Federal clearance of educational evaluation instruments: Procedural problems and proposed remedies. Educational Researcher, 1977, 6(6), 7-12.

Cook, T. D., et al. Sesame Street revisited. New York: Russell Sage, 1975.

Dalenius, T. and Lyberg, L. Bibliography on nonsampling problems. Institute of Statistics, University of Stockholm, Sweden, 1975.

Duncan, J. W. and Haber, L. D. The survey review process in social research. Presented at the NSF Conference on Solutions to Ethical and Legal Dilemmas in Social Research, Washington, D.C., February 23-25, Washington, D.C., 1978. (Available from the authors, Statistical Policy Division, U.S. Department of Commerce).

Ferber, R. The reliability of consumer reports of financial assets and debts. Urbana: University of Illinois, 1966.

Havens, .. S. Statement before the Senate Committee on Human Resources on Expenditures, Problems, and Prospects of Management and Utilization of Program Evaluation. Washington, D.C.: U.S. Government Printing Office, October 6, 1977.

Hedrick, T. E., et al. On ensuring the availability of evaluative data for secondary analysis. Policy Sciences, 1978, in press.

Kruskal, W. Taking data seriously. In: Y. Elkana, J. Lederberg, R. K. Merton, A. Thackray, and H. Zuckerman (Eds.) Toward a metric of science: The advent of science indicators. New York: Wiley, 1978, pp. 139-169.

Kulley, A. M. The reliability and validity of survey measurement of health related variables: A research bibliography. Publication 8, Health Services Research and Training Program, Department of Sociology, Purdue University, West Lafayette, Indiana, 1974.

Lesser, J. T. A double sampling scheme model for eliminating measurement process bias and estimating measurement errors in surveys. Department of Biostatistics, University of North Carolina, Chapel Hill, Institute of Statistics Mimeo Series No. 949, August 1974.

National Center for Health Services Research. Advances in health survey research methods: Research Proceedings Series. Washington, D.C.: U.S. Department of Health, Education & Welfare, Public Health Services, Human Resources Administration, 1977 (DHEW Publication No. (HRA) 77-3154).

National Center for Health Services Research. Experiments in interviewing techniques: Field experiments in health reporting/1971-1977. Washington D.C.: U.S. Department of Health, Education and Welfare, Public Health Service, Health Resources Administration, 1977 (DHEW Document No. (HRA) 78-3204).

Nunnally, J. C. and Durham, R. L. Validity, Reliability and Special Problems in Measurement in Evaluation Research. In E. L. Struening and M. Guttentag (Eds.), Handbook of Evaluation Research, Vol. I Beverly Hills, California: Sage Publications, 1975, pp. 289-354.

Riecken, H. W., et al. Social experimentation: A method for planning and evaluating social programs. New York: Academic, 1974.

Sinaiko, H. W. and Broedling, L. A. (Eds.) Perspectives on attitude assessment: Surveys and their alternatives. Champaign, Illinois: Pendleton, 1976.

Sudman, S. and Bradburn, N. M. Response effects in surveys: A review and synthesis. Chicago: Aldine, 1974.

U. S. Census Bureau. Indexes to survey methodology literature: Technical paper #34. Washington, D.C.: U.S. Department Commerce, Social and Economic Statistics Division, 1974.

U.S. General Accounting Office, Standards for Audit of Government Organizations, Programs, Activities, and Functions. Washington, D.C.: U.S. GAO, 1974.

U.S. General Accounting Office, Evaluation and Analysis to Support Decision Making. Washington, D.C.: U.S. GAO (PAD-76-9/Revision of OPA 76-9), December, 1975.

U.S. General Accounting Office. Federal program evaluations. A directory for the Congress. (1976 Congressional Sourcebook Series, PAD-77-5), Washington, D.C.: U.S. Government Printing Office, 1976.

U.S. General Accounting Office (Program Analysis Division), Background Paper for Use by the SSRC Committee on Audit and Research on the Need to Access by GAO Auditors in the Audit of Social Research and Experiments. Washington, D.C.: U.S. GAO, mimeographed, April 8, 1977.

U.S. General Accounting Office, Report to Congress. Finding Out How Programs Are Working: Suggestions for Congressional Oversight. Washington, D.C.: U.S. GAO (PAD-83-3), November 22, 1977.

U.S. General Accounting Office. The Concorde -- Results of a supersonic Aircraft's entry into the United States. Washington, D.C.: U.S. GAO (CED-77-131), September 15, 1977.

U.S. General Accounting Office. An assessment of the Department of Housing and Urban Development's Experimental Housing Allowance Program. Washington, D.C.: U.S. GAO, (March 8) 1978.

Warner, S. L. Linear randomized resonse model. Journal of the American Statistical Association, 1971, 66, 884-888.

ACKNOWLEDGEMENT

The committee wishes to express its appreciation to William Kruskal, Katherine Lyall, Henry Riecken, and to Charles Nelson and his associates at the Rand Corporation for their valuable comments on an earlier version of this report.

APPENDIX I

Report Contents and Contract Requirements

This appendix relates sections of the committee's report to specific requirements of the SSRC-GAO contract (No. 7130078) of March 1977, as amended on March 10, 1977 and January 16, 1978. Page numbers in parentheses below refer to pages in the contract, and paragraph numbers (e.g., ¶3) refer to the following work statements from the amended contract.

¶1: Identifying jointly with GAO a sample of field experiments audited by GAO and other organizations over the past eight years;

¶2 Exploring with the principal investigators some of those projects and the possible positive and/or negative effects of those audits on their research results with a view toward identifying and analyzing alternative methods by which GAO might meet its legislated responsibilities for performing program audits or evaluation of social research and social experiments;

¶3 Directing staff in the conduct of evaluation of potentially less intrusive or more beneficial audit procedures for field research projects;

¶4 Preparing a report on the results of the work that will include discussion of the work, its findings, conclusions, and recommendations; and,

¶5 Giving consideration in its conclusions and recommendations to implications of recent Federal and State legislation affecting privacy of information and fair information practices, and to any conclusions and recommendations of bodies such as the Privacy Protection Study Commission, insofar as they are relevant to GAO's reviews of social experiments.

SECTION I: EVIDENCE ON THE IMPACT OF GAO AUDITS

It is relevant to ¶2 and ¶3. In addition to the literature summarized in Section I, Council staff in conjunction with GAO staff attempted to locate audits by the GAO and/or state audit agencies that would be relevant to the committee's concerns. These efforts resulted in the identification of only one case involving reinterviews: The Experimental Housing Allowance Program (EHAP). Council staff conducted several interviews and reviewed files at Abt Associates (Cambridge, Massachusetts), the Department of Housing and Urban Development (HUD - Washington, D.C.), the General Accounting Office (Washington, D.C.), the Rand Corporation (Washington, D.C.), and at the Urban Institute (Washington, D.C.) in order to review the audits of EHAP by the GAO and the Office of Management and Budget (OMB). In addition, the committee has considered the audit experiences of the New Jersey Graduated Work Incentive Experiment and the Educational Performance Contracting Experiment.

SECTION II: ALTERNATIVES TO REINVERVIEWS

In direct response to ¶3.

SECTION III: THE ROLE OF GAO IN SOCIAL EXPERIMENTS

This section is intended to provide general background and context to the committee's recommendations.

SECTION IV: SOME COSTS AND BENEFITS OF AUDITS

This section is pertinent to contract requirements generally and lays some of the basis for the committee's conclusions and recommendations.

SECTION V: ROLE AND RESPONSIBILITIES OF RESEARCHERS AND SPONSORS

Like the preceding section, it is pertinent to contract requirements in general.

SECTION VI: RECOMMENDATIONS

This is in response to ¶4, as is the entire report. In preparing the report, the committee has taken into account ¶5, with privacy of information considerations being discussed where relevant.

APPENDIX II

REPORT OF COUNCIL STAFF ON AUDITS OF E.H.A.P.

Committee on Problems in Evaluation Research

Interim Report

on

Audits of EHAP

July 15, 1977

Prepared by Ronald P. Abeles

***

The Department of Housing and Urban Development's (HUD) Experimental Housing Allowance Program (EHAP) has been reviewed or audited three times since its beginning in 1970: once by the Office of Management and Budgets (OMB) and twice by the General Accounting Office (GAO). The following report will briefly summarize the purpose and nature of these reviews and discuss the possibility of testing the impact of the latest GAO audit on EHAP.

The OMB Review

Within a year of the passage of legislation authorizing an experimental program to test the feasibility of housing allowance, OMB instituted a review of EHAP that for several months blocked the implementation of the experiments. The review commenced in August 1971 and officially ended in October 1972. The Director of OMB (George P. Shultz) requested an abeyance of all but the Demand Experiment. This abeyance effectively stopped work on EHAP from progressing beyond the design stage during the period of March 29, 1972 until October 3, 1972.

The OMB review consisted mostly of discussions between HUD and OMB staff that centered around design issues and the implications of an eventual implementation of a national housing allowance program. OMB

directed a series of questions about EHAP to HUD on 3 occasions: in August 1971, March and July 1972. The first set of questions dealt with a series of mainly macro-economic questions about the design, price effects (e.g., price and income elasticities for the total housing market), budgetary impacts. The questions were quite general and broad, and they were accompanied by a request for a written report within six months. Many of the questions probably could not be answered properly until the completion of the experiments, which indicated a misunderstanding of the experimental program by OMB.

However, by the time of the second set of questions, this misunderstanding seemed to have been cleared up. These questions were also broad ranging, but addressed a variety of issues concerning the general rationale and need for a housing allowance experiment. In particular, the questions evidenced a concern with how the experiment would fit in with existing HUD programs and how a housing allowance program would fit in with President Nixon's proposals for welfare reform. This latter concern may have been a key consideration behind the review and the order to hold off implementing the experiment. In Shultz' letter of March 29, 1972 calling for the abeyance of EHAP, Shultz cited the following concerns:

- The experiment appears to a demonstration or pilot program rather than an experiment. He was anxious about the program laying the groundwork for an operational program and establishing a fait accompli;
- The budget costs of an operational program would be enormous and not fit into planned Federal budgets in the forseeable future;
- The desirability of earmarking funds for housing allowances goes against the proposed Nixon welfare reform of doing away with earmarked funds for welfare in general.

In closing his letter, Shultz commented that

"...my immediate concern is that the scale of the demonstration

> program together with the extent to which it deals with technical issues will leave many with the impression that a decision has been made to go in the direction of housing allowances, when no such decision has actually been made."

In the continuing discussions of EHAP, a variety of documents about the design and proposed implementation of the experiment were sent to OMB and discussed in person with OMB staff. Throughout this period, OMB continued to wonder whether the experiment was necessary and whether or not answers to EHAP experimental questions might not be obtainable through secondary analyses of existing data. The third set of OMB questions, which were sent to HUD in July 1972, raised this issue again and asked about whether the design would consist of the right mix of experimental sites to permit generalization to the national housing market, how EHAP fit with current HUD programs of housing assistance, and about the anticipated statistical accuracy and confidence of experimental results.

Finally, on August 31, 1972, OMB lifted its abeyance request by leaving the final decision of whether or not to proceed to Secretary George Romney (of HUD). In this letter, OMB stated its reservations about the reliability of the conclusions that EHAP would be able to provide and about the "timeliness of these conclusions with respect to the policy debate over housing allowances." OMB was concerned that EHAP (1) will not monitor systematically changes in the supply of new housing throughout entire housing markets; (2) will not be able to assess the impact of housing allowances in the absence of other subsidized housing programs, since these will be allowed to operate normally during the experiment; (3) will not be able to shed light on landlords' and investors' behaviors during a permenant housing allowance program -- OMB suggested that 5 to 10 years is too brief of a period for the impact of housing allowances to be felt;

(4) will not be able to generalize from just two SMSAs of 250,000 people; (5) will not be able to separate out the variety of causes influencing housing supply and costs in the Supply Experiment; and (6) will not be able to produce usable information soon enough to help in policy decisions. In closing, OMB recommended the stopage of the Supply Experiment, but the continuation of the Demand and Administrative Agency Experiments.

HUD chose to proceed with all three components of EHAP, and so informed OMB on October 3, 1972. Except for the occasional transmission of documents to OMB, this ended the review of EHAP by OMB.

The GAO Reviews

The First Review

The first GAO review of EHAP consisted mostly of an evaluation of the overall design of the experiments and a summary of the housing allowance concept and of previous housing allowance programs in the U.S. and abroad. The review process started (1/72) before the final design of EHAP had been completed and before contractors had been selected. For awhile, both the GAO and OMB audits were taking place simultaneously, although there appears to have been little to no exchange of information between GAO and OMB. By the completion of the first review in September 1973 (date draft reports were released for comments), all three EHAP experiments were beyond the design stage and were at various points of operationalization of the designs: the Administrative Agencies Experiment (AAE) was enrolling participants at all eight sites; the Demand Experiment was in the midst of its first year of operation; and the Supply Experiment was screening occupants of housing units in Green Bay and attempting to finalize negotiations in the selection of a second experimental site.

The purpose of the GAO review seems to have been to (1) evaluate the general design of the experiments; (2) review the operation of the contractors and HUD during the start-up phase; (3) make projections of the costs of the experiment over its anticipated lifespan; (4) review the concept and background of housing allowances. The audit consisted of interviews and discussions with personnel at HUD, Rand, Abt, Urban Institute, and at site offices of the experiments. The auditors received numerous documents spelling out the design of EHAP and procedures for implementing the design. The audit was conducted by Charles Stokes -- an urban economist at the University of Bridgeport -- who was a Faculty Fellow at GAO from about June 1972 until July 1973. This makes the initial review of EHAP an atypical GAO audit.

The relationship between HUD and GAO in regard to EHAP got off to a rather rocky start in the first review. In early 1971 the EHAP staff was quite small and consequently was heavily burdened during the start-up phase of the experiments. Thus, when GAO requested to set up appointments to discuss EHAP, the appointments were frequently postponed or shortened by HUD in order to deal with more pressing and relevant matters (from HUD's point-of-view). This lead GAO to perceive a lack of cooperation with their review, which tainted the atmosphere of the review for some time. Eventually, this scheduling problem was overcome, and frequent and regular meetings between HUD and GAO were held.

The Second Review

Near the beginning of the first review of EHAP, GAO indicated its intention of probably reviewing the operation of EHAP at a later date. That second review commenced with the transmission of several EHAP annual reports to GAO in February 1975. The scope of the second review was much

greater than the earlier audit in that it went beyond questions of design to include an evaluation of the actual implementation of the design and reinterviews of recipients of housing allowances at three of the experimental sites.

GAO audits consist of three phases of operation: Survey, Review, and Reporting phases. The survey phase consists of a general collection of background materials, discussions with personnel, observation of procedures, and like in an attempt to obtain general information about the audited project and to discover problem areas that might deserve more intensive investigation during the review phase. Thus, the survey phase is a reconnaisance expedition guided by very general questions. During this phase the auditors themselves are not positive about what they are looking for. If they find nothing that "looks funny," then the audit might well stop here. However, if problems come to their attention, then more intensive, directed probing occurs during the review phase.

During the survey phase, GAO and HUD (with its contractors) immediately collided over direct access to recipients of housing allowances. GAO believed that it was necessary to interview participants and to inspect their housing in order to test whether procedures and guidelines were being properly administered. In short, GAO wanted to investigate whether the "right" people were receiving housing allowances and whether their housing was indeed above standards set by HUD. From GAO's perspective, no amount of reviewing procedure manuals or of observing the processing of data would be sufficient to answer this basic question. In order to confirm the quality of the data, it was necessary to go straight to the source of the data (i.e., respondents), according to GAO. However, HUD and its contractors expressed deep concern about this violation of pledges of confidentiality

made to participants and about the subsequent consequences that this would have for the validity of the experiments themselves.

GAO and HUD did not have too much difficulty in arriving at arrangements for GAO to observe and evaluate the processing of data. GAO representatives were able to visit EHAP sites in Green Bay, Wisconsin; Salem, Oregon; Pittsburgh, Pennsylvania; Jacksonville, Florida and to observe their operations. In addition, GAO staff visited the central offices of Rand and Abt in order to review operations there. Here they were able to trace the processing of individuals' data, but with identifiers removed (i.e., they could follow a single person's file through data processing procedures without knowing who the person was). This was a procedure that was satisfactory to both GAO and HUD.

However, the impass over direct access to individuals (i.e., face-to-face interviews and review of their files) continued from about April 1975 until about June 1976. HUD and its contractors offered various compromises, including face-to-face interviews by a third-party (e.g., a CPA firm). During these negotiations, the general open-endedness of the survey phase of the GAO audit proved frustrating to the experimenters. From their perspective, they could not get a "straight answer" from GAO when they asked what it was that GAO wanted to see. A definitive answer was not possible, since GAO auditors did not know exactly what would be of interest to them. They had broad guidelines, but only experience with EHAP would provide them with specific questions and leads to follow. However, while aware of the potential hazards to the experiment, GAO continued to believe that direct access to participants was essential to its audit's integrity. In addition, GAO auditors argued that an audit would be part of the "real

world" of housing allowances, should the program ever be implemented. Thus, GAO staff reasoned, shouldn't audits be part of the experimental environment as well? Finally, GAO believed, from its past experiences, that its interviews would not be disruptive, since people distinguish GAO audits from other types of investigations by other governmental agencies.

Eventually, a compromise was developed that allowed GAO auditors to interview participants who had given prior permission to be reinterviewed. Samples of 100 participants were selected in Salem, Green Bay, and Pittsburgh. These individuals were invited by letter to be interviewed. Those consenting were eventually interviewed and their housing was inspected. GAO was not given the names of individuals who declined to be interviewed. During these interviews, GAO asked questions about the participants' amount and source of income and evaluated the condition of their housing unit. GAO didn't use a standard questionnaire, and each auditor team asked its own set of questions. (Each auditor team had received training from EHAP's local staff.) In most instances, GAO's questions were limited to ones that EHAP asked of its respondents, but at one site these questions were supplemented by others (i.e., instead of only asking an open-ended question about sources of income, GAO included a checklist as well).

By the time these reinterviews occurred, the danger of their impacting on EHAP had decreased considerably. Both the Demand and the Administrative Agencies Experiments were in their termination phases. Both had completed their collection of experimental data prior to the GAO interviews. At the time of GAO interviews, participants were being transferred from the housing allowance programs to other housing assistance

programs. In effect, the experiments were completed, and the housing allowance programs were nearly finished.

In the case of the Supply Experiment, the situation is slightly different. The Supply Experiment's sources of data consist of (1) samples of landlords, tenants, and homeowners tied to particular housing units (i.e, the sampling unit is the dwelling unit) and (b) the files and records of the housing allowance offices in Green Bay and South Bend. The Supply Experiment is scheduled to continue for another year (into 1978) and the housing allowance program itself will continue for another five years beyond the experiment's termination. Now, GAO's sample consisted of 100 participants (i.e., recipients) in the housing allowance program. Very few of the recipients (2900 households) in the housing allowance program are included in the samples of landlords, tenants, and homeowners. Thus, next to none of these individuals fell into the GAO sample. In short, the population of the GAO sample is the recipients of housing allowances; the population of the Rand Supply Experiment sample is the housing units with their associated tenants and owners. Thus, the experimental samples cannot be "contaminated" by the GAO interviews, except by some generalized indirect effects (e.g., negative publicity in the newspaper; and there does not seem to have been anything like this).

However, the Supply Experiment might afford the opportunity for a limited test of the impact of the GAO interviews. It may be possible to compare changes in the characteristics of the GAO sample with changes in those of the Rand experimental population. That is, the characteristics of the GAO interviewees is known as of the date of the interviews (July 1976) and they can be ascertained as of now. Similarly, the characteristics of the 2900 households enrolled to receive housing

allowances is known as of July 1976 and can be ascertained for now. One might compare changes in these groups over time in terms of (a) reported income and sources of income, (b) condition of housing, (c) participation rates in EHAP, (d) household size and composition, and (e) reasons for dropping out of EHAP. The meaningfulness of this analysis would be complicated by the fact that the GAO interviewees are a self-selected group to some extent (i.e., 67% of the 100 agreed and were actually interviewed). While in theory this analysis might be possible, it is difficult to predict how easily the data could be obtained and how meaningful any comparisons would actually be.

DRAFT of 7/77

CHRONOLOGY OF EHAP

DATE	HUD & Contractors	Review Agencies (GAO & OMB)
1967-1968	President's Committee on Urban Housing (Kaiser Committee) recommends initiation of an experiment to test housing allowances.	
1969	Urban Institute (UI) and HUD discuss housing allowances in general and design considerations of a housing allowance experiment. UI prepares various working and discussion papers for HUD.	
1970	Passage of Housing and Urban Development Act of 1970, which contains provisions for a housing allowance experiment (Section 504).	
March 1971	UI begins work to develop initial design documents for a set of housing allowance experiments. Objectives of experiments specified by HUD and UI.	
August 1971		OMB officially requests HUD to analyze and report to OMB on several questions prepared by OMB concerning EHAP. OMB raises questions about general design, necessity for experiment, relationship to other housing programs, potential effects on housing markets, and budgetary matters.
September 1971		HUD replies to OMB's requests by sending documents and meeting with OMB staff.
November 1971	Request for Proposals (RFP) is issued based on UI's working papers on design considerations. RFP specifies a single experiment, leaving open question of whether questions about impact on markets and inflation can be tested via experiments.	GAO staff first start receiving background materials on EHAP. Routine exchange of information with resident GAO staff.

DATE	HUD & Contractors	Review Agencies (OMB & GAO)
Novemer 1971	Rand Corporation subcontracted with UI to conceptualize an experiment dealing with market consequences of housing allowances. This eventually becomes the Supply Experiment (SE). Discussions with Secretary Romney (HUD) result in conceptualization of tests for administration of housing allowance programs. This becomes the Administrative Agencies Experiment (AAE).	
January 1972		GAO attempts to meet with HUD staff to obtain background information. Difficulties experienced in scheduling meetings results in letter of complaint from GAO to HUD.
February 1972		Design documents prepared by UI sent to OMB.
March 1972	HUD signs contracts with Stanford Research Institute (SRI) for the design phase of the Demand Experiment (DE) and with Abt Associates for the design phase of AAE.	HUD sends GAO copies of proposals received in response to RFP. OMB sends HUD a detailed list of questions about EHAP. After discussions resulting from this list, George Shultz (Director, OMB) requests abeyance of all but DE. This effectively stops work on EHAP at the design phase as of March 29, 1972.
April 1972	HUD and contractors start site selection process for all three experiments. Work begins on preparation of operations manuals, survey instruments, and final experimental designs. HUD signs contract with Rand for the design phase of SE. (Sole source contract)	Copies of SRI, Abt, and Rand contracts made available to GAO. HUD prepares counter-arguments to issues raised by OMB; meetings between OMB and HUD staffs; design documents sent to OMB.

Date	HUD & Contractors	Review Agencies (OMB & GAO)
June 1972	HUD selects Allegheny County (Pittsburgh) as Site I for DE.	HUD sends design documents (UI papers) to GAO. HUD sends OMB preliminary design documents on SE. Telephone discussions between HUD and OMB concerning AAE.
July 1972	HUD and Rand visit prospective sites for SE (Site II). Screening interviews are begun in Pittsburgh (completed November 1973): 50,000 families.	Two general orientation meetings between HUD and GAO. GAO officially informs HUD of intention to review EHAP during its start-up phase in order to prepare a report to Congress by January 1973. GAO also indicates that an evaluation of EHAP during its operational phase would most likely occur at a later date. Questions raised about site selection, need for the experiment, suitability of existing data for secondary analyses. Discussions between HUD and OMB about design of experiments (e.g., ability to generalize from limited number of sites; potential impact of housing allowances on housing markets; fit with existing HUD programs). SE design reviewed with HUD expresses concern about impact of further delays on the experiments.
August 1972		Two meetings between GAO and HUD to discuss general questions about EHAP and GAO visits to contractors. OMB officially states doubts about the usefulness and advisability of EHAP, especially in regard to SE. OMB advises HUD to stop the SE, but leaves the final decision to HUD (Romney). Letter of August 31, 1972.
September 1972	HUD sends letters to proposed agencies inviting participation in AAE. 5 agencies receive letter in September and 3 agencies in March, 1973.	

Date	HUD & Contractors	Review Agencies (OMB & GAO)
October 1972		HUD (Romney) informs OMB of decision to continue all three experiments and attempts to rebut OMB's criticisms. This ends OMB review of EHAP, although HUD continues to send documents to OMB.
November 1972	HUD and Rand designate Saginaw County, Michigan as Site II in SE, but encounter difficulties in obtaining permission from suburban jurisdictions. Year of fruitless negotiations follows.	GAO reports to HUD the results of its evaluation of the competence of HUD's contractors. Evaluations based on visits to headquarters of the contractors.
December 1972	Abt Associates replaces SRI as contractor for Demand Experiment. HUD tentatively designates Green Bay (Brown County), Wisconsin as Site I for SE.	
January 1973		GAO visit EHAP sites to evaluate progress and capabilities. Familiarize themselves with housing conditions at sites; assess relationships between EHAP and local communities. GAO and HUD discuss funding of EHAP and definitions of eligibility to receive housing allowances.
March 1973	Abt submits Final Evaluation Design for DE and starts baseline interviews in Pittsburgh; completed in January, 1974. AAE enrollment period starts; ends May, 1974. Operation phase of SE starts with opening of Rand office in Green Bay.	
April 1973	Baseline interviews for DE started in Phoenix; completed in December, 1973. First AAE survey of participants.	
May 1973	Rand submits draft of research design for SE.	

Date	HUD & Contractors	Review Agencies (GAO)
June 1973	Peer group review of final design for DE.	GAO drafting EHAP Report No. 1; completed late July, 1973 and undergoes inhouse reviews.
August 1973	Screening survey of occupants of housing units in Green Bay for SE; completed in October, 1973.	
September 1973	Operation procedures being developed and reviewed at all experimental sites. All eight AAE sites are enrolling participants.	GAO sends draft of EHAP Report No. 1 to HUD and contractors for comments. Correspondence between HUD and contractors regarding report.
October 1973	Housing Allowance Office (HAO) incorporated in Green Bay. Baseline survey of housing units starts; completed in December, 1973.	
November 1973	First periodice interviews start in Pittsburgh and Phoenix for DE; completed in November, 1974.	
December 1973	Baseline survey of landlords, tenants, and homeowners started in Green Bay for SE; completed in April, 1974.	
January 1974	AAE starts paying housing allowances.	Official HUD comments on EHAP draft report sent to GAO.
March 1974	Completion of enrollment operations in DE.	Revised and final EHAP report delivered to HUD.
April 1974	South Bend (St. Joseph County), Indiana is designated as SE Site II.	HUD repares rebuttal to GAO report for public distribution.
May 1974	Second periodic interviews begin at DE sites.	

Date	HUD & Contractors	Review Agencies (GAO)
June 1974	Enrollment of housing allowance participants begins in Green Bay for SE. Analysis review meeting held between Abt and HUD regarding DE.	HUD complains to GAO that its original comments on EHAP Report No. 1 were not reprinted in their entirety in the revised report; also HUD raises new criticisms of report.
July 1974	HAO is incorporated in South Bend as Site II of SE. Screening interviews of occupants of housing units starts; completed in September, 1974.	
September 1974	"Windshield" survey of housing and neighborhoods in South Bend; completed in November, 1974.	
October 1974	Green Bay HAO begins outreach program.	
November 1974	Baseline surveys of landlords, tenants, and homeowners start in South Bend for SE; completed in June, 1975.	
December 1974	Panel of residential buildings and housing units selected in Green Bay for SE. South Bend HAO begins enrollment of participants.	
January 1975	Green Bay wave 2 of tenants and homeowners begins; completed in September, 1975.	
February 1975		First GAO request for documents since end of first review. HUD sends annual reports. By May, 1975 over 50 documents are sent to GAO.
March 1975		HUD and GAO prepare for "entry interview" prior to formal review of EHAP.
April 1975	Baseline survey of residential buildings in South Bend for SE; completed in July, 1975. Green Bay Wave 2 survey of landlords starts; completed September, 1975.	Confidentiality of data raised with GAO.

Date	HUD & Contractors	Review Agencies (GAO)
May 1975	Third periodic interview wave started in Pittsburgh; completed in March, 1976. Termination procedures started at the AAE sites; completed June, 1976.	Briefing session for GAO is held by HUD to provide overview of EHAP and of GAO's concerns. More documents given to GAO. Rand raises questions with HUD about GAO access to records and to recipients; questions of confidentiality raised. HUD staff meet with GAO staff to discuss growing conflict over access to data. HUD requests contractors to prepare materials specifying the need for confidentiality and potential effects of not honoring confidentiality pledges. Abt refuses to release confidential data to GAO.
June 1975	Green Bay HAO begins second year of open enrollment, first annual recertification cycle, and first annual housing reevaluation cycle.	GAO requests that HUD instruct contractors to cooperate with GAO review. Without denying its authority to access data, GAO temporarily agrees to review files purged of participant identifiers. GAO indicates intentions to review operations in Pittsburgh, Green Bay, and Salem. HUD sends letter to contractors requesting their cooperation with GAO to the extent that it does not violate confidentiality pledges. Documents on the Integrated Analysis by UI sent to GAO, along with other technical reports from Rand and Abt.

Date	HUD & Contractors	Review Agencies (GAO)
July 1975	Third periodic interview wave starts in Phoenix for DE; completed in March, 1976. Last AAE survey of participants is completed.	HUD distributes copies of background document ("Working Notes") on EHAP and GAO, which contains a summary of EHAP's design and operations, GAO's concerns, speculation on consequences of GAO having direct access to recipients. 50 more technical reports sent by UI to GAO. 30-40 more reports sent to GAO by Rand and Abt. GAO staff visits Rand, Santa Monica & Green Bay, to discuss the Supply Experiment. Questions about design and operation of SE. GAO site visit to Abt in Cambridge & Salem occurred sometime between May and July, 1975. Pittsburgh & Jacksonville visited, too.
August 1975	South Bend HAO begins active outreach to attract participants. Green Bay Wave 2 survey of residential buildings; completed end-October, 1975.	Consultant to GAO delivers memorandum to GAO discussing GAO audit responsibilities and confidentiality.
September 1975		Meeting between GAO and HUD and its contractors to discuss direct GAO access to data. Attempt to work out a compromise, which fails. GAO officially requests access to data with identifiers and cooperation in arranging reinterviews and inspection of housing units. GAO states these data are necessary for its audit, cites relevant, laws, and says will arrive in six days to start its audit. HUD, Abt, and Rand refuse to allow such access to GAO. HUD indicates willingness to cooperate, but not to supply identifiers.

Data	HUD & Contractors	Review Agencies (GAO)
October 1975	Preparations begin for Wave 2 surveys in South Bend for SE.	Rand informs HUD that GAO audit has cost them sofar $50,000.
	Preparations begin for Wave 3 surveys in Green Bay for SE.	Comptroller General (Staats) writes to Secretary of HUD (Hill) asserting GAO's right of access and complaining about lack of cooperation. Demands access to recipients, and claims GAO was never consulted about confidentiality clauses in contracts.
	First semiannual recertification of housing allowance recipients is started by South Bend HAO.	GAO staff visit South Bend HAO to review operations (10/6-8/75).
November 1975		Secretary of HUD (Hill) responds to Comptroller General's October letter by reemphasizing HUD's concerns about confidentiality and consequences of GAO review, cites history of prior contacts with GAO, and requests meeting to resolve differences.
		Correspondence between HUD and GAO stating positions on confidentiality issues.
December 1975		GAO requests data about participants (without identifiers) from Abt for DE. Data sent by Abt with warning about misinterpreting the data.
		HUD and GAO meet to work out a compromise to allow direct access to participants, but only after they have been given prior permission to be interviewed.

Date	HUD & Contractors	Review Agencies (GAO)
January 1976 to April 1976	Wave 3 survey of tenants and homeowners starts in Green Bay and South Bend (SE); completed in July, 1976. End of data collection on agencies' performance records in AAE. (January, 1976)	Negotiations between HUD and GAO on the implementation of the compromise for access to participants. Numerous meetings and memoranda. Disagreements about wording of letter asking for permission of respondents for reinterviews; contractors seeking pledges of confidentiality of data from GAO and the firm of certified public accountants (CPA), which will draw the samples and correspond with respondents.
March 1976	End of last periodic interviews of DE participants in Pittsburgh and Phoenix.	Letters of agreement and workstatements on procedures exchanged between HUD and GAO.
April 1976	Wave 3 survey of landlords in Green Bay for SE; completed mid-August, 1976. South Bend starts second year of enrollments, first annual recertification cycle, and first annual housing reevaluation cycle.	HUD, contractors, and GAO meet to arrange GAO site audits at Salem, Pittsburgh, and Green Bay, which are to include reinterviews of participants. Visits scheduled for July, 1976.
May 1976	End of HAO operations in AAE. South Bend HAO begins direct mail advertising to attract participants (SE). Wave 2 survey of landlords of rental properties begins; completed end of August, 1976.	GAO conference on "Social Experimentation" held in Washington, DC to discuss the general issues regarding access to participants. HUD, contractors, and outside consultants (including SSRC) participate.
June 1976	DE starts terminating participants.	Abt and Rand prepare staff for GAO visits.
July 1976	All control group participants terminated from DE. Pretesting of Wave 3 survey of tenants and landlords in South Bend (7/16-23/76) for SE. Pretesting of Wave 4 survey of tenants and homeowners in Green Bay (7/19-23/76).	GAO field representatives and FHA housing inspectors receive background information and training from Abt in preparation for reinterviews and housing inspections (Salem AAE and Pittsburgh DE). Samples drawn (7/12/76) and reinterview appointments set (7/26/76). Reinterview sample drawn for Green Bay SE (7/23/76).

Date	HUD & Contractors	Review Agencies (GAO)
August 1976	South Bend HAO begins billboard ads.	GAO reinterviews of participants begin (8/3/76) in Pittsburgh DE; completed on September 9, 1976 with a response rate of 57%. GAO reinterviews at Salem AAF begin on August 10 and are completed on September 3, 1976 with 80% response rate. GAO reinterviews in Green Bay commence on/about August 23 and are completed about September 26, 1976 with response rate of 67%. All GAO reinterviews are limited to questions about eligibility to receive housing allowances (i.e., income amount and sources and condition of housing).
September 1976	Wave 3 of survey of tenants, landlords, and homeowners in Green Bay SE nears completion.	Green Bay HAO discusses cases of disagreement in evaluation of housing standards (for respondents in reinterviews).
October 1976		Green Bay SE sends GAO case-by-case reasons for participants refusals to be interviewed.
November 1976		GAO submits draft copy of EHAP Report No. 2 to HUD. States that the report only covers activities preceding the resolution of the confidentiality conflict. It reviews only design and operations of EHAP, but does not include data from reinterviews. Copies of relevant portions of draft report are sent by GAO to contractors, who subsequently receive full report from HUD: 11/29/76.
December 1976		HUD and contractors send criticisms of draft report to GAO (late December & early January, 1977).

Date	HUD & Contractors	Review Agencies (GAO)
January 1977	Wave 4 interviews of landlords, renters, and homeowners commences in Green Bay.	GAO issues revised draft report.
February 1977	Last participants in DE terminated.	GAO debriefing on visits to AAE and DE sites. Oral presentations and discussions with Abt staff.
March 1977		GAO staff discusses Integrative Analysis with UI staff.

APPENDIX III

References to Research Cited in the Text on Confidentiality of Respondent Records in Social Research

Illustrative Cases

Boruch, R. F. Educational research and the confidentiality of data: A case study. Sociology of Education, 1971, 44, 59-85.

Dalenius, T., & Klevmarken, A. Personal integrity and the need for data protection research in the social sciences. Stockholm: Swedish Council for Social Science Research, 1976.

Kershaw, D. N., & Small, J. C. Data confidentiality and privacy: Lessons from the New Jersey Negative Income Tax Experiment. Public Policy, 1972, 20(z), 258-280.

Oyen, O. Encounter with the image of sociology. Sociologiske Meddelser, 1965, 10, 47-60.

Spergel, I. A. Community action research as a political process. In I. A. Spergel (Ed.) Community organization: Studies in social constraint. Beverly Hills, CA: Sage, 1969, pp. 231-263.

U.S. Senate, Committee on Labor and Public Welfare. Hearings on S.3835: Comprehensive alcohol abuse and alcoholism treatment and rehabilitation act of 1970. 91st Congress, 2nd Session, Part 2, 1970.

Walsh, J. News and Comments. Science, September 10, 1969, 1244.

Walsh, J. Antipoverty R & D: Chicago debacle suggests pitfalls facing OEO. Science, 1969, 165, 1243-1245.

Illustrative Surveys

Reuter, J. M., Richards, J. A., & Hartsock, S. L. The prevalence of drug abuse in the army: A comparison of urinalysis and survey

rates. HUMRO Technical Report 15-17. Arlington, Va: Human Resources Research Organization, June 1975.

Krotki, K. J. & Fox, B. The randomized response technique, the interview, and the self-administered questionnaire. Proceedings of the American Statistical Association: Social Statistics Section, Washington, D.C.: ASA, 1975, pp. 21-27.

National Academy of Sciences, Committee on Federal Statistics. Report of the Panel on Privacy and Confidentiality as Factors in Survey Response. Washington, D.C.: NAS, 1978.

Illustrative Experiments

Goodstadt, M. S., & Gruson, V. The randomized response technique: A test on drug use. Journal of the American Statistical Association, 1975, 70, 814-818.

Locander, W., Sudman, W., & Bradburn, N. An investigation of interview method, threat, and response distortion. Proceedings of the American Statistical Association: Social Statistics Section, Washington, D.C.: ASA, 1975, pp. 21-27.

National Academy of Sciences, Committee on Federal Statistics. Report of the Panel on Privacy and Confidentiality as Factors in Survey Response. Washington, D.C.: NAS, 1978.

Singer, E. Informed consent procedures in surveys: Some reasons for minimal effects on response. Presented at the Conference on Solutions to Ethical and Legal Problems in Applied Social Research, Washington, D.C., February 23-25, 1978. (Available from E. Singer, Columbia University, New York, New York).

Zdep, S. M., & Rhodes, I. N. Making the randomized response technique work. Public Opinion Quarterly, 1977, 41, 531-537.

APPENDIX I V

Selected General References on Social Experiments

Adams, S. Evaluative research in corrections: A practical guide. Washington, D.C.: U.S. Department of Justice, Law Enforcement Assistance Administration, 1975.

Bennett, C. A., & Lumsdaine, A. A. (Eds.). Evaluation and experimentation. New York: Academic Press, 1975.

Boruch, R. F., & Riecken, H. W. (Eds.). Experimental testing of public policy: Proceedings of the 1974 Social Science Research Council Conference on Social Experiments. Boulder, Colorado: Westview Press, 1975.

Fairweather, G. W., & Tornatsky, L. G. Experimental methods for social policy research. New York: Pergamon, 1977.

Geismar, L. L. Thirteen evaluative studies. In E. J. Muller, J. R. Dumpson, and associates (Eds.). Evaluation of social intervention. San Francisco: Jossey-Bass, 1972.

Glaser, D. Routinizing evaluation: Getting feedback on effectiveness of crime and delinquency programs. Rockville, Maryland: National Institute of Mental Health, 1973.

Lyons, G. M. (Ed.). Social research and public policies: The Dartmouth OECD Conference. Hanover, New Hampshire: Public Affairs Center, Dartmouth College, University Press of New England, 1975.

Reicken, H. W., Boruch, R. F., Campbell, D. T., Glennan, T. K., Pratt, J., Rees, A., & Williams, W. Social experimentation: A method for planning and evaluating social programs. New York, N.Y.: Academic Press, 1974.

Rivlin, A. Systematic thinking for social action. Washington, D.C.: Brookings Institution, 1971.

Sechrest, L. (Ed.). Emergency medical services. Washington, D.C.: U.S. Government Printing Office, 1978 (in press).

Suchman, E. A. Evaluation research. New York: Russell Sage, 1967. Chapter I, The Evaluation Research Design.

Weiss, C. H. Evaluation research: Methods of assessing program effectiveness. Englewood Cliffs, New Jersey: Prentice-Hall, 1972.

APPENDIX V

ANNOTATED LIST OF GAO AUDITS

INVOLVING INTERVIEWS AND/OR REINTERVIEWS

OF PARTICIPANTS

M E M O R A N D U M

TO: Committee on Evaluation Research

FROM: Adrienne Armstrong and Ronald Abeles

RE: Interviews and reinterviews conducted by GAO

Copies of GAO reports were collected from the library of Northwestern University, Government Publication room, and from the Social Science Research Council. The library began receiving reports in 1969. However, the number of reports received was minimal. Most early reports were on foreign aid and health care matters. Not until 1974-75 did the Government Publication room begin to receive reports on a consistent basis. Not all reports have been received. Thus, the illustrations listed below were assembled from GAO reports dating from 1969 to the summer of 1977. The majority of these reports were taken from post-1975 materials.

According to the GAO reports, interviews were conducted during each of the project evaluations cited. However, it is often unclear from the reports whether or not these were interviews or reinterviews. In addition, the GAO reports do not state consistently whether or not the agencies audited conducted interviews prior to the GAO audit.

April 22, 1977
(rev. July 13, 1977)

1. FOLLOW THROUGH: LESSONS LEARNED FROM ITS EVALUATION AND NEED TO IMPROVE ITS ADMINISTRATION (MWD-75-34; 10/7/75)

Follow Through was an experimental program designed to find more effective approaches to teaching young children from low-income families. Intended to "follow through" on results of Head Start experiences.

Local education agencies (LEAs) were required to select one of fourteen different educational programs or approaches. The approaches were developed primarily by colleges, universities, and private educational research organizations. These institutions (sponsors) contracted with the Office of Education (OE) and LEAs to provide curriculum materials, teacher training, and other assistance. Program lasted 7 years, including 6 years of evaulation studies.

Scope of GAO Review: Made at OE headquarters in Washington, DC, and at nine selected project sites. Reviewed OE policies and procedures, project applications, initial results of a national evaluation of Follow Through. Interviewed Federal, State, and local officials responsible for administration and operation. Observed classroom activities and interviewed selected teachers and parents of Follow Through enrollees.

Major GAO criticisms: (a) lack of random assignment of LEAs to sponsors (and thereby to "treatments") and (b) lack of control groups.

Evaluations conducted by Stanford Research Institute, Menlo Park, California.

2. PRELIMINARY COMMENTS ON THE NEW JERSEY GRADUATED WORK INCENTIVE EXPERIMENT. June 1970

Experimental program designed to assess the impact on work effort (employment experiences) of 8 plans of "negative income tax" and "guaranteed income." Plans reflect combinations of a "guarantee" and "benefit reduction rate." Experiment initiated by the Office of Economic Opportunity (OEO) in 1968 and was conducted by Institute for Research on Poverty and Mathematica (Princeton, NJ). Four-year experiment.

Scope of Review: Most of the work done at offices of Mathematica. Objections were raised by OEO and contractors over GAO's access to data. GAO agreed to proceed without accessing these data and "to test certain of the data presented in the (OEO) report by means of sampling procedures..." Disagreement was partially over reinterviews of participants in the experiment.

Major GAO criticisms: (a) OEO preliminary report describes data inadequately to allow any independent interpretation; (b) OEO reaches premature conclusion on effects of incentive plans as an aggregate; (c) OEO reaches premature conclusions in comparing experimental and control groups; (d) attrition compromises results; (e) lack of comparability of control and experimental groups at start in terms of percentage employed (95% vs. 89% respectively); (f) OEO fails to take into account community-wide rates of wage increases.

3. EVALUATION OF THE OFFICE OF ECONOMIC OPPORTUNITY'S PERFORMANCE CONTRACTING EXPERIMENT (B-130515; 5/8/73)

"Performance contracting" defined as agreement between a local educational agency and a private educational firm (contractor), wherein payment to the contractor is related to some measure of student achievement. Performance contracting is not a program, but a method of organizing programs. The OEO experiment was conducted during 1970-71 and was designed to assess the overall impact of remedial reading and mathematics programs conducted by private educational firms in comparison to regular school programs (controls). OEO concluded that there was no difference between experimental and control groups.

Scope of Review: At OEO headquarters in Washington, D.C., and at the main offices of the test and analysis and management support contractors. Also visited 8 of the 18 school districts to observe operations of instructional programs. Interviewed officials at OEO, contractors, school districts, and educational firms. Employed two consultants.

Major GAO criticisms: (a) Design and implementation shortcomings invalidate OEO's conclusion; (b) lack of comparability of students in experimental and control groups; (c) lack of monitoring of control groups; (d) length of instructional periods was not coordinated; (e) continuous change in educational programs of the contractors during the experimental period; (f) problems in implementation of programs on short notice; (g) poor testing conditions for the administration of standardized tests.

4. EXPERIMENTAL HOUSING ALLOWANCE PROGRAM: ITS RESULTS AND QUESTIONABLE FUTURE (Draft Report: 11/29/76)

OBSERVATIONS ON HOUSING ALLOWANCES AND THE EXPERIMENTAL HOUSING ALLOWANCE PROGRAM (B-171630; 3/28/74)

Question addressed by Experimental Housing Allowance Program (EHAP) is whether it is feasible and desirable to provide low-income families with housing allowances to enable them to obtain adequate housing. Experiment is sponsored by the Department of Housing and Urban Development (HUD). Composed of three component experiments, each located at different sites: supply, demand, and administrative agency experiments. (1) Supply: designed to analyze how the housing market will respond to a housing demand created by a full-scale housing allowance (Rand Corporation); (2) Demand: to examine how households use housing allowances (Stanford Research Institute and Abt Associates); and (3) Adminstrative Agency: to address how a national program might best be administered (Abt Associates).

Scope of Review: Carried out at HUD in Washington, DC, HUD's regional offices in Chicago, Philadelphia, and Seattle; Rand Corporation in Santa Monica, California; Abt Associates in Cambridge, Massachusetts; the Housing Authority for Salem, Oregon; housing allowance offices in Green Bay, Wisconsin; South Bend, Indiana; Pittsburgh, Pennsylvania; Salem, Oregon; and Jacksonville, Florida. Consisted of direct observation of procedures and operations, as well as review of design documents. GAO was refused access to records identifying individual program participants at the time of the writing of the draft. Subsequently, access to participants on a voluntary basis was arranged. Data from these participants will be reported in a separate document.

Major GAO criticisms: (a) EHAP is being operated under very strict and controlled conditions which would probably not exist in a national program; (b) the experimental sites are not representative of large urban areas where housing problems are most prevalent; (c) too few households are participating in supply experiment to measure how a city's housing market responds to a full-scale allowance program; (d) demand experiment indicates that the allowance has no effect on improving participants' housing quality.

5. EFFECTIVENESS OF VOCATIONAL REHABILITATION IN HELPING THE HANDICAPPED (B-164031(3) April 3, 1973)

The Vocational Rehabilitation Act provides assistance to the States in rehabilitating handicapped persons to prepare them for gainful employment. The Rehabilitation Services Administration (RSA), an agency of HEW, is responsible for providing leadership to the States in planning, developing, and coordinating State programs. RSA designates target groups that the States should emphasize in providing rehabilitation services. However, the States are not required to include RSA's target groups in their programs. Review attempts to evaluate program's effectiveness.

Scope of Review: Conducted at HEW Headquarters and regional offices in Atlanta, Georgia, Dallas, Texas, and in Chicago, Illinois; the State offices of the Division of Vocational Rehabilitation in North Carolina, Michigan, and Oklahoma. Toured training centers, workshops, hospitals, mental institutions, prisons, and schools and observed vocational rehabilitation services being provided. Discussed program activities with personnel at these facilities and at some regional and local agency offices. Randomly selected for review 820 of the 31,650 cases 3 States reported as closed in 1970. Additional information was obtained from questionnaires sent to program participants whose cases had been closed 1 to 2 years prior to GAO review (N=403 completed questionnaires). Response rates to questionnaires from general agency clients were 73% and 59% for successful and unsuccessful cases, respectively.

Major GAO criticisms: (a) HEW should consider extent to which needs might be met through other Federal programs when planning growth of vocational rehabilitation program; (b) States should institute better follow-up programs and measures; (c) HEW should improve data keeping records.

6. EXPERIMENTAL SCHOOLS PROGRAM: OPPORTUNITIES TO IMPROVE THE MANAGEMENT OF AN EDUCATIONAL RESEARCH PROGRAM (MWD-76-64 4/27/76)

The Experimental Schools Program (ESP) was designed to test the hypothesis that comprehensive changes to existing educational systems will result in improvements in the way students are educated. The program was also designed to increase and to improve basic knowledge about the process of education and to implement the results of research, demonstration and experimentation in actual school settings. The program was administered by the National Institute of Education (NIE)since 1972. ESP has funded 18 projects and each project is planned to operate 5 years.

Scope of GAO's review: Review was conducted at HEW and NIE headquarters in Washington,D.C. GAO visited projects and their evaluators in California, Minnesota, South Carolina, Texas and Washington. They interviewed HEW and NIE officials and reviewed policies, regulations,procedures and practices for administering ESP.

Major GAO criticisms: (a) lack of prepared plans to effectively carry out and evaluate comprehensive educational changes, plans were written in conceptual rather than operational terms; (b) evaluations did not produce adequate information on projects' impact on students, teachers, administrators and communities; (c)projects failure to collect "baseline" data on student achievement and attitudinal levels before the comprehensive changes were made, thus they were unable to determine the impact of the program over the 5 years; (d) lack of any cost analysis for many of the projects at the time of GAO's visit;(f) lack of specific measurable objectives for evaluating effectiveness of program (f) lack of data necessary to determine compliance with special program financial regulations.

HEW agreed generally with GAO's assessments.

7. PROJECT HEADSTART: ACHIEVEMENTS AND PROBLEMS (MWD-75-51 5/20/75)

Project Head Start was an experimental demonstration program providing health, educational, nutritional,social and other services primarily to economically disadvantaged preschool children, their families and their communities. Head Start was also required to provide for direct parental participation in the programs development, conduct and overall direction. The Office of Child Development (OCD) and HEW administered Head Start through grants to local non-profit organizations,ex. community action agencies, school districts.

Scope of GAO's review: GAO reviewed administration of Head Start by OCD and 8 grantees and related administrative activities of 3 HEW regional offices. GAO reviewed program activities such as parent participation, eligibility, recruitment, average daily attendance and services to the handicapped at 4 of the 8 grantees.

Major GAO criticisms: (a) lack of parental involvement in the program; (b) lack of professional staff, training facilities and equipment needed for Head Start to adequately serve severely handicapped children; (c) grantees failure to obtain documentation demonstrating eligibility from familites applying to Head Start to insure that no more than 10 percent non-poor families are served;

(d) failure to emphazise early and continually recruit children to insure full enrollment; (e) failure of HEW regional offices to effectively follow up on problems identified in monitoring reports to insure that grantees were taking corrective action.

8. THE WELL BEING OF OLDER PEOPLE IN CLEVELAND,OHIO (HRD-77-76 4/19/77)

GAO assessed and measured the overall well-being of a sample of older people in terms of their social and economic status, mental and physical health, and ability to do daily tasks and gathered information on the services and other factors that could affect the well being of individuals in the sample. This report looked across agency lines at how 23 Federal programs affected the target population. Currently no evaluation has been undertaken to measure the combined impact of Federal programs on the people they are trying to help. GAO attempted to demonstrate that multiprogram evaluations performed by a single agency looking across agency lines at different departments are necessary.

Purpose and scope of GAO's review: (a) to discuss the well-being of older people; (b) describe the help they receive from others and; (c) explore issues relating to the many programs designed to help them. The purpose is to demonstrate what can be learned by assessing the wellbeing of a target population and looking across agency lines at how these people are affected by the programs which are designed to aid them.

A scientific random sample of 1600 people of age 65 or older in Cleveland, Ohio were interviewed during June to November 1975. Many Federal, state and local agencies serving older people in Cleveland also cooperated in the study.

This report contains no conclusions or recommendations. Recommendations will be forthcoming in a second report when data on the impact of the program is available.

9. ASSESSMENT OF READING ACTIVITIES FUNDED UNDER THE FEDERAL PROGRAM OF AID FOR EDUCATIONALLY DEPRIVED CHILDREN (MWD-76-54 12/12/75)

Title I of the Elementary and Secondary Education Act of 1965 authorizes Federal financial assistance for programs designed to meet the special educational needs of educationally deprived children. Emphasis was placed on developing reading skills. Federal funds are provided to state educational agencies (SEAs) which make grants to local educational agencies (LEAs). The Office of Education (OE) administers the program at the national level. Most of the funds have been used to provide instructional services for deprived children.

Scope of GAO's review: The review was made at OE headquarters, Washington,D.C.; 7 HEW regional offices; 15 LEAs in 14 states and the SEAs in these states. Examination of a wide variety of documents relating to the program. Interviews with parents and teachers involved in program; visits to classrooms to observe reading activities.

Major GAO criticisms: (a) available achievement data showed that most students were not reading at levels sufficient for them to close the gap between their reading level and the national norm; (b) most students were not retaining gains after they left the program; (c) lack of adequate information from state and local agencies for measuring the national impact of reading programs on improving student's achievement; (d) lack of adequate training to teachers and aides.

10. OBSERVATIONS ON EVALUATION OF THE SPECIAL SUPPLEMENTAL FOOD PROGRAM FOOD AND NUTRITION SERVICES (RED-75-310 12/18/74)

The Child Nutrition Act of 1966 authorizes the Special Supplemental Food Program to provide cash grants to states to provide supplemental foods through health clinics to women, infants and children suffering from inadequate nutrition and income. The Food and Nutrition Service (FNS), Department of Agriculture, administers the program. State and local agencies which operate the program are to maintain adequate medical records on program participants to enable the Department of Agriculture to evaluate the benefits of the nutritional assistance provided. Evaluations of the program are to determine (a) medical benefits of the nutritional assistance provided,and (b) cost efficiency of various methods of distributing food. The School of Public Health of the University of North Carolina at Chapil Hill and the Department of Commerce's National Bureau of Standards have agreed to evaluate the "medical benefits" and cost efficiency respectively.

Scope of GAO's Review: Review at Department of Agriculture, at University of North Carolina, and at 8 projects included in the medical evaluation. Review of legislation and agency records concerning planning and implementation of the program's operation. GAO received assistance from consultants in the fields of nutrition, biochemistry, pediatrics, obstetrics and biostatistics.

Major GAO criticisms: (a) lack of precise definition of good health and adequate nutritional status; (b) lack of precise determination of the types or quantities of nutrients necessary to maintain or improve a given nutritional level; (c) lack of control groups which precludes reliable judgments on how much the foods or other services contributed to the findings; (d) lack of adequate indicators of mental development which precludes any reliable conclusions on the program's effects on infants' mental development;(e) an a result of weaknesses in training, pretest procedures and procedures for controlling data quality, the reliability of the data remains questionable. FNS and the university are unable to insure data reliability.

At present the GAO report does not contain recommendations.

11. EDUCATIONAL LABORATORY AND RESEARCH AND DEVELOPMENT CENTER PROGRAMS NEED TO BE STRENGTHENED (B-164031(1) 11/13/73)

The Office of Education (OE) funds a variety of activities designed to seek solutions to educational problems. Included in the support is the Educational Laboratory and Research and Development Center Programs. These are independent non-profit institutions designed to make results of experimentation in education available to schools. Research results were to be developed into educational products which could be used in classrooms. Since 1972, the National Institute of Education within HEW became responsible for the expermintal program. OE delegated to the laboratories and centers the responsibility for evaluating the products, but did not designate evaluation guidelines.

Scope of GAO's review: A review of pertinent legislation and documents relating to the laboratory and center programs. Review of OE and NIE headquarters, Washington, D.C. at 5 educational laboratories and 3 research and development centers. Interview of personnel involved in the laboratory and center probrams.

Major GAO criticisms: (a) failure to provide guidelines for contractors to follow in their evaluations; evaluation procedures varied significantly among the contractors. Contractors failed to 1. state product objectives in measurable terms; 2. control for factors influencing the validity of the experiments; 3. design evaluations to determine product impact on student learning. (b) failure to provide follow up evaluations to determine the long term impact on student learning; (c) failure to demonstrate the product's marketability.

12. FACTORS THAT IMPEDE PROGRESS IN IMPLEMENTING THE HEALTH MAINTENANCE ORGANIZATION ACT OF 1973 (HRD-76-128 9/3/76)

Health Maintenance Organization Act of 1973 was designed to provide a trial Federal program to develop alternatives to traditional forms of health care delivery and financing by assisting and encouraging the establishment of HMOs. HMOs provide specific health services for its members. The act authorized Federal financial assistance for a period of 5 years for the HMO program.

Scope of GAO's review: A review was conducted at HEW, Washington, D.C., Health Services Administration headquarters (HSA) and at all 10 HEW regional offices. Questionnaires were sent to 809 entities which had been sent grant application packages between January and May, 1974 to determine why potential HMOs had not requested financial assistance and what problems were encountered by successful HMO applicants in complying with the act requirements.

Major GAO criticisms: (a) lack of staff, especially in the regions, with expertise in marketing, actuarial analysis and financial management; (b) failure to issue all final regulations and guidelines required by the act to more effectively administer the nation-wide HMO program; (c) restriction of the development of HMOs as a result of restrictive state laws.

13. THE PILOT CITIES PROGRAM: PHASEOUT NEEDED DUE TO LIMITED NATIONAL BENEFITS (GGD-75-16 2/3/75)

An objective of the Law Enforcement Assistance Administration (LEAA) of the Department of Justice is to develop new ways to improve the Nation's criminal justice systems through direct financing. The Pilot Cities Program, which began in 1970, was one of LEAA's first major directly funded programs. LEAA selected 8 locations to research, demonstrate and integrate new and improved projects into their criminal justice systems to prevent or reduce crime. The program was to demonstrate that improved research on local criminal justice problems could result in better programs which would reduce crime. The 8 locations (cities) were to demonstrate to the nation how to develop better planning processes. Each city was to have a 5 year term to implement the experimental program. Grants were awarded to nonprofit organizations or universities for a team to do research and plan projects.

Purpose and Scope of GAO's review: GAO wanted to determine whether LEAA adequately planned and managed the program to demonstrate that improved research could result in better programs to reduce crime. Careful review of operations of pilot cities teams in Albuquerque, Dayton, Norfolk, Omaha and Santa Clara. Brief review of operations in Charlotte, Des Moines and Rochester. LEAA headquarters and appropriate regional offices were also reviewed.

Major GAO criticisms: While individually the 8 teams benefitted from the program from a national standpoint, the overall program did not accomplish its goals. Various problems: (a) lack of consistent objectives; (b) different interpretations of the program; (c) participating organizations experienced instability; (d) guidelines were too general; (e) regional offices of LEAA used different management methods.

14. THE LEGAL SERVICES PROGRAM: ACCOMPLISHMENTS OF AND PROBLEMS FACED BY ITS GRANTEES (B-130515 3/21/73)

The Legal Services Program (LSP) seeks to provide representation which will benefit the poor and help alleviate their problems through legal processes. The program which began as a small experiment within the Office of Economic Opportunity (OEO) funded 265 grantees which operated 934 offices in 50 states during 1971.

Scope of GAO's review: GAO's review covered 12 months and ended during calendar year 1971. GAO reviewed seven standard program grantees which employed attorneys to provide legal assistance. GAO also reviewed the Wisconsin Judicare Project, under which legal services were provided by private attorneys and paid for by the project. Nineteen randomly selected annual evaluation reports for the 256 standard program grantees operating in 1971 were reviewed in order to ascertain the grantees achievement of program objectives. Applicable legislation and records were analyzed and officials of the grantees, local bar associations and local community action agencies were interviewed. In addition, 138 clients were interviewed to obtain their views on the services received and 18 judges to obtain their views on the competence of grantee attorney's representation of clients.

Major GAO criticisms: (a) Grantees need clearer and more detailed plans to achieve program goals; (b) OEO needs to develop a more reliable system to gather data on grantee's accomplishments; (c) only limited achievements by most grantees in areas of economic development and law reform areas; (d) Grantees provided the poor with the same scope of representation that was available to thos able to afford attorneys; (e) grantees should define objectives in operational terms. GAO had difficulty interpreting and analyzing grantee's results.

15. DIFFICULTIES OF THE NEIGHBORHOOD YOUTH CORPS IN-SCHOOL PROGRAM AND ITS MANAGEMENT PROBLEMS
B-130515 Dept. of Labor (2-20-73)

Neighborhood Youth Corps (NYC) Program provides training and work experience and other services to youths from low income families. It tries to encourage youths to stay in school and provide them with training for productive jobs. This report is concerned with the In-school part of the program: which provides work experience and support services to youths to encourage their continued enrollment in school. Sponsors of the program are both public or private nonprofit agencies. GAO reviewed the 1970-71 in-school program to see if it had improved the drop-out tendencies. GAO's review suggests that program has not changed drop-out levels.

Scope of review: Review of 1970-71 NYC in-school program in Washington, D.C., Virginia, and Texas. Interviewed officials of the Manpower Administration. Visited NYC program work stations and interviewed enrollers and their supervisors. Pandomly selected 279 enrollers from 21,116 enrollees at the three locations. Interviewed enrollees in sample, their supervisors and school guidance counselors.

16. EMPLOYMENT OPPORTUNITIES IN THE FEDERAL GOVERNMENT FOR THE PHYSICALLY HANDICAPPED
B-164031(3) Civil Service Commission (9-16-74)

GAO wanted to know how the Federal Government was providing employment opportunities and serving as an exemplary employer of the handicapped.

Review conducted in Washington, D.C., and San Francisco areas and included two questionnaires. One questionnaire was sent to Federal Agency coordinators to determine agency effects to employ handicapped. The other questionnaire went to handicapped federal employees to assess their view of the Government's program.

17. SUPPLY AND DEMAND CONDITIONS FOR TEACHERS AND IMPLICATIONS FOR FEDERAL PROGRAMS
B-164031(1) Office of Education HEW (3-6-74)

GAO reviewed supply and demand conditions for elementary and secondary school teachers and federal programs affecting the supply of such teachers because reports identified teacher surpluses and project that the teacher job markets will continue to worsen.

Scope: GAO attempts to identify all federal programs which provide assistance to develop school teachers. This was accomplished through a questionnaire survey of federal agencies. GAO sent questionnaires to various colleges and school districts. Follow-up interviews with selected questionnaire respondents were conducted.

18. RESTRUCTURED NEIGHBORHOOD YOUTH CORPS OUT-OF-SCHOOL PROGRAM IN URBAN AREAS
B-130515 Dept. of Labor (4-2-74)

The restructured out-of-school program objectives were to place enrollees in suitable jobs, advanced training, or further education after they leave the program. To test whether the Department of Labor's restructuring of the Youth Corp out-of-school program had actually improved the program, GAO reviewed five NYC-2 projects.

Scope of review: GAO interviewed officials from the projects, sponsors, school systems and regional manpower Administrator Offices.

19. PROBLEMS OF THE UPWARD BOUND PROGRAM IN PREPARING DISADVANTAGED STUDENTS FOR A POSTSECONDARY EDUCATION
3-164031(1) HEW (3-7-74)

Attempts to test the effectiveness of the Upward Bound Program, which is administered by HEW's Office of Education Program designed to provide low income students with skills and motivation necessary to succeed in education beyond high school. Program corrects student's faculty academic preparation by providing remedial instruction, altered curriculums, tutoring, cultural exposure encouragement.

Scope of review: Interviewed OE officials in Atlanta, Boston, and San Francisco. Also at project offices (15 projects were reviewed) GAO interviewed tutors, counselors, teachers, and administrators.

20. LEARNING DISABILITIES: THE LINK TO DELINQUENCY SHOULD BE DETERMINED, BUT SCHOOLS SHOULD DO MORE NOW
GGD-76-79 Department of Justice, HEW (3-4-77)

GAO investigated underachievement among juvenile delinquents in institutions and found that 1/4 of those tested in Connecticut and Virginia institutions had primary learning problems or learning disabilities. Correctional institutions in states visited by GAO were not effectively identifying and treating the learning problems of delinquents.

Scope of review: GAO made the review in five states. Interviewed 373 classroom teachers and over 300 other school officials. Consultants were hired by GAO to test juveniles chosen randomly from institutions in Connecticut and Virginia.

21. TRAINING EDUCATORS FOR THE HANDICAPPED: A NEED TO REDIRECT FEDERAL PROGRAMS
HRS-76-77 HEW (9-28-76)

HEW needs to improve its programs which assist in preparing teachers for the handicapped. Most classroom teachers do

not generally receive training in the skills needed to teach the handicapped.

Scope of review: Questionnaires sent to a sample of 757 public school districts throughout the nation and a sample of 155 universities having special education teacher programs. Interview of HEW officials.

22. INDIAN EDUCATION IN THE PUBLIC SCHOOL SYSTEM NEEDS MORE DIRECTION FROM THE CONGRESS
HRD-76-172 HEW (93-14-77)

GAO examined the two major programs under the Indian Education Act of 1972. Projects were designed to meet the special needs of Indian children in elementary and secondary schools. GAO reviewed 16 projects in operation during 1974-75 school year.

Scope of review: Interviewed officials or members of parent committees responsible for the projects.

23. THE EMPLOYMENT SERVICE - PROBLEMS AND OPPORTUNITIES FOR IMPROVEMENT
HRD-76-169 Department of Labor

The Employment service has provided a labor exchange for persons seeking work and for employers with jobs to fill. The agency serves those jobs and persons characterized by low pay. GAO elevated the role of the service to determine the role of the service in today's job market.

Scope of review: Interview of labor officials and local ES officials personnel. Questionnaires were sent to random sample of 800 employers, 600 applicants who were still seeking employment through ES asking them to comment on their relationship with ES their evaluation of how effective ES was in meeting their needs. GAO received replies from 570 employers, 762 applicants.

24. MORE CAN BE LEARNED AND DONE ABOUT THE WELL-BEING OF CHILDREN
MWD-76-23 Social and Rehabilitation Service, HEW (4-9-76)

GAO reviewed the Federal Child Welfare Services program and recommended that HEW develop a system for evaluating the welfare of children and focus research programs on the greatest obstacles to improvements.

Scope of review: Questionnaire developed to extract information from case files. To follow trends in program accomplishments over time, sample cases were selected from each of the 10 locations. Case records were discussed with local child welfare agency officials.

25. CONCERTED EFFORT NEEDED TO IMPROVE INDIAN EDUCATION
CED-77-24 Bureau of Indian Affairs, Department of Interior
(1-17-77)

GAO evaluated the educational programs in schools operated by the Bureau of Indian Affairs. According to GAO, there was little evidence that the Bureau had made progress since 1972 in improving educational achievement of Indian children.

Scope of review: Seven schools were visited. Bureau of Indian Affairs officials and school administrators were visited and interviewed.

26. NEED TO MORE CONSISTENTLY REIMBURSE HEALTH FACILITIES UNDER MEDICARE AND MEDICAID
B-164031 (4) HEW (8-16-74)

GAO reviewed reimbursements to proprietary hospitals and skilled nursing facilities because of the many controls which have been built into the reimbursement process.

Scope: interviewed officials of the Social Security Administration, social and rehabilitation services, intermediaries, state Medicaid agencies, local agents, hospitals and SNFS.