United States General Accounting Office

GAO

Program Evaluation and
Methodology Division

February 1986

Designing Evaluations: A Workbook

INTRODUCTION

This document, which presents designs for three specific evaluations, is intended to be used as a companion-piece to the transfer paper, Designing Evaluations. That is, these three designs are real-world examples of the different blueprints for evaluation explored in more abstract form by the transfer paper.

Designs are routinely produced for most jobs undertaken in GAO's Program Evaluation and Methodology Division. These designs are not in any sense cookbooks for the conduct of a job; rather they are iterative plans for doing a job which often change in accordance with the accumulating job experience. The designs we complete in PEMD have four purposes: to serve as guidance to managers and staff in the execution phase of the job; to minimize time that would otherwise be lost on a job when staffing changes must be made later on; to give job sponsors (or users) a clear understanding of what information will be produced by the job (as well as to document that understanding); and to help me judge what the achievements of the finished evaluation are likely to be so as to decide on the merits of proceeding with the work.

Given these four purposes, each PEMD design normally includes:

- the question(s) the job will seek to answer and a discussion of why the questions are important;
- a review of the knowledge already accumulated on the subject of the particular evaluation, including an analysis of studies or evaluations already performed and the lessons to be learned from their successes and failures;
- the design developed to answer the question(s) based on the type of question involved (i.e., whether the question is descriptive, normative, or cause-and-effect) and on other issues (generalizability, for example) that are important;
- a discussion of strengths and weaknesses of the design in terms of the conclusiveness of the information to be produced (this usually involves some discussion of why a particular design was chosen, versus possible alternatives);
- a review of likely data sources and expected data problems;
- an analysis plan;
- a statement about the intended use of the information produced; and
- the resources required for the job's execution (management plan).

The three designs presented here are typical examples of the design work generally done in PEMD* but they have been chosen for their diversity of subject matter and for the different types of evaluation questions they seek to answer. The designs are:

(1) Error Correction in Pell Grants: An Evaluation of the Effects of the Department of Education's Validation Method (Authors: Fritz Mulhauser and Catherine Baltzell).

(2) An Evaluation of DOD's Implementation of Technical Risk Assessment in the Area of Weapons System Acquisition (Authors: Luis Gonzalez, Marcia Gilbert, and Joan McLaughlin).

(3) The Women's Retirement Project: An Evaluation of the Factors that Influence the Timing of Retirement for Women (Authors: Joanne Frankel and Scott Crosse).

These designs thus represent study efforts in the areas of education, defense, and retirement and they focus on cause-and-effect (Pell Grants), normative (DOD's Technical Risk Assessment), and descriptive (Women's Retirement) questions. That is, while each design includes more than one question or type of question, the central focus of the Pell Grant study is on the effects the Department of Education's validation method has had on the reduction of error in grant award, on students' pursuit of higher education and on the administrative burdens of institutions; the central focus of DOD's Technical Risk Assessment is on whether that assessment is being performed according to DOD's own specifications; and the central focus of the Women's Retirement Project is on identifying and describing the factors that affect the timing of retirement for women.

Since there is no standardized approach for designing an evaluation, one design may differ markedly from another, as a result of differences in the job question or who poses them, in the maturity of the issue addressed, in the time or cost constraints present, in user emphasis, or other considerations. These three designs are no exceptions; they vary along a number of dimensions. The Pell Grant questions, for example, were mostly posed by the study's congressional sponsor (the House Subcommittee on Post-secondary Education, Committee on Education and Labor), whereas in the defense and retirement evaluations, the questions were determined by the researchers. Each design paper has it own format, based on the specific focus of the job. Individual emphases are

---

*To cut down on volume, however, some discussions have had to be shortened here, and all administrative sections (management plans, schedules, resources, etc.) have been deleted.

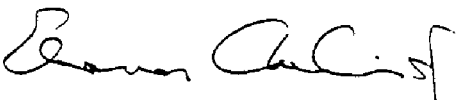
different: for example, the review of the literature in the women's retirement area had to be much more complete, extensive, and meticulous than for the other jobs because of the immaturity of the topic area; we had to determine, first, whether enough methodological groundwork had been laid to enable us to proceed with our study, and second, whether in fact the study we wanted to do was not already begun or ongoing. Neither of these state-of-the-art issues was a problem in the other two evaluations, where the literature review could be more generally limited to legislative histories and studies of the individual issues and programs.

The designs also have many similarities as well. All three show tight linkages between the question asked and the technical approach chosen; all three carefully examine the areas of power and of limitation in their selected approaches; and all three strive to find practical approaches for deriving the information sought.

The divisional decision on whether to proceed is always driven, however, by the usual considerations involving the cost of producing the information versus its importance, and the likely conclusiveness of the information produced versus the sponsor's need. The Pell Grant and DOD's Technical Risk Assessment designs had favorable cost/benefit ratios on both of these dimensions and were therefore implemented. The Women's Retirement project, on the other hand, turned out to be overcostly and complex relative to its likely usefulness, and therefore had to be discontinued.

The fact that an evaluation design led to the discontinuation of the job, however, should not be construed as a mark against it. On the contrary, the persuasiveness of the design paper's recommendation not to proceed is a reflection of the logic of the work performed. Indeed, I believe all three of these designs present successful examples of the kind of review and analysis that usefully precede the implementation of an evaluation. As such, they should be helpful as specific illustrations to the more general guidance offered in Designing Evaluations.

Comments or questions about the design papers presented in this volume should be addressed to their authors, listed above, or to me. We welcome the interest of all readers and look forward to suggestions that can help us improve subsequent documents in this series.



Eleanor Chelimsky
Director

Table of Contents

I. DESIGN PAPER FOR ERROR CORRECTION IN PELL GRANTS: AN EVALUATION OF THE EFFECTS OF THE DEPARTMENT OF EDUCATION'S VALIDATION METHOD

Frederick Mulhauser

Catherine Baltzell

December 1983

Table of Contents

Page

APPENDIX

I. INTRODUCTION

In 1983-84 about 1.3 million students at 6,000 universities, colleges, and other schools will be asked to provide documentary evidence of details of their family finances to show their need for federal student aid. The Department of Education (ED) requires this "validation" or proof of need from about half of those who receive the Department's Pell Grants, and ED has proposed to expand validation to other major federal aid programs in 1984-85.

At the request of the Chairman of the House Subcommittee on Postsecondary Education, the Program Evaluation and Methodology Division of GAO (PEMD) will conduct an exploratory evaluation of several aspects of this ED policy, to aid the subcommittee members and staff in their review of federal student aid topics leading towards reauthorization of the Higher Education Act of 1965.

This design paper presents the broader context within which this study is located, and then gives details of the PEMD plans for work on the five study topics outlined in the letter of request reproduced in Appendix A from Subcommittee Chairman, Hon. Paul Simon (D-IL). To give the subcommittee early indication of the direction of the findings, PEMD staff will give a briefing in late April 1984 based on partial data, with a written report to follow later.

II. GENERAL APPROACH TO THE STUDY

A. Evaluation of strategies against error and fraud in entitlements

Growth in the number and complexity of federal entitlement programs has been accompanied by problems at each of the steps of determining eligibility, computing benefits, and disbursing money. In addition to the problem of mistakes by applicants or administrative agencies, there is a growing concern about deliberate error, or fraud and abuse.

Agencies involved in entitlement programs need strategies to cope with these problems through:

- prevention
- deterrence
- detection
- error-correction
- enforcement

And despite the importance of these agency actions, there is remarkably little general knowledge about what works. Several years ago, the PEMD Director wrote:[1]

> It is clear that the lack of knowledge about the effects and costs of typical strategies and techniques used in combating fraud and abuse are major barriers to the use and expansion of countermeasures....A great deal of evaluative work remains to be done....Without this information the development of sound entitlement programs which limit fraud and abuse to a minimum will be impossible.

The basic problem, as the quotation suggests, is to gauge and to balance the effort invested in strategies against error -- the costs of time, attention, and money for those involved at every level -- against the effects, and to insure that the effects are desirable and in proportion to the efforts.

On the matter of the appropriate balance to be struck, Herbert Kaufman notes in his essay on Red Tape[2] that although

---

[1] Eleanor Chelimsky, "Reducing fraud and abuse in entitlement programs: an evaluation perspective," GAO Review (Summer 1981), p. 32. A task force in GAO's Human Resource Division is planning a long-term work program on the general subject of verification of applicant eligibility in entitlements, to start in 1984.

[2] Herbert Kaufman, Red Tape: Its Origins, Uses, and Abuses (Washington, DC: The Brookings Institution, 1977), p. 53.

it is said that prevention of financial abuses sometimes costs more than the ailment, public mon brings with it a special point of view:

> Not only are public property and public discretion held to have a special moral status; they occupy a special political position because abusing them eats away at the foundations of representative government. So we are willing to put up with a lot to safeguard their integrity.

The economist Arthur Okun once remarked, in the same vein, The Office of Management and Budget should spend $20 to prevent the theft of $1 of public funds."[3] But no matter what the eventual judgment of the right balance between the effort and the effects of strategies against error and fraud, at the least an informed opinion must rest on good information about the extent of the problems in the first place, and on the specific costs and results of the methods employed as solutions or preventions. Exploratory data-gathering and analysis of just this sort in a particular program context is precisely the goal of this study.

B. The Pell Grant program context

The Pell Grant program is an ambitious attempt to assist a very wide range of individuals to pursue education beyond high school. About five million people each year apply for the grants (formerly called Basic Educational Opportunity Grants), and about half, or 2.5 million, are found eligible by a central computer analysis of detailed information about family income and assets, submitted by students on complex application forms. (Appendix D explains the Pell Grant program in more detail.) The actual award of funds is disbursed by the specific school an eligible student decides to attend, after a second stage of computation of school and living costs since the award cannot exceed one-half the cost of attendance (or $1800, whichever is greater). Over $2 billion per year is awarded.

Concern for error in the complex chain of events from application to eligibility-determination to calculation and disbursement of the awards, has led ED to several kinds of action. First, longstanding rules require participating schools to meet procedural standards (to verify that a student is properly enrolled, not already a degree-holder, making satisfactory progress, and other items of school status), and to use all information in a school's possession when considering an award -- that is, not to ignore discrepancies if some other document fortuitously at hand at the school casts doubt on some item of data in the student's Pell eligibility calculation. A

---

[3] Arthur Okun, Equality and Efficiency: The Big Tradeoff (Washington, DC: The Brookings Institution, 1975), p. 60.

second strategy involves checking the internal consistency of application items by the central computer processor, and requiring some students to re-examine their application and verify -- that is, alter or reassert -- data items which appeared discrepant. And third, some students have been required to provide documentary evidence of key data. This last group, required to "validate" their application data, was about 300,000 students in 1980-81.

In that year, however, a new study commissioned by the Department found evidence of very extensive errors by applicants and by institutions, despite the three strategies described. A combination of internal pressure from the Office of Management and Budget and public concern over the misallocation of funds (the study estimated that hundreds of millions of dollars were awarded in error though no data were gathered on the extent of fraud or deliberate abuse), clearly called for new policy and procedures.[4]

For the 1982-83 year, ED greatly expanded the validation process which previously had touched a small percent of applications, and now directed 1.66 million students (over half those eligible) to bring supporting documents to campus aid offices for review. The 1.66 million included the first 1.3 million eligible applicants, and 300,000 more after that whose applications showed oddities that seemed to merit further review. An extensive handbook gave campus officials detailed instructions on what to do with students' evidence. Similar procedures are in use again in the current year, 1983-84, with a target of 1.3 million students to be validated, chosen according to four separate methods by the central computer processor.

In a letter to the higher education community in August 1983, ED officials proposed to expand validation to include applicants to the other federal aid programs known as "campus-based" in which individual schools make many more of the eligibility and award rules, and to the Guaranteed Student Loan Program, in which aid decisions are made by lending institutions after a school certifies that a student is properly

---

[4] A GAO report five years earlier had recommended increased and strengthened action to verify applicant information ("Office of Education Basic Grant Program Can be Improved," HRD-77-91, September 21, 1977). A more recent GAO report comparing Pell Grant and other federal student aid programs recommended verification of application data in the others as well ("Inconsistencies in Awarding Financial Aid to Students Under Four Federal Programs," HRD-79-16, May 11, 1979).

enrolled. Since for those programs there is no equivalent of the Pell Grant central processor, each school would be required to select its own group of students to be validated, at random, but in a quantity to be specified by ED based on the error rate at each school. This error rate would be determined by individual schools' internal quality control studies, to be done according to an ED-prescribed schedule and format.

Clearly, financial integrity has taken on appropriate new importance in the Pell Grant program. However, strategies towards this goal must always be considered in an integrated way with the program aims, in this case assisting needy students to further their education beyond high school. The extensive new procedures required of students, institutions, ED contractors and ED staff, all raise the specific questions of balance noted earlier -- what effects the expanded validation effort is having, how the effects match the costs of the effort, and how the several goals being pursued can be conciliated and integrated.

C. Basic elements of the exploratory study

The validation requirements of the last two years are controversial. The Department characterized the results so far as "progress in eliminating abuse while maintaining equity in the distribution of funds" in a recent letter announcing plans to expand validation. The head of admissions and financial aid at the University of Chicago testified to the National Commission on Student Financial Assistance that validation is a "futile and time-delaying process," which could be seen as "a planned and callous approach to eliminate eligible students from the aid program." Consequently, the House Subcommittee on Post Secondary Education has asked GAO to provide information that would help in weighing such diverse views.

The primary goal of the validation effort is of course, to improve the accuracy of Pell Grant awards, so that scarce funds reach truly needy students. That objective is being assessed through a major ED study due this fall from a contractor, based on a detailed study of 4,500 student Pell recipients and their families.

The Subcommittee is interested in further data on topics that will not be covered in the ED work, such as other effects of the policy, and its costs. In negotiations with Congressional staff, we agreed that in the short time before hearings in 1984, PEMD could at best explore the subject and gather some initial data of limited generalizability. The study objectives are:

1. To provide information on the origins, goals, and costs of the Education Department's error-detection and error-correction activity.

2. To evaluate the Department's data on award error and the statistical methodologies for selecting students for validation to prevent error.

3. To gather preliminary data on institutional and student impacts of the validation process.

4. To search for promising alternative approaches to the detection and correction of error in the Pell Grant program.

Congressional staff have emphasized that the Department's chosen methods place main reliance on campus aid officials to receive students' documentation, review it and judge the need for revisions of eligibility. The staff encouraged PEMD to look especially for information concerning the impacts on institutions of higher education as they took on these new tasks.

We plan to meet these objectives with a variety of inquiries and data-gathering:

1. We will interview officials and review records in the Department and with its contractors to establish details of the current policies, their origins and aims, their costs, and the degree of consideration of burden in their design. We will obtain expert review of error data, analyses, and targeting methodologies.

2. We will conduct detailed on-site cost analyses of validation activity at about fourteen campuses, with nine of the schools forming a panel where earlier data together with ours allows a comparison of costs of aid administration before and after the introduction of enlarged validation requirements.

3. We will survey a national sample of financial aid administrators at institutions participating in the Pell Grant program to learn about their experiences with validation and its impacts, and to gather their views on related topics.

4. We will gather data on students' experience with validation, and any burdens, delays or other impacts, through a survey of students at several of the schools involved in the on-site cost studies. We will also attempt to gather information from students who applied to the schools but did not enroll, to see if validation played any part in changes of plans.

The exploratory nature of the esent study needs to be stressed. Though studies of short-term requirements associated with regulations or paperwork are well-known, tracing events or activities and attributing their cause to a set of required procedures is much more complex than simply gathering information on the hours of time spent on the procedure. In the case of validation it is precisely those later effects on students' education and on institutions' functions and activities that have been suggested in testimony and correspondence as important effects of validation. We will explore several different ways of estimating these effects, since there is no established method available.

Our study is exploratory in the further sense that we will not be able to project national estimates of impact or the incidence of various effects, because of the small samples we will use. In studying institutional effects, we will have a national survey combined with detailed case studies of about fourteen schools. In our search for student effects, we will have reports from campus officials, and data from several hundred individuals who went through validation. In addition, we will analyze data in ED files for an earlier year, to see if administrative data can show effects on those validated.

As McGrath has pointed out, research decision-making involves trying to maximize three desiderata: generalizability with respect to populations, precision in control and measurement of variables related to the behaviors of interest, and realism for the subject or respondent, of the context within which those behaviors are observed.[5] However, trying to maximize any one will weaken the other two, and aiming for doing well on two will inevitably harm the third. Our design for studying effects includes elements with varying patterns of strengths and weaknesses, with elements chosen to compensate for each other as follows:

---

[5] Joseph E. McGrath, et al., Judgment Calls in Research Decision Making (Beverly Hills, CA: Sage Publishers, 1982).

Strength of Design Elements on
Three Criteria

Element	Generalizability	Precision	Realism
Institutional effects			
a. Nine-school panel study of costs before and after new requirements	weak	strong	mixed
b. Additional case studies of costs in 1982-83	weak	mixed	mixed
c. National survey of campus aid officials	strong	mixed	mixed
Student effects			
a. Survey of Pell recipients at five schools	weak	mixed	mixed
b. Telephone interviews with applicants who did not enroll	weak	mixed	mixed
c. Campus officials reporting student effects (included in survey)	strong	mixed	weak
d. Secondary analysis of ED data files	strong	weak	weak

Our study design must also be seen in the context of the other evaluation and policy research work under way. We are not planning to evaluate the impact of validation on errors in Pell awards, because ED is just completing a major study which may help address that topic. A contractor is replicating the 1980-81 Quality Control Study (whose results led to the current expanded validation policy), with a report due in spring 1984. The replication involves gathering extremely detailed information on 4,500 students who received Pell grants in 1982-83 and about half of whom were validated. From school records, student and parent interviews, and official tax and bank records, the contractor will gather data to compute proper Pell awards, compare these with actual awards to find the extent of error, and trace its sources in students, institutions, and the processor. If there is less error than before, it may be due in part to detection of errors as students pass through the required validation, or to improved deterrence of others under the general warning of potential validation. If the new research is reported in enough detail while we are doing our own study of other aspects of validation, we will examine the contractor's data and methods to judge if they provide a sound basis for further policy, but we will not attempt our own study of errors in awards or dollar savings through validation. However, the ED study is limited in that it pays very little attention to the costs and burdens of the validation policy and procedure, where we will have substantial new information. The two studies are thus complementary.

In addition, the two studies mirror the general problem of choosing criteria for judging action in this area. By limiting its study to the single program goal and evaluation criterion of financial integrity, ED ignores other aims of the Pell program which may be affected by new policies of validation. This question of criterion choice will be apparent at several points in the plans to follow, and will be treated in different ways. We first plan to find out the ED goals and related criteria for success, and whether the activities seem to be designed at least in logical relationship to those criteria. Later, we will judge the ED data on errors and the ED validation-selection methodology on criteria outside the system, drawn from principles of research study design and statistical science. In looking for institutional costs and student effects, we implicitly select criteria of great diversity, including dollar costs and less tangible items related to the educational purposes of the funds. In the final segment of the work we will examine criteria used in other government contexts, and also take a larger view of the overall goals of the Pell Grant program, in order to present a matrix of various outcomes of interest, and diverse approaches to the error problem, in order to see how approaches may affect the various goals.

The rest of this paper presents in detail the plans for the PEMD study. For each of the five main study topics we discuss the specific questions posed by Congress, our understanding of what they mean and their rationale, together with the results of the planning period in which we examined alternative approaches to gathering information about the questions.

III. THE FIVE STUDY TOPICS

A. Current policy on validation

1. Description of the questions

Congress requested that PEMD look at three questions on this topic:

What is the goal of the current ED policy on validation of Pell Grant applications?

How did the Department decide on the current goals and methods, and with what consideration of burden?

What does it cost ED to carry out its current policy and methods of validation?

Answers to these questions provide the context for data to be gathered in other parts of the study; they are prerequisite to comparisons with alternatives.

2. Rationale

Validation of information on student applications is a means to the end of reducing error in awards. However, there are many kinds of errors -- some made by applicants, some by institutions; some made inadvertently, some made deliberately; large ones and small ones. Since error comes in many guises, it is not immediately clear what should be done about it. Resources can be spent in a wide variety of ways to address the different problems of error in awards, and analysis of policy must begin with understanding what the current approach is intended to do. Strategies can be directed to different aspects of the overall problem, such as those listed in the introduction (prevention, deterrence, detection, error-correction, enforcement). And strategies can be directed at errors (total errors or high-dollar errors) or at the subset of deliberate deception known as fraud.

Understanding the policy aims of ED will allow judgments about whether these are appropriate in the first place, and will further allow analysis of the match between the chosen validation methodology[6] and the stated aims.

---

[6] Methodology in a broad sense here includes the decision rules for choosing individuals for validation, the numbers of applicants in the aggregate thought necessary to validate, the data items required to be validated, the tolerances to be applied in deciding if an error requires formal correction, and the rules governing disbursement of funds while validation is pending.

As the study was initiated out of concern for the burden of the validation process, it is important to examine the sensitivity of ED officials to this criterion, especially the extent it was considered in decision-making. Examples could include pilot tests of alternate methods to determine burden, representation of affected groups in decision-making, or preparation of burden estimates for explicit weighing by decision-makers together with costs and projected benefits.[7] Various kinds of "impact statements" have been developed in recent years, as decisions seemed not to be adequately considering important effects; data on this question will help the Congressional committee judge if such mandated consideration could be needed.

We are also interested in the cost of the current methodology. The costs borne by schools and by students will be captured in later sections of the study. Here we understand Congress to be interested in the costs to the federal government, so that suggestions of alternative methods with varying mixtures of public and government costs can be examined against a base of knowledge of current government outlays for validation.

3. Approach to gathering information

In the planning period, we looked for available data that would help us answer these context or background questions concerning goals, decision processes, and costs. We came up empty-handed in each case.

We looked for common documents where federal agencies display policy in the making, or where action is justified. But our search revealed that validation procedures are not included in regulations, so no options have been advertised nor public comments formally considered. Nor do the paperwork or data-gathering requirements seem to have been justified and reviewed under the Federal Reports Act, the Regulatory Flexibility Act, the Executive Order requiring OMB reviews of

---

[7] Validation requirements in the school lunch program of the U.S. Department of Agriculture were recently challenged in Federal court on grounds that they were arbitrarily imposed, without pilot tests to gauge the cost-effectiveness of diverse approaches. The requirements were allowed to stand, as the judge found that Congress had not made completed pilot tests a specific prerequisite, and that USDA had considered the burden imposed in deciding the size of the validation sample each school district must use.

burdensome rules, OMB's other paperwork control processes, or the special regulatory impact reviews done in recent years by the Vice President's Task Force On Regulatory Reform.

Neither the policy in general, its mechanics, nor its cost details, seem to have been treated in appropriations hearings except for brief exchanges, and we have found no report language on the subject. The Department in 1981-82 made several requests to reprogram funds from student aid to support validation procedures -- requests which Congress rejected each time. Repeated inquiries to Department officials during our planning period have not produced copies of justification packages or other materials sent to Congress with these requests that might help answer our questions.

Department staff have told us that they know of no data maintained separately on the costs of validation.

We have concluded that individual interviews with ED officials will be needed, with specific follow-up requests for documents mentioned by informants. We have completed a chronology of events, beginning with the first Quality Control Study of 1980-81 awards which initiated the latest round of policy, continuing through the conversations with Congress in 1981-82, to the diverse approaches to validation put in place in 1982-83, 1983-84, and suggested for 1984-85.[8] (See Appendix D for a brief version of the chronology.) We will use this to interview people in ED who took part in the decisions at each stage. About 10 interviews will be needed, as several different offices have had responsibility for aspects of the issue. Interviews may be held with some or all of the following, as well as others named by them:

Undersecretary (responsible for all policy and procedural recommendations to the Secretary)

Deputy Undersecretary for Planning, Budget, and Evaluation (which prepared the Congressional requests and analyzed the initial error data as requiring strong response)

Deputy Undersecretary for Management (includes Office of the ED Comptroller where a Credit Management Project aims at correcting a wide range of problems in student aid administration including errors)

Assistant Secretary for Postsecondary Education (which oversees the Pell Grant program)

---

[8] Clearly policy in this area has an even longer history if GAO first urged validation in 1977. The 1979 GAO report notes some early HEW/USOE steps that year, but from 1978-79 through 1981-82 validation affected fewer than one in ten recipients.

Deputy Assistant Secretary for Student Financial Aid

Director, Management Services (includes the Validation Branch)

Director, Program Coordination (oversees all policy development concerning student aid including Pell Grants)

Director, Division of Policy and Program Coordination (responsible for all Pell Grant policy development in recent years)

In addition, we have been told that the Office of Management and Budget has played a strong role in ED policy (including mandating the initial Quality Control Study). Pre-award screening of applicants in all Federal programs is one of the initiatives in OMB's "Reform '88" program of government-wide management improvement; thus continued involvement with ED seems likely, and we would like to know more about the goals and purposes which OMB has set for ED efforts concerning error, fraud and abuse in Pell awards.

Interviews will follow an interview guide, to insure that comparable information is gathered from informants on their perception of the goal(s) of current policy, the evolution of goals (if any), and the process of deciding on the current goals and methodologies (with explicit reference to any consideration of burden). Tolerance levels, for allowable error in individual application items or award calculations, will be another focus of interest, as they have a direct effect on the amount of effort to be spent searching and correcting errors. Data recording will be in the form of verbatim notes taken during the interview, and a summary memo written immediately afterwards. Re-interviewing of some may be needed as discrepancies arise and new questions need to be recycled. Documentary references will be requested, to allow substantiation of recall wherever possible.

Concerning federal costs, we note two different areas for data-gathering. First, although the vast majority of validations are done by campus aid officials reviewing documents brought by student applicants, the Education Department itself performs about 13,000 validations a year (in both 1982-3 and 1983-4) for schools electing not to administer the Pell Grant program themselves, and for special cases. This set of costs can be gathered by review of records and interviews with staff at ED headquarters. Staff time allocations for the 13 people involved will be gathered from supervisors, and ED records will show other direct costs such as data-processing (equipment and time allocable to validations), added mail volume, special training for staff, additional forms and materials, etc.

In addition, there are ED costs to administer the validation effort, such as printing extensive instructions for the campus officials, training them, answering questions about the process and supervising the contractor's work on validation. We will search for data on these through interviews and ED records as well.

More difficult to gather, though much larger in potential dollar amount, are the added costs owing to validation incurred in the federal government's contract with System Development Corporation, Santa Monica, CA, for computer processing of all Pell Grant applications. As the Department cannot provide this cost data, we expect to gather it on-site in California. The Education Department pays the processing contractor for a variety of work either unique to validation, or which increases in volume as a result of validation. Student applicants must be flagged for validation as their application is processed, special messages must be printed on return forms giving the directions for changing added documentation to campus officials, and the resulting corrections will swell the volume of corrections submitted for other reasons. We have identified the following kinds of costs to examine in our on-site review:

- staff time developing selection criteria (including evaluation of past ones)
- programming to direct the actual selection, and message instructions
- additional computer time needed for initial validation routines
- additional printing needed for validation messages
- additional costs at every stage of processing (mail room, keying and data entry, computer time, printing, mailing) associated with corrections submitted to the processor after student and campus aid official review documents
- staff time to answer validation questions, by phone and mail (and related phone and mailing costs)
- validation-related file-maintenance, statistical summarizing, and report-generation as required by ED

The time-period for these estimates, both at ED and at SDC, would best be 1982-83, the first year of substantial validation activity (1.66 million students compared to .3 million in 1981-82). Validation-related corrections carry a special code in the SDC system, so that a count of the total volume for our purposes, or even by examining a sample, should be possible, though 1982-83 summary statistics are not available generally until mid-1984.

In consultation with the ED staff knowledgeable about the cost elements in the processor contract (which amounts to about $10 million per year) we will prepare a data guide for use in our on-site study -- to be supplemented if SDC staff tell us of other costs we've overlooked. By sharing the cost categories and our specific questions in advance, we can encourage SDC to have much of the data ready for our review on arrival. Interviews will be needed to gather staff time estimates.

B. Methodology issues in analyzing error and selecting students for validation

1. Description of the questions

The subcommitte has asked PEMD to draw on its special expertise in statistics and study design, to provide an independent analysis of several points that have been controversial. Thus the first question on this topic:

> Does ED have reliable data on award errors on which to base policy, and have those data been interpreted using appropriate methods?

Moving to the central matter of choosing individuals for validation, the subcommittee also wants a methodological review (joined to a broader analysis):

> Are the methods for selecting students for validation statistically sound? Are they suited to the policy goals?

The third question reflects an interest in whether continuing scientific inquiry is being done on the chosen methods and policies:

> Does the Department evaluate its methodology and use the findings in regular improvement of its approach?

2. Rationale

The first question in this topic area results directly from controversy in late 1981 over the initial Quality Control Study report on errors in Pell Grant awards. The contractor reconstructed every data item for 4,500 applications using the best possible sources of information to contrast with the actual data reported or calculated at every step. Using such intensive methods, as well as a narrow $2 tolerance level, the report concluded there were hundreds of millions of dollars awarded in error. Word of the magnitude of the problem no doubt sent shock waves throughout ED, the Executive Office of the President and OMB, the press and the financial aid community. Although a wide range of sources of error was identified in the report, and although an equally wide range of "corrective actions" was suggested in a third volume of the study, the report's data and specific findings concerning applicant or student error were

interpreted as support for expansion of validation from 300,000 students in 1981-82 to 1.66 million in 1982-83. With such a dramatic enlargement of policy and methodology, the supporting data and analysis became a focus of controversy. Thus Congress has asked PEMD for a technical review of the quality of this key information on error in Pell awards.

A policy focus on detecting and correcting applicant error may have been appropriate since such error was a major contributor to overall award error. But there remain sizable methodological questions about how to choose applicants of whom to require documentation. The Department has used several different methods:

- Random selection;[9]
- Selection according to ad hoc criteria, based on internal inconsistencies within the application data (for example, an applicant reports federal income tax paid in an amount that is too low when compared to the tax-table amount for the reported income of the family);
- Selection based on matching applications against an "error-prone model" (a statistical procedure to study error patterns in past applicants to discover combinations of application characteristics that together predict the presence of error);
- Selection based on discrepancies found in comparinga student's current and previous year's applications.

Each of these involves technical questions of design and application, and as they are the basis for requiring added burdens of many students and officials, we plan a review of their statistical soundness, and the evidence that each selection method has a valid and plausible connection with the policy goals of the Department.

The validation procedures have changed in each of the years since 1981 when discussion of the problem of error reached new heights. A systematic approach to policy concerning error would involve pilot tests or simulations of possible solutions, choice of a method, and evaluation of its effects leading to revision for a subsequent year. An iterative approach, based on continuous evaluation, seems especially warranted since the chosen policy, as noted repeatedly, involves extra work for many people. We plan to review the changes in validation over the

---

[9] This is the ED term, which strictly speaking is misused here. In fact, the process did not give each applicant an equal chance to be selected. In 1982-83 the "random" method was set to select every one of the first 1.3 million applicants then none. Later applicants had no chance to be selected by the random method.

years (changes in numbers targeted, changes in targeting or selection methods, changes in data items to be validated, and more), and to examine the context of these changes, to see if they result from analysis and evaluation, at least in terms of ED goals if not in any broader context or strategy.

3. Approach to gathering information

Unless further interviewing in ED discloses other significant, recent sources of data on Pell Grant errors, we assume that the Quality Control Study three-volume report on award errors in 1980-81 is the major source that has been relied on as policy has been developed in the last three years. To evaluate the ED data, we will focus on that published report. The same contractor is at work on the replication now, and the report is due in fall 1983. If the new study report becomes available while we are at work, and if resources permit we will also review it. (The earlier report was delayed many months because of controversies over the data and the analysis.)

To evaluate the data on award errors and its interpretation, we will carry out expert review of the report, Quality in the Basic Grant Delivery System: Volume 1, Findings (McLean, VA: Advanced Technology, Inc., April 1982) and Volume 3, Methodology. Review of the study will require expertise in several areas:

General research expertise
study design
sampling
instrument development
field procedures and coding
data analysis
statistics
interpretation of data

Specific expertise in quality control research
sources of accurate data
tolerance levels for error
attributing cause of errors
reporting cumulative errors

Criteria for expert review would be drawn in the first case from generally accepted standards of social research, and in the second from the state-of-the-art in that particular field in which this work takes its place, on quality control in administrative data. There is a substantial body of the latter kind of work, as other federal programs have done both internal and contract studies for more than a decade.

We will locate expertise in the needed areas to carry out the requested review of the data and analyses. Individuals may have all or most of the needed knowledge, or we may need to

break the review task into smaller pieces for a larger number of separate evaluations by specialists. We will use experts from PEMD, GAO generally, and outside GAO, as needed to cover the various aspects.

In our planning period we identified several critics of the earlier quality control study, and several papers about it. None of the published discussions are detailed and rigorous enough to allow us to meet the Congressional request simply by meta-analysis or synthesis of existing critiques.

The second methodological question is whether the methods of selecting students for validation are statistically sound, according to general norms of statistical science, and whether the methods seem suited to the expressed policy goals of the Department. In the planning period we have repeatedly sought clear and comprehensive descriptions of the selection methods used by the Department -- their rationale and exact characteristics. Such descriptions do not exist at ED headquarters, though they may at the processor's headquarters in California. This means that the first task in our approach is to develop our own description of the selection methods in operation over the last three years. Table 1 shows the general pattern of methods by year.

Table 1
Numbers of Pell Grant applicants selected for validation by method and year
(in thousands)

Method [a/]	Year 1981-82	1982-83	1983-84 [b/]
Random	80	1,300	50
Pre-existing criteria (PEC)	220	300	350
Error-prone model	--	--	300
Cross-year checks	--	--	500
	300	1,600	1,200

Notes to table

[a] ED terminology; See p. 25 for discussion
[b] Target figures or estimates for current year

Once PEMD statisticians have developed descriptions of the methods, their development and application in each year, they can analyze their adequacy. The main criterion will be validity: the extent to which a method seems likely to target applicants making errors. Targeting can be more or less efficient when examined in relation to a specific policy goal

about error: reduction of total numbers of errors, reduction of the aggregate total of dollars awarded in error, or reduction in high-dollar errors. We will not be actually checking on validity through data, but inferring the likely validity of the methods by examination of their development and internal characteristics.

We refer to the models "in actual use" several times here, since we have discovered another aspect of the selection procedures as implemented. The Department in 1982-83 in instructions to the processor set quotas and ceilings for each selection method, and also set the order in which the selection criteria will be applied to each student file that comes in. (We understand there are no ceilings for 1983-84.) The PEMD review needs to consider the statistical soundness of the complete selection process, including each method separately and all of them together as applied. For instance, because of the ceilings, an application submitted early in 1982-83 had more chance to be selected under several criteria than the same application would have later in the processing period when some criteria have "filled their quota" and are no longer in use. We would like to know why the selection method called error-prone modeling is only used in 1983-84 after review of the application on several other selection criteria; if the model was developed to predict error-prone applications, why it is placed so "late" in the sequence, and with so low a target figure?

The third question in this topic asks whether ED evaluates and continuously refines its methodology for selection. We had hoped to review ED evaluations, but ED staff have not identified any studies of the selection methods.[10] The current replication of the Quality Control Study is cited as a general test of the validation process, and we may be able to observe during the period of our study whether the new findings are used in decision-making. (In fact, some decisions about 1984-85 validation in the Pell Grant program must be in process now, although the study data and conclusions are not yet delivered to the Department.) And the major expansion of

---

[10] We have been told that the PEC have been evaluated by the current processor and the previous contractor, but no reports of such studies are available at ED. We will need to interview contractor staff to locate prior studies.

validation to include other aid programs has been proposed before there are even data on error rates in those programs.[11]

Our approach to the third question will be to interview ED and processor staff to locate formal evaluations or other data such as information system reports from the processor, and to ask whether those are regularly used in refining the validation methods to improve targeting. (We will not be asking about information use for routine program development and refinement, such as in data processing or in clarifying communication with students or schools, important as these are.) Our focus will be to look for any direct study and improvement of the selection process. The processor interviews can be done during the on-site study of costs under the first topic; ED interviews can be done at any time, and in some cases by adding questions to the interview guide for study of the decision process described under the first topic. Our study of this subquestion must be limited, to allow most resources to be concentrated on the next two major questions, on effects of the process of validation.

---

[11] Decision-making on the expansion of validation to the campus-based aid programs is an interesting case for our review. The ED letter of August 1983 announced a proposed plan for 1984-85 in which every school or college would conduct its own study of error rates in applications to the campus-based federal aid programs each year. Each school would be required to validate a specific number of applications, varying from a handful to hundreds, chosen at random. The actual number to be validated would be set by ED, based on the error rate found at each school. Thus the plan does away with central selection and requires no criteria for selection even at the institution. This approach maximizes local autonomy, but at some unknown cost in precision and efficiency of targeting of validation. Knowledge of the differential success of the current targeting methods could help in deciding if local, random choice was to be preferred. Our interviews with ED officials described under the first topic, will explore how the development of these new policies were linked to formal or informal evaluation of the results of the current Pell validation policy.

C. Effects on institutions

1. Description of the questions

The questions here are straightforward:

> What costs are incurred by the diverse types of institutions of higher education in doing the validations?
>
> Are there other effects of the validation process on institutions? Are these more serious at particular kinds of institutions?

2. Rationale

Institutions participating in the Pell Grant program as disbursers of funds must validate student application data -- for as many students, as many data items, and using such tolerances as ED requires. The impacts of these requirements on the institutions -- both dollars of cost incurred, as well as other effects -- are one of the major aspects of this PEMD study. Itcan be argued that a citizen seeking a government benefit based on need should be expected to submit to almost any degree of burden associated with verification of that need. But the assignment of the work of doing this verification is a much more open question: Who should do it? Who should pay the financial costs of the extra work involved? These are questions for legislative and executive branch judgment, with many possible answers. In the first years of the validation activity, the Education Department has answered clearly: the Pell Grant disbursing institutions will do the work and will shoulder the costs.[12]

As the scope of the task has become clearer, with millions of students now bringing various documents to campus offices for review each year, officials of schools, colleges, and universities have raised concerns about the size of the effort required, about its appropriateness to the educational mission of the institutions, and about the difficulties staff face in becoming expert in detailed review processes covering taxes, tax forms, other assistance programs of all kinds, and family economics. In addition to direct burdens, educators have noted that the validation cycle as a whole (including requirements of recomputing student eligibility by the central processor after

---

[12] GAO also has recommended in its study of campus-based aid in 1979 that "verification should be made by financial aid officers before awards are made" ("Inconsistencies in Awarding Financial Aid to Students Under Four Federal Programs," HRD-79-16, p. 36).

review on campus reveals any needed corrections) has resulted in delayed awards to students and related cash-flow problems at institutions.

Congress has heard testimony and received correspondence on these impacts, and so has the National Commission on Student Financial Assistance established by Congress in 1980. But without better data concerning impacts, oversight and legislative action are handicapped. Thus GAO has been asked to explore the question of the institutional costs and other impacts of validation, at a variety of institutions, with data gathered so as to be comparable, to aid in review of the ED policy and methodology of validation.

3. Approach to gathering information

Our activity in the planning period began with cataloguing institutional impacts that observers and participants had mentioned. To learn about instances of effects that we should consider as candidates for study, we searched in many places including:

- Interviews with a half-dozen financial officials at diverse schools
- Interviews with officials of the National Association of Student Financial Aid Administrators
- Interviews with the head of the National Student Aid Coalition which monitors validation
- Review of materials from the current Education Department Quality Control Study of validation in 1982-83, including the instruments and a report of debriefing of contractor staff who interviewed campus aid staffs at 300 schools
- Review of correspondence to Members of Congress and national associations with complaints about validation burdens
- Review of testimony to Congress and the National Commission on Student Financial Assistance on problems of validation
- Interviews with representatives of national associations of various types of schools and colleges.

Since the Quality Control Study did include a few questions about institutional burdens of validation, we hoped to get an

early view of the data to aid in our planning.[13] However, the contractor doing the study has been using all the data and the available computer facilities on a tight schedule to complete their main report, and the project director told us that it would be extremely inconvenient to provide any tabulations or even to allow access for PEMD staff to review and tabulate data from the original instruments from the 300 schools. We will try again to examine the survey materials if the contractor's report does not fully exploit the burden data items. Whenever we can see them, these data would be useful as comparisons to our own.

As we gathered ideas about costs and other impacts, it became clear that validation was being carried out differently at different schools. That is, within the general ED requirements, schools varied in their implementation according to their history and preferences. For example, some schools validate all aid applicants, not just the Pell applicants ED selects for validation. We decided that we would need to gather information about process features, in order to understand the meaning of variations in costs and other impacts.

Also during our planning process, ED sent the letter already mentioned to schools and colleges participating in federal student aid programs, outlining the proposed expansion of validation to the campus-based aid programs and the Guaranteed Student Loan program. Though similar in the technique of validation, the proposal suggested a new approach to selection of students for validation. The ED validation policy and procedure has changed each year in the last three years and more changes are being suggested. Thus we decided that it could be useful to the subcommittee to ask institutional officials not only about their experience with the existing versions of validation, but also about their suggestions for change, their opinion of certain specific alternatives, and their judgment of the desirability of the present system or the alternatives.

Thus we concluded that the institutional information to be gathered would include:

1. The validation process itself (history at a school, scope or extent)

2. The dollar costs of that process

3. Other effects of validation on the institution

4. Opinions of institutional officials on the present policy and procedure, suggestions for change, and alteratives.

---

[13] The interview guide for use at the 300 schools sampled asked generally about "problems," then "unusual delays for recipients" (numbers and an average length of delay), methods of accommodating students delayed, and presence or absence of staff burdens resulting from validation.

Since the ED contractor for the Quality Control Study had already drawn a national sample of institutions, we initially planned to use a representative subgroup of that sample, and to gather our data through telephone interviews asking about 1982-83 validation. Review of that approach by experts in cost and burden analysis suggested that detailed cost data would only be available through on-site study. Further, since we aimed to make an inference about the impact of the newly-imposed validation process, we needed a base of comparison rather than data on a single year.

Accordingly we have revised our approach to include two separate kinds of data-gathering methods, each suited to its purpose, and each embedded in a different design.

(a) Cost case studies

First, we plan a set of intensive case studies of institutional costs, with data gathered through on-site study. To allow the strongest inference of impact, we have located case studies of the costs of administration of student aid in the year 1981-82, before substantial validation began. These case studies of institutions were done by the public accounting firm of Touche Ross and Co. under contract with the National Commission on Student Financial Assistance. The study gathered data on all types of costs, for all aid administration functions, for all types of financial aid, federal and other. The nine schools were chosen to include at least one of each major type and mode of control. They are in different parts of the country. We will engage the same study team to revisit the nine institutions to gather data on 1982-83 aid administration costs, using the same instruments, definitions, cost-allocation methodology, and so forth. (Special care is taken in gathering personnel time-allocation data as these offices are labor-intensive.) These panel data will allow comparisons of costs before validation and after the new requirements were implemented at each school, and will also allow tracing shifts in amount of activity in six different aid administration functions, to note any displacement of effort caused by validation burdens. (We checked ED records and found that each of the nine schools studied did do validations in 1982-83, in numbers ranging from 180 to 6,000.)

Interpretation of the cost data will require caution, and solid inference may require certain other data items to be gathered. For example, increases in costs may be due to inflation in rates of staff pay or in prices of purchased services such as data processing; increased overall expenses may simply reflect a bulge of applicants or changes in administrative routines apart from anything to do with validation. These are familiar problems in cost studies, and will be explored with our contractors at Touche Ross and Co. before going into the field, to be sure that directions for gathering information to

check on these other factors will be included in the field data guide. The earlier study which serves as our baseline already included some relevant procedures, such as not including the full charge of capital expenses made in the year under study, but rather prorating them across many years.[14]

To be more confident of our understanding of the range of costs incurred by institutions in 1982-83, as validation increased to its current proportions, we plan to engage our contractor to do additional on-site cost studies in schools beyond the nine in the panel study described above. These studies would be done using the same methodology, but for the 1982-83 year alone, with no possibility of formal comparison to 1981-82. (Reconstructing staff time-use for a period that began sixteen months ago is too problematic.) Pell Grant administration costs did not vary much across the wide variety of the original nine schools (from $58 to $69 per student recipient), but our study will be more useful if we have looked at more than the original nine schools to estimate the range of costs. (See below for the criteria for choosing additional schools for case study.) The case study data will be based on a wide-ranging search for direct and indirect costs associated with validation procedures, as well as unusual costs we have heard about in scattered instances such as costs to the institution of borrowing funds to take the place of anticipated student payments from Pell grants that were delayed because of validation problems; or costs to the institution of making short-term loans to students whose Pell awards are delayed through validation.

There are strong arguments for contracting with Touche Ross and Co. to conduct the data-gathering and analysis in the cost case-study part of this study. Inferences about impacts of validation on costs rest heavily on having data on the year prior to validation -- data already collected by this firm at nine diverse schools. We doubt we could collect data using any method that would be as strong as these data. Second, this firm has perfected a methodology for capturing the costs of financial aid administration. This is not impossible for others to do, but there are clear costs of time for PEMD or another contractor to start again from the beginning with definitions of terms, piloting of data-gathering forms and approaches, revising the

---

[14] We sent the original Touche Ross and Co. study report and instruments to an experienced evaluator, a college aid official familiar with administrative cost studies, and to GAO's Accounting and Financial Management Division for review. All agreed the study could provide a sound baseline for our purposes. The sponsor of the original study, the National Commission on Student Financial Assistance, also invited testimony on the report, which was favorable.

tools, and so on -- not to mention developing the analysis plans and statistical routines. By contracting, we gain access immediately to both important data and specific, targeted expertise. The firm also has the confidence of the study sites, through sensitive handling of the earlier study, as confirmed by the recent comment from staff at one of the proprietary schools in the earlier study that they would be very willing to take part in a follow-up study similar to the earlier one.

We plan to add cases for on-site study beyond the nine which form the before-and-after panel for several reasons. As Table 2 shows, several kinds of schools were omitted from the original nine -- most prominently, four-year colleges, either public or private. These schools together enroll about 20 percent of the Pell Grant recipients, and may have substantial validation responsibilities without the administrative resources of the larger universities. Continuing interest in the condition of traditionally black institutions suggests including one in the group of colleges to be studied. Second, we would like to add one or two schools to several of the categories which presently include only one or two institutions. Both community colleges (public two-year schools) and proprietary schools are so diverse that we would prefer a larger set of examples than provided by the original nine. (This was the only critical point made by several reviewers of the study in testimony to its original sponsor, the National Commission on Student Financial Assistance.) Finally, as automation of aid processing affects costs greatly, we want examples of a range of practice in this area if not already provided in the nine-school set.

We cannot generalize to any type of school; the purpose of adding schools is not to approach a more robust conclusion of that sort. Rather, we want to be a bit more confident that we have bracketed the range of costs, by selecting schools of the greatest diversity. Such diversity would lend credibility to an estimate of minimum or maximum cost impacts, for instance.

If resources permit the addition of five case study sites, we propose to distribute them as shown in Table 2, adding three 4-year colleges (two private, one public and one of these traditionally black), and two two-year-or-less schools (one public and one proprietary).

Table 2

Institutions for validation cost case-studies cross-classified by institutional type and control.[a]/

TYPE	CONTROL Public	Private	Proprietary
5+ year (university)	3 (3)	2 (2)	- -
4- year	1	2 -	1 (1)
2- year or less	3 (2)	- -	2 (1)

Note to Table 2

[a]/ Numbers in each cell are the total number of institutions to be included in the case study group. (In parenthesis are the numbers of schools already included in the nine-school panel.) Thus, the lower left cell shows that, of three public community colleges to be studied, two are already in the Touche Ross panel; one other remains to be selected. Empty cells represent types of schools that taken all together enroll 3% of Pell Grant recipients.

(b) National survey of institutions

The second approach to understanding institutional impact will be to gather survey data on validation processes, costs and impacts, and opinions from student aid officers at a nationally representative sample of institutions. We will consider asking a few cost questions of this large group, if we are convinced that aid officials are able to answer them. We doubt that dollar cost estimates would be accurate, but there may be some way to gather limited cost data that could be compared to some of the case-study data to provide a cross-check. We will work on this possibility with the Touche Ross staff who know the institutional cost analysis aspect, and advisers from the community of aid officials. See Figure 1 for a display of the information to be collected in the survey.

Figure 1

Information to be collected by
institutional survey

1. Validation history and process

   Did the institution validate aid applications before the ED policy?
   How many and what kinds of staff are available for validation tasks?
   Does the institution have data processing support for any aspect of aid administration?
   Does the institution validate more Pell Grant applicants than those selected by ED? How chosen?
   Did the 1983-84 process go more or less easily than 82-83, and why?

2. Validation effects and institutional response

   Was the 1982-83 expanded validation effort an unusual burden or not? What factors contributed to that impact (understanding the rules, interpreting tax records, verifying information from other sources, advising students or parents)?
   If used training, from what source and how effective?
   If used ED assistance, how effective?

3. Possible effects of validation on students

   Were any students delayed in receiving Pell awards as result of validation?
   Did institution make any accommodation to students delayed because of validation (loans, deferred tuition payments, award other aid even if Pell delayed, award part of Pell Funds)?
   Did validation cause other burdens for students (problems of understanding rules, obtaining documents, getting answers to questions from ED, getting corrected eligibility figures from processor after validation)?
   Did validation process or delays discourage any students from enrolling at institution, or in higher education generally?
   Did validation process or delays cause any changes in students' educational plans?

4. Opinions about validation policy and procedure

Is it necessary to continue Pell Grant validation?

What would be appropriate administrative allowance to cover costs of validation?

Are there aspects of validation that need improvement (application form itself, instructions to students, Handbook of rules for campus officials, timeliness of rules, consistency of rules from year to year)?

Does institution support certain specific ideas for change in validation (100% validation on campus, local error rates used to determine numbers to be validated; validation by central processor; validation by computer matching; shorter application; allow schools to retain percentage of funds saved through catching errors in validation)?

Does institution support ED proposal to expand validation to campus-based and Guaranteed Student Loan programs?

We will select a survey sample of 400 from the universe of about 6,000 institutions participating in the Pell Grant program in 1982-83. (Details of the sample are described in a separate technical paper.) An ED data tape provides access to such a listing, together with name and address and telephone number of the campus aid official. Each institution is classified on the tape by type and control, and we will choose the 400 institutions from the cells of a type x control matrix, in proportion to the number of Pell grant recipients at the type of school denoted by each cell. (A sample size of 400 is required for accurate projection of estimates to the universe at the strongest level of statistical confidence. A total sample of 1800 would be needed for equally accurate projection to the universes of each of the twelve cells of Table 3; a survey of that scale is ruled out by the time and resource limitations in this exploratory study.)

We would have liked to know something about the incidence of the burden of validation, to use in designing the sample of schools for our survey, and we will use the ED contractor's data if they become available in time. But for now, we cannot find any data in ED that aggregates the students chosen for validation to show where they apply or where they attend school. Choosing a simple random sample of institutions involved in the program would yield a very biased group, as a third of any such sample would come from the more than 2,000 proprietary schools which enroll only about 12% of the recipients. We elected to weight the institutional sample according to the proportion of Pell recipients at the types of

schools, so that we would mostly be hearing from places where the bulk of students do attend. Table 3 shows the numbers of schools of each kind in our institutional sample.

Table 3

Sample design for the institutional survey, cross-classified by type and control

TYPE	CONTROL			
	Public	Private	Proprietary	Total
5+ year (university)	128	36	1	165
4- year	24	44	1	69
2- year	108	8	16	132
less than 2-year	1	1	32	34
Total	261	89	50	400

We plan to skip the usual first mail follow-up, and to go immediately to telephone calls after two weeks, only re-mailing as needed. We will contact non-respondents to check for any systematic patterns which may bias the data.

Detailed analysis plans are in Appendix C. Because of the judgment sample of institutions for our case studies of validation costs, we can only present the range of costs we find; we will have to forego answering the obvious follow-up questions concerning typical costs at different types of institutions. Our discussion of the impact of validation on specific elements of cost from one year to the next, such as staff allocations to tasks, will be based on the panel of only nine cases, so it will be even more carefully limited in its generalizability. Of course we will not project a total validation cost across all schools. The survey data, while cross-sectional, will permit analyses of reported processes, effects, and opinions by type and control of school, and according to the numbers of validations done, extent of automation, and other classifying variables. From the survey we could more confidently project overall estimates of particular variables studied, and with some limitations, the significance of differences between specific school types.

D. Effects on students

1. Description of the questions

The subcommittee questions here are again straightforward:

What are the effects of the process of validation on students selected?

Are there effects of the validation process that fall disproportionately upon particular groups of students?

The preceding set of questions asked about burden and impact on the institutions doing the validations; our attention now turns to the students involved.

2. Rationale

Validation represents an explicit policy choice, to focus on applicant error, to make students the main focus of attention. Thus it is obvious to say that there will be "effects on students" of validation -- for that is precisely the point: if the policy is achieving its goal, some students will lose eligibility or have awards reduced, after careful review of documented eligibility and need. But as we reviewed correspondence and testimony about validation during our planning period, we found observers of higher education raising questions about effects on students beyond the intended one of improved accuracy of application data and award calculation. Those effects are posited to be more than simple annoyance, inconvenience, or the frustration of newcomers to adult bureaucratic rituals. The subcommittee has been told of potentially serious effects, and is asking GAO to check them out.

The need for a look at student effects is underscored by the limited design of the ED replication of the Quality Control Study. Although that study is looking in great detail at 4,500 aid recipients in 1982-83, the first year of enlarged validation, the design does not call for any information on effects of validation on those students except in financial terms. (The study will have useful details about how students filled out the original application, what sources of information they used, what errors are common in each of the dozens of data elements, what corrections resulted from validation and their impact on awards, and a further analysis of the remaining extent of error, through comparison of corrected student applications with "perfect" applications assembled by the study team from independent sources of information.)

Burden is inevitable in an award system that must base decisions on information about individuals--information that someone has to provide. The point is not to estimate burdens and other impacts on student applicants with an eye towards eliminating them. Rather, what is needed is to understand the full range of effects of the ED policy -- effects on the accuracy of awards and on the people involved in the error-detection and -correction process. With that understanding can come the most sensitive choices of policy goals and methodologies.

The second question in this topic reflects an interest in a finer grain of analysis of effects: Who is most burdened? The ideal system wastes no effort, but efficiently finds those with errors and does not disturb those even in very similar

circumstances who have not made errors. At present, ED cannot describe the group of students being chosen for validation, except in the simplest terms, much less talk about the effects of the experience.[15] Thus we have designed our work to provide an initial analysis of those being selected, and how the effects of the process are distributed among subgroups -- as a first step towards asking if the system can distinguish error-proneness from circumstance.

3. Approach to gathering information

The first question we considered in our planning period was the definition of "effects of the process." After discussions with our clients and review of the comments of observers of validation, we concluded that we should focus on the original purpose of student aid: the encouragement of active pursuit of education, learning, and training. Processes of aid administration should be examined, at least in part, to see if they support this purpose, and the specific version of this question for us would be: Does validation affect students' educational plans and progress?

Despite the numerous suggestions from higher education associations, individual officials, and scholars, that such effects were occurring, we found no data. Scattered schools had done cost studies, which aided in developing a list of variables for that part of our study, but none had more than anecdotes about student effects. The chief proxy measure suggested was delay in students' receipt of funds, but we saw no data even on such a measure.

Concluding that we would need to collect new data, we turned to the problem of what information to ask for and from whom.

(a) Student survey at selected campuses

The ED study is asking officials at the 300 institutions in their sample to estimate the number of students with delayed awards, but nothing else about what those delays mean for students' educational plans and progress. We decided that we would need to gather data directly from students, about their experiences with the validation process, and about subsequent

---

[15] The routine ED data do show students flagged for validation by income, and for each of the ad hoc (PEC) selection criteria, the average change in eligibility that has resulted from validation. No other summaries or analysis have turned up in our search so far. We are considering asking for special analyses for our use during the study, though solid data tapes lag the program by a year.

events which might have been effects of that process. Figure 2 shows a series of questions about students' validation experiences, possible effects of any delay, and other effects of the process.

Figure 2

Questions for students about their validation experience and possible effects

1. Students' experiences with the validation process

   Does the student (S), recall being selected for validation?

   Did S understand why, and what was wanted?

   Could S get whatever was needed to back up application data? Any problem obtaining documentary evidence? Which documents? How solved?

   Did S have questions? Where turn for help (at school, on campus, to central Pell processor by letter or phone, etc.)? How helpful?

   Did S have to submit corrections? Did processor make corrections? Problems with that transaction (errors in correction)?

2. Later events which might be effects of undergoing validation

   Did S Pell eligibility change as result of validation -- or actual grant amount expected? How much? Effect of that?

   Was there delay in Pell award from time expected? How long? Why (process delay, prior delay in obtaining documents, delay on campus after corrected forms received, etc.)?

   Did S have to change educational program or schedule owing to delay?

   Did S have to borrow money as result of delay? From whom? How much?

   Did S lose other financial aid owing to delay in setting Pell amount? (Since some schools won't allocate other aid until Pell amount known, funds may be exhausted for term if validee is last in line.)

   Were there effects on family (impact on other family members' education plans, or intrafamily disputes over access to records)?

Are there other projected/predicted educational impacts of delay or other aspect of being validated (such as weakened articulation of education program with job or further education plans owing to delay)?

The main goal of the analysis would be to understand these students' experiences of validation. But to be confident that we note the extent of delays in awards and other effects which may chiefly be the result of that experience, we plan to gather information from students who were not validated, on their experiences of obtaining their Pell Grant awards and the frequency of delays and educational consequences. As we considered the varieties of experience with financial aid that occur across the millions of applicants and recipients of Pell Grants, we added further elements to our design.

- We must search for effects on those who may not have become Pell Grant recipients -- who may have been influenced by the validation process to alter educational plans to the extent that they do not enter higher education. This may seem far-fetched, but has been suggested repeatedly in commentary about students in groups with the least experience in formal record-keeping and procedural detail, who most need aid to permit post-secondary education of any kind. We have planned a special study to check on this.

- We would like to contrast the validation experience of recipients at different types of higher education institutions.

Statistically valid comparisons among many diverse groupings require sizable samples. Design of the present study is constrained by the time available (only until spring 1984) and the resources tentatively allocated (three staff in PEMD headquarters and some temporary regional office help). We developed fully adequate designs, incorporating all the desirable comparisons, but they would have required gathering information from several thousand students. The ED study has reached that many students (and their parents as well), but it has taken over a year and several million dollars. As we considered the details of contacting students and gaining cooperation even for group interviews at our fourteen case-study sites, much less a national school sample, it became clear that completing those tasks during December in order to have research teams at work in January would be impossible.

Thus although we would prefer other methods, we have concluded that for our exploratory purposes, a survey of students is the only feasible alternative.[16]

Because our sampling plan is complex, we decided to limit our survey to students at five schools drawn from the case study group of fourteen schools. While a survey might seem to offer the potential of a nationally representative sample, our data requirements for sampling make it impossible to develop the sampling frame for a national survey under study schedule constraints. To have student data for the first briefing, we will try to complete our data-gathering and analysis on students at two of the schools; analysis of data from the remaining three student groups will be finished after the briefing.

Our plan is as follows:

1. We will survey recipients of Pell Grants at five of the case study schools, chosen to represent diversity in student characteristics, validation processes, type and control.

2. We want to compare the validation experiences of students of different types, as the subcommittee requested. A major distinction is in degree of financial need shown on the original Pell application. Students also differ in whether validation caused any change in their eligibility for a Pell grant. In addition, we need to compare validated students as a group with other students who were not flagged by the processor and so did not go through validation, as a check on problems students at each school may have in general, quite apart from validation.

3. These necessary comparisons dictate an eight-cell matrix to divide the students we will survey at each school. We propose to survey 30 students from each of the eight cells of Table 4, for a total of 240 at each school. (This number is the smallest acceptable for statistical comparisons between two cells; the cell sizes and the overall school sample size of 240 are not adequate to allow generalization or projection.)

---

[16] We discussed problems of survey research on student-age populations with an official at ED's National Center for Education Statistics (where several national studies of high school graduates have included financial aid items), and with a director of research in a local school system which surveys its recent graduates. Both had warnings of difficulties. However, PEMD survey experts are confident that proper design and follow-up can yield usable results. And practically, we have little choice.

4. We will be surveying students enrolled in 1983-84, who went through validation in the summer or fall of 1983. We will check to see if the processor's records by December 1983 are complete enough to allow them to draw our samples on the schedule we need. If not, we will use school records (manual or automated) to provide the information for sampling. The first institution of the two to be studied for the initial briefing will be in the Washington, D.C. area, to ease the problems of developing the sampling methods (and follow-up techniques as well).

5. Using school records allows us to obtain strong information on current home and school addresses with telephone numbers. This will be a better starting point for our mailing than if we had to rely only on processor records (which contain an address on a Pell Grant application that could be months out of date).

6. The actual sampling process will probably start with creating a subset of about 1000 randomly chosen Pell recipients at a school. The first sort would then be to separate those in the top quartile of need, according to original Pell eligibility, and all others. Subdividing these two sets into those who went through validation and those who did not[17] yields immediately the two non-flagged groups from which to randomly select our survey samples. The flagged group needs to be further subdivided according to change in eligibility, which we expect to have to do manually if we are using school records. We would simply perform the calculations seriatim until we reach the target numbers for each of the six groups. (We will choose replacements in each cell as well.)

---

[17] There is at least one school in the original nine which validates all aid aplicants, not only those Pell applicants selected by ED. Clearly we will have to screen schools so that we have some in our set of five where there are non-validated Pell Grant recipients.

Table 4
Pell Grant recipents to be surveyed at one school

Original Eligibility	Not Flagged	Flagged by Processor For Validation: Large Eligibility Change	Moderate Eligibility Change	Small Eligibility Change
Eligibility prior to validation in top 25% of need	30	30	30	30
Eligibility prior to validation not in top 25% of need	30	30	30	30

Clearly there are major limitations of the chosen approach, including drawing our survey students from such a small number of schools, selected from the judgment sample used for the cost case studies. The survey method of data-gathering has drawbacks as well, in obtaining full information on thoughts and behaviors which occurred some time before. Further, the design is not constrained to select students with equivalent aid histories (i.e. first-year students only); there may be some students who have been through validation more than once. And finally, the design is not constrained to provide for analysis of subgroups of students chosen for validation by each of the four different methods used by ED in 1983-84. (We considered trying to draw only students selected by the random method, but that group is small this year, and there would probably be few cases of students chosen that way at our small number of schools, only part way through the award year. The target for the total number of random choices of students for validation over the whole award year is only 50,000.)

(b) Special study of "disappeared" students

Our design thus far aims to understand the student experience of validation by querying those who did end up enrolling at some institution. Though the institution at which we find them may not be the students' first choice, and though they may have encountered delays or other obstacles along the way to enrollment, nevertheless by sampling recipients we are surveying a group that is pursuing higher education as the aid program was designed to encourage.

Some people, however, may have been affected more seriously by the validation requirements. It has been suggested by some observers that some students may have been deterred from higher education altogether. We have designed a special study of people who originally applied for a Pell Grant for 1983-84, who were found eligible, but who were flagged for validation and who have not reappeared in the processor's records as grant recipient. The problem of discouragement may be most acute with low-income students, so we plan to concentrate our special study on those applicants in the quartile of highest Pell eligibility.

The elements of our plan for the special study are:

1. We will choose three institutions (either from the 14 in the case-study group or from a wider pool), and aim to contact students who indicated on the original Pell application that that school was first choice.

2. From the processor records of 1983-84 applicants with first choices of any of the three schools, we will ask for applicants in the highest quartile of eligibility (according to each school's distribution). These should be applicants who subsequently "disappeared" from the records: that is, they do not appear in the processor's files as having received a Pell Grant at any school.

3. From each school's pool of "disappeared" but highly eligible applicants, we would sample 50 who were flagged for validation and 50 who were not flagged (plus replacements). The total sample would be 300 - 100 from each of three schools.

4. We would use the last address shown on the processor's record as the starting point for telephone tracing, aiming to complete a telephone interview with the applicant to find out the course of events since the Pell application, whether the individual is pursuing higher education, and the influence of the validation process (if any) on the individual's educational career.

There are several difficulties with this approach. First, the aid award year is not over at the time of our sampling, so there could be many months yet in which this year's "disappeared" but eligible student could come in from the cold, enroll, and become a Pell recipient. That is, it may be too early in the year to find the effect that some claim is present. (Sampling students from an earlier year seemed unpromising, as the tracking information in the processor's records would be very out of date, forecasting a low completion rate.) Second, there are a very great many reasons why a

student would not enroll at a school listed months ago as a first choice, including some as routine as not being accepted at the school, or choosing marriage rather than education for now. Then, too, students' testimony on the effects of validation on plans, among all the other vicissitudes of life, will be at best a partial source of data. The effects we are looking for are likely to be rare, our ways of finding the people we need to talk to are limited, and the credibility of the information is only fair: clearly this aspect of the study is a long shot, though intriguing.

(c) Secondary analysis of ED student data

As one more attempt to locate effects on students, we plan a secondary analysis of the most recent ED file on Pell applicants and recipients for a complete award year. Because the file is massive (with over 5 million records) and complex, spanning 2 years from the opening date for applications to the close of awards, it takes the processor months to deliver. The tape we will have to use is probably 1981-82, which is the year before the significant expansion of validation. Still, we may be able to create some useful subfiles from the 300,000 students who were validated that year, to compare them with the rest. We could examine whether the students flagged for validation become recipients at any different rate than non-flagged students (controlling for change in eligibility), or whether flagged students attend their first-choice school at any different rate. During our planning period we have not had access to the ED file, so our design is sketchy at best, and if further reconnaissance proves unpromising, we will consider dropping this activity.

The general plan of analysis for the student information is to provide a description of the validation experience of the 900 flagged students in our sample, both the process questions and the effects, and to compare the effects with experience of the 300 non-flagged students where we can look for rates of delay or change of plans arising from other causes. Beyond the comparison of validated students with others, we will look for any differences in the validation experience and its impact for subgroups: by school, initial eligibility, or extent of change in eligibility. We will be especially interested in impacts on those with small eligibility changes (those for whom validation probably yielded no payoff on the main policy goal of detecting error), and on students with high eligibility (limited family resources) who may be expected to have the least ability to cope with the process. Detailed analysis plans are again in Appendix C.

E. Alternative Approaches and Methodologies

1. Description of the questions

In the final three of the thirteen questions posed in the Congressional request letter, attention turns to setting the current approach to Pell Grant error correction in a larger context. The questions are:

> What are various approaches to the problem of error in Pell Grant awards?
>
> Are there experiences of other Federal agencies, or of the private sector, that offer useful suggestions on this problem?
>
> Are there alternative methods for preventing or correcting award errors that could offer a better balance of positive and negative effects?

2. Rationale

There is clear reason to examine alternatives to any policy and methodology which, like validation, affects a very large number of people and which has aroused contradictory comments. Even by trial and error, without benefit of thorough-going research or evaluation, alternate methods may be found that would improve on current practice. There is even more reason to look at alternatives as part of this inquiry, as we may be finding evidence of burdens on institutions and individuals serious enough to call plainly for relief.

Comparison is an essential aspect of the process of understanding also (as parents of only-children are constantly reminded by those with larger families). Education is not alone in facing the problem of verifying information as part of proper awarding of a need-based benefit, or the problem of deciding appropriate strategies to cope with errors in that information. The original Quality Control Study noted that at least five other Federal programs were very similar; recent OMB concern with verification suggests the ubiquity of the issues (and error rates quoted in recent news stories for other programs show that other programs are far from perfect too -- 13 percent for Food Stamps and 8 percent for AFDC). Since our concerns in this study range from the basics of gathering and interpreting error data, to goals and methods of a detection and correction effort, to the impacts on those involved in the methodology, it makes good sense to examine these generic topics in comparable program contexts to see how ED stacks up.

Most generally, we are concerned that debate over errors and their correction be placed in a larger context, which this set of questions can provide. A benefit program has multiple goals, and whether we like to admit it or not, some level of

error will inevitably have to be tolerated, because the price of perfect virtue would be too great in defeating other program goals. We understand the subcommittee to be asking for help with the kind of trade-off analysis this suggests. For instance, we could begin with a list of general Pell Grant program goals suggested by the Department's contractor in the first Quality Control Study of 1980-81 awards:[18]

- Educational effects (behavioral effects)
- Simplicity (number of forms, data elements to be provided)
- Equity (sensitivity to student/family differences)
- Integrity (minimum fraud/abuse)
- Governance neutrality (states rights and institutional autonomy)
- Cost of delivery (share of appropriation)

Approaches to error (to maximize the integrity criterion) could then be weighed for their impact on the other criteria. We have seen no ED analyses of this sort, though it is to be one purpose of the work under Topic A, current policy, to be sure we look systematically for any such ED analyses.

In addition to comparing alternatives for their effects on other program goals, we would like to think about how alternatives could be arrayed as between short- and long-term implementation, and between diverse purposes such as prevention, prediction, detection, and correction of errors. We have seen no comprehensive matrix that set out options in this way.

This conceptual work, linked to some concrete exhibits of practice elsewhere, and both joined to data on the ED current policy, practice, and impact -- could form a unified package with potential to be very useful in sorting out next steps in the administration of the Pell Grant program.

3. Approach to gathering information

The basic tasks are simple to state: to develop a set of goals or criteria which describe important values in the Pell Grant program and which can be used in thinking about strategies to combat error; to gather information about alternative strategies to allow arraying them against the criteria; formulating suggestions about ones that seem to have substantial advantages, such as balancing effects across criteria or minimizing negative

---

[18] Items taken from Volume 2: Corrective Actions (McLean, VA: Advanced Technology, Inc., 1982), p. 7-4.

effects altogether. To reach the basic framework of goals or criteria characterizing the Pell Grant program, we will review a number of sources, especially ones which have discussed quality control issues. These include:

- The Quality Control Studies (both current and 1980-81)
- Materials from the Credit Management project in the ED Comptroller's office, which includes a complete redesign of the aid delivery system, which has been commented on in public hearings
- Materials from the GAO/HRD project examining issues in the verification of eligibility in many Federal programs including education aid programs
- Materials from Congressional sources containing reactions to past proposals for validation initiatives (such as the multiple rejections of ED proposals to reprogram funds in 1982 to cover expanded validation).

These will aid us in developing a full understanding of the several criteria that are important in the program, and initial views of various parties about how error-correction fits among them.

We will then search for a wider set of examples of methodologies for treating the diverse stages of the error problem. Initial review suggests that at least three other Federal programs have similar problems of getting good information to aid in award determinations: AFDC, Food Stamps, and the National School Lunch Program. All three have a history of work on quality control issues such as understanding error in the first place, taking action, and evaluating impacts. We plan to visit individuals in the appropriate agencies to find out details of program size, error rates and tolerances, interactions with applicants, the nature of validation (pre- or post-award, documentation required, selection of individuals if not 100 percent), and any prevention efforts that have been attempted.

We will visit IRS to consider an often-mentioned possibility of matching tax data from returns to Pell Grant application figures in the Pell Grant program. Also at IRS we want to learn more about their version of discriminant function analysis or what ED calls error-prone modeling, used to select tax returns for special reviews.

The Congressional request suggests review of private sector experience as well, but there will be few resources for much work along this line. Nor are we sure what private sector transactions are usefully comparable. We will consult with GAO experts in banking to see if banks' policies on financial transactions provide clues to acceptable error rates and how the private sector weighs information costs against error.

Washington-based associations in the banking field may provide another set of sources for information about the state of the art. Accounting generally must deal with the problem, in the long-running debate between examining individual transactions to check for errors, versus examining systems of control intended to prevent or detect error. We can explore the area with GAO colleagues in AFMD. In all the non-government forays, though, we will keep in mind Kaufman's observations about the special moral and political status of public money and processes for distributing it, which make it hard to compare practice across public/private sector boundaries.

The conclusion of work under this heading is to estimate the impact of diverse strategies concerning the problem of error in Pell awards on the program goals and criteria, through a matrix of both. Analysis will focus on tradeoffs across the criteria, and analyses of long- vs. short-term approaches, mechanical vs. systemic approaches, prevention vs. after-the-fact approaches. We will not have time or comparable data to do formal tests of approaches using Pell Grant data, to see precise predicted outcomes. But our analysis will be suggestive, pointing out key alternatives that have shown promise in related circumstances.

IV. POTENTIAL USEFULNESS OF THE STUDY

A. Relevance to user information needs

The plans for the study follow closely the questions listed in the letter requesting the GAO evaluation (included in Appendix A). One or two of the subquestions will receive less emphasis than the rest, but we expect to be able to address all the questions with evidence not now available. The work plans give priority to the survey of institutions and the institutional cost case studies, in response to committee staff comments that the effects of validation on schools, colleges, and universities is the topic of most interest.

The broad sweep of the full set of questions, with their attention to executive branch policy processes, the quality of research used as the basis for policy, the costs and impacts of the policy on various groups, and alternatives to the present policy, should provide a comprehensive approach to the subject of validation, which will help the committee begin oversight on a subject that has not been the focus of much Congressional activity up to now. Thus, usefulness of the overall study may be enhanced by its comprehensiveness.

Research and evaluation studies always have limitations, which may affect their use by intended audiences. One major safeguard against disappointments of this kind is clear communication of limits at early stages of project planning. We will talk with our clients about the several cautions noted in earlier pages of this plan, on such issues as:

- limitations of generalization from the institutional cost case study data

- limitations of generalization in the institutional survey sample to sub-groups of institutions

- the absence of independent data in our study on the extent of error reduction attributable to validation (obtainable from the ED Quality Control Study, not from our study)

- likelihood of having only part of the student data available at the time of our briefing, owing to schedule constraints

B. Timeliness of study results

The subcommittee requesting the study is in the process of reauthorization hearings on federal student aid programs throughout the period 1983-84. GAO information on the validation of Pell Grant applicants needs to be available to the subcommittee as quickly as possible, so that issues may be considered in development of authorization legislation. In the short time available from the date of the request (September 1983) to the end of the legislative term some time in mid-1984, GAO can complete certain limited data-gathering and analysis in time for use in spring 1984.

The information thus developed will be presented in an extensive briefing, now scheduled for late April 1984. (See Appendix B for an outline of the briefing topics.) As mentioned just above, it will be important to communicate to the requestors the limitations of data and analysis that will be available at that point in our work.

Further data collection and analysis will continue in the months after April 1984, concluding in August 1984, when a report will be drafted from the total body of information. This report will still be timely and useful when it appears early in 1985 following internal review and agency comment, as most observers agree that Congressional action to reauthorize federal student aid will not be completed until the 1985-86 session of Congress. Executive branch action on validation policy for 1985 will also be developed in mid-1984 and may be able to use ideas or directions from our study as the study draft is sent for agency review.

C. Presentation

The briefing in April 1984 is the major method of communicating our findings and analyses to the requestor. Since the study is exploratory, the issue will no doubt arise as to further work which may be needed as a result of our findings so far. It is possible that the subcommittee will request

testimony from GAO, if the results seem to warrant further exploration by various witnesses.

We plan a formal written report to the subcommittee chairman, based on all of our work, to be released early in 1985.

D. Possible impacts

The policy and procedures for validation of Pell Grant applicants are not now covered by legislation. Thus, as one result of our review, Congress may decide to create a statutory framework for policy on verification of applicant information, in the Pell Grant program or for federal student aid programs generally. Or, Congress could signal its intentions with specific report language at the time of the 1985 reauthorization of the Pell program.

If our information and analyses are considered in ED, there could be impacts on policy and procedures of validation as well, as in the absence of specific statutory guidance, the ED approach is constantly changing. This offers the chance of many different kinds of impact on agency action.

Through legislative or agency action, the kinds of impact we might have could include:

- reduced burden and dollar costs of institutions
- reduced burden and less effect on educational plans for students
- decreased costs to ED for the validation process
- improved efficiency of targeting of Pell Grant funds through improvements in validation process.

These are hard to be more precise about at this point, since the major purpose of our work is to determine if substantial burdens exist, which may then merit development of alternative policies and methodologies for validation.

COMMITTEE ON EDUCATION AND LABOR

U.S. HOUSE OF REPRESENTATIVES

320 CANNON HOUSE OFFICE BUILDING

WASHINGTON, D.C. 20515

SUBCOMMITTEE ON POSTSECONDARY EDUCATION

September 7, 1983

Mr. Charles A. Bowsher
Comptroller General of the
United States
General Accounting Office
Room 7026 441 G Street, N.W.
Washington, D.C. 20548

Dear Mr. Bowsher:

The Subcommittee on Postsecondary Education is preparing for reauthorization of the Higher Education Act, including the Pell Grant program of student financial assistance.

This program has grown in the last decade, so that it now serves over two million students and distributes over two billion dollars each year. As a result of several Department of Education studies, concern has been increasing about errors in awarding funds in the Pell Grant program. In response to this concern, the Department developed a methodoligy known as "validation," which requires over one million of the applicants for Pell Grants to provide additional detailed documentary evidence to support their applications. We understand that the Department is collecting information on the effectiveness of this methodology as measured by the errors corrected this year through validation, and already intends to expand the procedure to other student aid programs as well.

However, testimony to my Subcommittee and to the National Commission on Student Financial Assistance, has raised questions about the validity of the Department's methodology, its costs, burdens, and other impacts and effects, though witnesses have lacked extensive data.

I am, therefore, requesting that the GAO conduct an exploratory study to see whether these concerns merit more detailed examination. The Subcommittee is interested in obtaining information on the questions presented in the attachment to this letter.

Discussions between the Subcommittee staff and the staff from your Institute for Program Evaluation have indicated that the Institute would be able to provide us with a briefing and preliminary informa-

tion from your exploratory study by early next year. A written report could follow as soon as possible after that if we later decide that is useful.

Thank you for your cooperation in responding to this request. If you have any questions please have a member of your staff contact Bud Blakey or Maryln McAdam of the Subcommittee on 225-8881.

Cordially,

Paul Simon
Chairman

PS/mmg

attachment

Questions for Analysis Based on Exploratory Study

Topic 1: Current policy

(1) What is the goal of the current Education Department (ED) policy on validation of Pell Grant applications?

(2) How did the Department decide on the current goals and methods, and with what consideration of burden?

(3) What does it cost ED to carry out its current policy and methods of validation?

Topic 2: Methodology

(4) Does ED have reliable data on award errors on which to base policy, and have those data been interpreted using appropriate methods?

(5) Are the methods for selecting students for validation statistically sound? Are they suited to the policy goals?

(6) Does the Department evaluate its methodology and use the findings in regular improvement of its approach?

Topic 3: Effects on institutions

(7) What costs are incurred by the diverse types of institutions of higher education in doing the validations?

(8) Are there other effects of the validation process on institutions? Are these more serious at particular kinds of institutions?

Topic 4: Effects on students

(9) What are the effects of the process on students selected for validation?

(10) Are there effects of the validation process that fall disproportionately upon particular groups of students?

Topic 5: Alternative approaches and methodologies

(11) What are various approaches to the problem of error in Pell Grant awards?

(12) Are there experiences of other Federal agencies, or of the private sector, that offer useful suggestions on this problem?

(13) Are there alternative methods for preventing or correcting award errors that could offer a better balance of positive and negative effects?

APPENDIX B

OUTLINE OF BRIEFING TOPICS TO BE COVERED

1. Review of Congressional questions and highlights of key findings

2. Findings from survey and case studies concerning costs and impacts on institutions from current validation policy

3. Findings from survey and interviews with students concerning impacts of current validation policy

4. Findings concerning ED aspects of current policy
   -- costs of current policy (at ED and processor)
   -- quality of error data used in policymaking
   -- quality of student targeting methods
   -- extent of consideration of burden in designing policy and targeting methods

5. Findings from review of other agency strategies against error in entitlement applications

6. Tentative conclusions and review of remaining work to be completed

DATA ANALYSIS PLAN

This appendix provides a brief presentation of basic data analysis plans for each of the five study topics in the congressional request letter and the main text of this design paper.

A. Current policy on validation

Questions under this topic (Section A of the main text) include the goals of validation and the decision process leading to the current approach and the costs of validation to the Department of Education. Data sources for understanding complex multi-faceted organizational actions such as the Department's validation effort (and any other responses to Pell error) will include documents and interviews. By the end of our fieldwork these will be voluminous, and we will be to some degree, "insiders" in certain units of the Department and elsewhere. The analytic task will be to draw from the set of materials assembled (documents, interview transcripts, etc.) a factual picture, both useful to the client and faithful to how insiders see the situation, of what the Department has been up to and why. At the descriptive level, we will require multiple observers' agreement to phenomena that are otherwise undocumented before we include them in our account, though perspectives of a single important official may be significant enough to be discussed alone. Important matters that are viewed one way by some, another way by others, may require re-interviewing (even confronting one observer with the discrepant perception of another) for confirmation or clarification before analysis is done. Thus, analysis and fieldwork are not entirely distinct phases of the project plans. For more inferential analyses, such as causes of actions or outcomes, we will move from hunches developed during the data gathering to block displays of elements that seem to be related, which can guide further retrieval of evidence from documents and notes. Initial texts developed from the data may need to be chronological to capture the flow of events before revising into more analytic formats such as around persistent themes that cut across a chronology. In this case, writing the initial text in light of the voluminous field data from which it will be assembled will be a process of discovery as much or more than a process of presentation of discoveries already made. Review of the arguments and evidence by others will be an important check on their plausibility -- in contrast to the review of procedural steps taken (sampling, analytic method used) in order to warrant that quantitative data are correct.

(See M. Miles and A. Huberman, Qualitative Data Analysis, Sage, 1984, generally on methods of analyzing the type of data we will have in several of the parts of this study.)

The final part of the study of current policy, is study of its costs. Here we will request data from the Department and its contractors, and accept that data without extensive verification (since our purpose is exploratory, rather than an actual audit). Our aim is to describe the funds spent by the Department, and the addition of the various categories of funds spend will accomplish that. The only difficulty we foresee is getting the costs of validation-related corrections sent to the Department's application-processing contractor, but that is a data-gathering problem more than an analysis problem. In order to gauge the size of the costs we find using these methods, we will gather Department data on the full range of Pell grant program expenses and compute a fraction that validation's costs represent. The costs (at the Department, at the schools or both) can also be compared to the "savings" in reduced Pell grant error, both intended and actually realized. These savings can be estimated from the Department's error research, which was done in 1980-81 before validation was expanded and again in 1982-83, the year of the expanded effort we are focusing on in much of the review.

B. Methodology issues

Questions here (Section B of the main text) center on the Department's error data and their quality, the Department's methods of selecting students for validation, and the Department's general use of evaluation in improvement of validation and related error policies. In order to judge the error research, we will commission independent statistical consultants to apply standard social science research norms in reviewing the published reports of the Department's "quality control" research. Where GAO's own analysis and that of the consultants agree, we will consider the strongest points for our findings. If there are points in more dispute, we would be more cautious.

To evaluate the statistical soundness of the Department's selection methods, we will first develop a technical description of these methods from documents and interviews. Then we will examine the procedures to see if they match the stated policy goals of the Department (if any). If no Department policy exists against which to judge the selection methods, we can judge them by comparison to social science norms (as in the case of one

method, called "random selection" by the Department and used for some research purposes). Alternatively, the standard could be the congressional mandate that the selection process find those with the greatest likelihood of error.

To answer the question regarding the Department's use of evaluation data, we will gather data about the Deaprtment's evaluation of its efforts from contractors and from internal research or information providers and from officials who may use it in decisions. The analysis of the documentary and interview data will be done as described under the previous topic area.

C. Effects on institutions

Questions about the effects on institutions (Section C of the main text) are straightforward:

> What costs are incurred by the diverse types of institutions of higher education in doing validation?
> Are there other effects of the validation process on institutions? Are these more serious at particular kinds of institutions?

Analysis of validation costs and effects will proceed in three independent stages: (1) analysis of the case study information; (2) analysis of the survey questionnaire information; and (3) integration of the case study and survey findings. Each of these stages is discussed below.

1. Analysis of Case Studies

The case study analysis will be conducted by Touche Ross (according to our specifications), and will focus first on a detailed description of all direct and indirect costs of financial aid administration at about 14 schools in 1982-83 for each individual case and the distribution of costs by aid administration function. Comparative analyses of 1981-82 and 1982-83 costs will be conducted for nine (of the total of 14) schools that provided this information in 1981-82, again for each individual case.

Once the individual cases are analyzed, we will examine all the cases for general patterns of costs and possible factors contributing to differences among the schools in costs,

functional distributions, and changes in costs. Our search for patterns will not rely on the computation of aggregate statistics across the cases. While we may compute some aggregate statistics that are of greatest interest (such as the cost of an average validation) for use as descriptive summations of findings, the analysis itself will concentrate on what is happening at the individual case level. For example, we will examine changes in the costs of Pell validation in each of the nine before-and-after cases to draw general conclusions about patterns of cost increases or decreases rather than taking the alternative approach of computing a cross-case mean change and relying on its directionality to tell us whether or not costs have increased or decreased. We will deflate the most recent cost data so that multi-year comparisons can be made in constant dollars.

This focus on individual cases is the most desirable approach for the analysis of data across several case studies. On the one hand, it exploits the richness, depth, and uniqueness of each individual case, which are lost in aggregate statistics. On the other hand, it preserves a recognition of the limitations of case studies (e.g., limited generalizability and small sample sizes), which tend to be obscured when cross-case analyses rely upon aggregate statistics.

Once general patterns in costs and cost changes are identified, we will examine the specific validation procedures at each school to see if there are factors associated with increases or decreases in costs.

2. Analysis of Survey Data

The analysis of the survey data from a sample of 400 schools that participated in the Pell program in 1982-83 will proceed in three phases. In the first phase, the survey data will be prepared for analyses.

The surveys will be edited and open-ended responses coded prior to keypunching for computer analysis. To edit the surveys, all items on each returned questionnaire will be reviewed by project staff to check for missing data and inconsistent, ambiguous or confusing answers. We will call survey respondents for additional information and clarification. To code the open-ended responses, we will extract and record the verbatim responses for all open-ended items from the first 50 or so surveys returned. These responses will be studied carefully and coded into categories that summarize and describe the responses.

The codes will then be applied to all surveys by project staff. (Code categories will be added as necessary as more surveys are returned.) A single coding supervisor will insure consistency and resolve specific problems. Once editing and coding are completed, each survey will be given a final check by project staff before being sent to keypunch.

Standard practice at the GAO keypunch contractor is to key an entire survey, then do it again and check the two together for 100 percent consistency. This process of verification will be used for all of our surveys.

After keypunching is completed, we will use the computer to conduct additional checks of the data for accuracy. First, all of the data will be listed and visually inspected for inconsistencies and inaccuracies. Second, we will conduct range and logic checks on each survey item. Whenever inconsistencies and errors are identified, we will go back to the original survey questionnaire to trace and correct the problem.

The second phase of survey data analysis will focus on general exploration of the data. This will involve computing a number of statistics on each item in the survey, including dispersion statistics such as ranges, means, medians, and variances, as well as constructing contingency tables (or cross-tabs), which will use the classifications of schools by type and control used in drawing the sample. The dispersion statistics will provide information about the overall response patterns for each variable. The contingency tables will provide information about whether response patterns vary by kind of institution. During this phase of analysis, we will also compute weights for the returned questionnaires and develop a "weighted" data set. These weights will allow us to use the sample to estimate population or universe responses. In other words, we will be able to extend our analyses beyond our sample of 400 and generalize to the population or universe of responding schools. Developing the weights will involve adjusting the universe to exclude non-respondents and hence must wait until survey data collection is completed.

We have chosen to adjust the universe downward in size to exclude non-respondents because it is the most conservative way of adjusting for non-response. Some analysts prefer the opposite approach of adjusting the sample upward to include non-respondents in the universe estimates. However, this approach rests on the assumption that the non-respondents'

answers to the survey questions would have been exactly the same as the respondents'. PEMD's sampling statisticians agree with us that this assumption is generally inappropriate.

We do plan to compare non-respondents and respondents on several variables taken from a third data source, such as the average number of Pell recipients and average level of recipient need. If we do not find great differences between the two groups on these variables, that will suggest the possibility that their survey responses may not have differed greatly, and that our findings may well apply to the non-respondent portion of the total universe as well. This tentative extension of conclusions to the non-respondent portion of the universe is considerably more conservative, and more appropriate, than adopting the assumption that the two groups' responses are identical and building this assumption into all of the analyses.

All of the dispersion statistics and cross-tabulations computed during this second phase of survey data analysis will be computed for both the sample alone before the weights are applied and for the universe after the weights are applied. This will allow us in essence to "double-check" the general picture of the data at both sample and universe level.

The dispersion statistics and contingency tables will be inspected closely in order to refine strategies and approaches for the third and final phase of the survey data analysis: examination and testing of the data to identify patterns of validation activity, resource allocations, and opinions about validation procedures.[1] For example, the contingency tables may suggest data reductions and relationships among variables that should be explored. The dispersion statistics (and the tables) will show which (if any) variables should be dropped from further analyses because they lack sufficient variation. In addition, these basic analyses will probably reveal distributional characteristics that must be taken into account in subsequent analyses.

Depending on what the dispersion statistics and contingency tables show, we may compute additional exploratory statistics

---

[1]We will not analyze the school questionnaire items on possible student impacts at this point, but will reserve them for inclusion in our analyses of effects on students, which are discussed later in this plan.

before moving into the third and final phase of analysis of the school survey data. For instance, we may compute correlation matrices for all or selected variables and data reductions such as cluster or factor analyses. At this time, it is impossible to say exactly what sorts of additional basic analyses might be appropriate. Nevertheless, it is important to note that we do not intend to truncate the initial basic explorations of the data.

The third phase of survey data analysis will focus on answering the evaluation questions specified earlier. This phase will begin with a search for effects. Next, where we find effects, we will examine the data to see whether or not these effects differed for different kinds of schools.

There are many ways to think about effects in the complex regulatory and policy environment of Pell validation. However, the information we gained in planning this study suggests that the effects of validation on schools should follow strongly from the Department's actions. For example, the overall pattern of changes in Pell validation activity at the schools between 1981-82 and 1982-83 should mirror changes in the Department's requirements: the number of students validated should increase while the number and complexity of application items validated should decrease. This directionality might "wobble" or vary a bit because of unique, institutional level policy decisions. For instance, some institutions may choose to move beyond the Department's requirements and validate more complex, non-required application items. Others may chose to limit validation to the application items required by the Department but validate 100 percent of their Pell applicants. Similarly, the schools' resource allocations for validation should shift in the direction of greater effort. For instance, more training might be required, more overtime demanded of staff, and the use of automated data processing might have increased.

Given these expectations, we plan first to examine the data to see what kinds of shifts in validation activity and resource use occurred in schools between 1981-82 and 1982-83. Our questionnaire will contain several items designed to measure activity level and resource allocations. Again using both the unweighted and weighted data sets, we will examine the patterns of change for each individual variable and across this entire group of variables. (Statistical tests will be performed as appropriate.) Should an effect be observed, we will turn to contingency table analysis to assess whether or not the effect varied for different types of schools.

We will then examine changes in validation activity and resource allocations between 1982-83 and 1983-84. This will allow us to make estimates about the stability of validation effects. (We plan to use the same approach as for examining patterns of change between 1981-82 and 1982-83.)

Next, we will examine the questionnaire items on opinions about various aspects of validation. For the most part, these items will not be replicated across years, but will either concentrate on aspects of 1982-83 validation or elicit an overall opinion about validation experiences in general. Again, we will examine each individual opinion item and the pattern of responses across the set of several opinion items.

We plan to look at the levels of satisfaction as well as dissatisfaction on these items. This will allow us to capture a more complete range of opinions about specific aspects of validation and about validation in general (as revealed by both the "general opinion" questionnaire items and the pattern of responses across all of the opinion items). We plan to use contingency table analyses for all of the opinion items to assess whether or not opinions vary for different types of schools.

As the final step in the school survey analysis, we will look back across the analyses of validation activity, resource allocations for validation, and opinions about validation, and see whether (and how) it all "hangs together." This will be a very important step, for the coherence of findings across these three domains may be very revealing. For example, validation activity and resource allocations may have increased dramatically, and opinions about validation may be very positive. This is a counter-intuitive pattern, and would merit additional analyses (e.g., deeper examination of original, open-ended responses and comments on returned questionnaires) to seek an explanation.

3. Integration of Survey Findings with Findings from the Case Studies of Financial Aid Administrative Costs

Once the survey analyses and the case study analyses are completed, our task will be to integrate the two sets of findings. This will be a critical step in the analysis of effects on schools, for it will not only complete the picture of school level impacts, but also will serve to "cross-validate" these two separate snapshots of validation's effects.

The case studies and the survey measure different aspects of validation impacts: dollar costs and other, more general costs. Yet, the two are related, and some general congruence of patterns of findings should emerge. For example, if the case studies show dramatic and severe increases in validation costs as a result of the Department's expansion in validation, the survey might be expected to show similar increases in the other, more general costs of validation. If this sort of congruence does not occur, it will raise questions about the "validity" of our case study and survey samples, particularly the smaller, non-random case study sample. If congruence does occur, our confidence in both our survey and case study findings and conclusions will increase substantially, for in effect we will have discovered cross-confirmatory patterns in two separate, independent samples and using two separate, independent methodologies.

Integrating the two sets of findings will be in essence a "data-grubbing" task. There are no formulae for accomplishing this. Rather, it requires careful, in-depth examination and comparison of the two sets of findings, and much discussion and debate. Where the two sets of findings seem to be in conflict, we will have to go back to the data underlying the findings to attempt to resolve the apparent contradiction. Where the two sets of findings seem to be congruent, we will have to take special care to determine that the patterns are broad-based and strong and not a function of isolated variables.

D. Effects on students

As we also discussed earlier in the main text of this paper, the questions about effects on students are again straightforward:

> What are the effects of the process of validation on students selected?
> Are there effects of the validation process that fall disproportionately upon particular groups of students?

The analysis of the student effects data will be very similar to the analysis of the school effects information. Each of three student data sources--the survey of school financial aid officers we discussed in the preceeding section, a mail survey of students at several of the 14 schools that participate in our case studies, and telephone interviews with a sample of 1982-83 applicants who were found to have the highest eligibility for but

our measurements). If congruence does occur, it will increase substantially our confidence in our findings, because, as was true for the school effects analysis, we will have discovered cross-confirmatory patterns in three separate, independent data sources using a variety of different indices of student impacts.

E. Alternative approaches

In this final topic, we are asked to locate various approaches to the problems of error, in a general sense, as well as search for specific experiences of other organizations facing similar problems. From these explorations we are asked to explain any alternatives that may better balance positive and negative impacts of an applicant data-control system.

The data sources include literature and interviews. Our analysis will begin by sorting ideas discovered from these sources for relevance to specific problems found in the Pell error research and in our analysis of the Department's response to the findings. We will examine specific ideas for their potential benefit, as a second criterion, but we acknowledge that we can only go so far with this analysis. Our data on potential benefit will come from our informants (or the literature) and will most likely reflect experience in other sectors beyond education -- with differing funding history, program purposes, and history of relations between the federal government and others in the program delivery system, all of which can affect the implementation and effects of quality control (QC) measures. And the data on that experience may only be participants' opinions; we will seek evaluations of QC measures elsewhere, but we have been told informally that even in more mature program areas with QC efforts, formal evaluations that conclusively show the effects of specific QC measures are rare. We will informally test the alternatives for their potential success in balancing positive and negative effects in the Pell grant setting. We will include a question about specific alternatives in our national institutional survey to campus aid officials, to see if any receive widespread support. But we will not be able, within our time and resources, to gather even more solid information, such as would be needed for firm recommendations of specific alternative strategies. Such information could include the results of specific pilot tests (as when GAO actually tested on a sample of cases an error-prone model for its potential usefulness, compared to methods already in use to locate problem cases in a D.C. welfare program). Another method we will not be able to use would be simulation of a novel approach, using a

database of Pell applicants or recipients. We will look for any such efforts by the Department, and report on their results if they seem promising, but we will not be able to do new simulations ourselves.

Further, since the error data is rather new to the Pell program, our analysis of policy and decision-making may show that there are prior or more general questions of the role of quality control in the Pell program that need attention, which will make it less important to test in detail specific technical remedies to reach firm recommendations. Thus analysis plans in this area depend somewhat on results of our study in other areas.

F. Further analyses for conclusions and recommendations

We will need to bring together the data and findings from each of the topics already discussed in order to reach general conclusions about Pell grant validation for our congressional audience. In the most general sense, the study is a search for hitherto unmeasured side effects of a treatment, which (once measured) can be compared to the main effects. That is, we must analyze our data to see the impacts of validation on error (the desired main effect) and also its side effects or impacts on the costs and other burdens borne by those performing the validations--schools and students.

This final analysis has two parts. The first is the comparison, or balancing, of main and side effects. This was suggested in the brief discussion of comparisons of validation costs and returns under topic A above. School costs will be determined from our school case studies and that analysis can be corroborated by the few cost questions on the national survey. We will also measure costs incurred by the government from Department and contractor records. The returns can be crudely measured by data on the projected dollar consequences of various errors, found in the Department's Pell grant error research reports. We have dollar error figures from two different studies, one done in the year before the expansion of validation (1980-81), and the other done in the year of expanded validation (1982-83). Thus, the marginal or added effect of validation can be derived from comparing the error figures in the two studies, as "before and after" measures. (This analysis will have to be subject to many technical caveats, since the programs were of different sizes and operated under different rules in the two different years being compared, and since research procedures for gathering the estimates may have differed.) Especially if we

find major side effects (impacts on costs especially, but also in less tangible dimensions such as delays or changes of academic plans for students, or dramatically altered work priorities for school staffs), it will be important to see if these are balanced by reductions in grant award error.

We have thought ahead about the possible results of this analysis. For example, we will be interested if there is a weak effect on error because Department officials have spoken with great confidence to Congress and the higher education community about the error-reduction effects they expected from expanded validation. Campus officials have scoffed at these predictions, based on their criticisms of the underlying error research, which they believe inflated error rates and aroused unrealistic hopes for the potential impact of corrective actions. The side effects are an unknown; we heard predictions by associations representing school officials that they will be heavy and reassurances from Department officials that they will be negligible. Our study will attempt to bring together the data on both kind of effect and to clarify the balance that has come about between error-reduction and burden.

The second part of this final analysis will follow and depend upon what we learn in the first, as we search for explanations of the effects we see, and their comparative balance. One type of possible outcome we may observe from the expansion of validation is a strong impact on error rates, but side effects of such size as to raise questions. This will lead us back to our data on Departmental decision-making to see whether the side effects were considered in designing the validation methodology and to see if alternatives could better balance positive and negative effects. A quite different possibility is that validation will be seen to have had small effect on error, which will transform the analysis from one of tradeoffs to a more general one concerning the Department's overall approach to the problem of error and how such an ineffective methodology came to be the chosen policy. Again, we will turn to our data on the goals of the Department's error policy and how that policy is set, but the focus will broaden as we ask "of all the alternatives we are aware of (as we learned about under topic E), is the Department aware of these, and how have some been chosen and others discarded in the process of setting the current direction?" Were other potentially effective alternatives not considered? Especially if the results of policy have been meager, we will be returning to our data from work under topic B on the Department's evaluation of policy in this

area to see if weak results had been signaled or not, and the Department's interpretation of any signals.

From these analyses we may observe a decision process not sensitive to the side effects, and Congress could consider if burden reduction is an objective that should be weighted more heavily by the Department in future. And if error is little reduced, we may conclude that even the main effect is so weak that Congress also may want to guide the Department concerning the priority that should be given to award accuracy. Congress has not had the opportunity to consider the issue of Pell grant award accuracy at reauthorization, since the reliable measurement of Pell error is new since the last reauthorization in 1980. Thus, we cannot perform a simple analysis to report on whether or not the Department's validation effort meets a congressional criterion. Our work will be useful in a more general way to aid Congress in considering for the first time, the importance of award accuracy among other Pell program objectives and priorities, and what can be done about it if current approaches are not fully effective.

DESCRIPTION OF THE PELL GRANT PROGRAM AND VALIDATION PROCESS, WITH CHRONOLOGY OF POLICY EVENTS

The Pell Grant Program

This program, administered by the Department of Education's Office of Student Financial Assistance, is the largest of the student aid programs authorized under Title IV of the Higher Education Act of 1965 and its amendments. In the 1981-82 academic year, the program offered grants ranging from $200 to $1,800 to help eligible individuals further their postsecondary education. Since its inception in 1973, the number of recipients has grown over twelve-fold. During the 1981-82 school year, 2,709,000 individuals shared over $2.4 billion in Pell Grants. About 66% of the recipients attend public institutions; about 21% attend private non-profit institutions; the remaining 12% attend private profit-making institutions.

The primary feature which distinguishes the Pell Grant program from other forms of financial assistance is its entitlement concept. All students meeting certain ceriteria are guaranteed aid, with the amount determined by financial need and educational cost.

How the program works

Participating institutions are responsible for the day-to-day administration of the Pell Grant program. As part of an agreement entered into with the Secretary of Education, each school is responsible for ensuring that the program is administered in accordance with rules and regulations. Specific rules the institutions are committed to enforcing include determining student eligibility, calculating and disbursing awards, enforcing standards of academic progress, and calculating refunds and disbursing them.

An Office of Program Review in ED conducts periodic on-site reviews to see that schools maintain appropriate administrative capabilities, comply with rules, and apply satisfactory accounting practices to keep track of federal funds. ED requires that each school receiving Title IV funds be audited by an independent public accountant at least once every two years. Accreditation, without which a school cannot receive Title IV funds, is a means used by ED to assure a basic level of quality instruction and consumer protection through peer review. Like accreditation, a state license is necessary for a school to receive Title IV funds and implies conformance with minimum standards governing the quality of education.

Student award calculation

To be eligible for a grant, an individual must meet certain residency requirements, be enrolled at least half-time in an eligible program in a Pell-participating school, and have sufficient financial need. Financial need is determined from a formula developed annually by

ED and reviewed by the Congress. Applied consistently to all applicants, this formula considers such indicators of financial strength as income, assets, and family size to produce a student eligibility index. The greater the financial need, the smaller the index.

Students must be pursuing undergraduate education only and cannot hold a bachelor's degree already. If male, an applicant must be registered for the draft. And applicants cannot be in default on any federally-guaranteed or insured student loan. Students who have attended higher education at any other school than the one where they are applying or receiving a Pell grant must present a Financial Aid Transcript to demonstrate that no loans are in default or refunds owed. Pell funds may only be used for educational purposes; recipients must sign a statement certifying that a grant will be used only for expenses of attending school.

A student's Pell Grant amount is determined from the eligibility index, together with information on the cost of attendance at the student's chosen school and enrollment status (full- or part-time). The size of the grant increases as the eligibility index decreases, so that an applicant with an eligibility index of zero may receive the maximum award. However, the maximum award is limited to one-half of the educational costs, not to exceed $1,800 in award year 1982-83.

Pell Grant recipients may also receive financial aid from other Title IV programs.

The program year cycle and application processing

Schools may disburse funds to students in an award year from July 1 through June 30. Students may apply beginning in January preceeding the July 1 date, and may apply at any time through March 15 of the next year. Students may apply using a federal application form, or may use application forms provided by other aid administration services such as the College Scholarship Service or the American College Testing Program, or through several state higher education assistance agencies including Pennsylvania and California.

Initial processing of applications each year is done by a contractor, where applications are received, edited, reviewed for internal consistency, and the eligibility is calculated. The contract for this processing has been held in recent years by System Development Corporation, of Santa Monica, CA; beginning in 1984, the contract will be with Westinghouse Information Services of Iowa City, IA.

Applications not sent directly by students to the processor, are sent on computer tape from the other agencies.

Validation

For some years the processor developed computer editing routines to scan the data sent by students on applications. Discrepancies between items, or miscalculations, could trigger a special message to an

applicant requesting verification of information. A small number of applicants were also asked to bring supporting documents to the aid official at their school for further review. This type of validation has been greatly expanded in recent years, as described in the body of the Design Paper.

Students are selected for validation by the processor, following rules set by ED. The selection process has been different in 1982-83 and 1983-84, with a wide variety of decision rules.

All students selected for validation must provide documentary proof of the data elements in their applications on: income, federal income tax paid, and sometimes other items including student's or parent's Social Security benefits, independent student status, household size, or assets. Documentary proof of public benefits such as welfare, Veteran's Administration, or Social Security is required.

Campus financial aid officials must review the student's evidence. ED provides a lengthy manual as guidance, and sponsors training through the National Association of Student Financial Aid Administrators. If the aid official decides that the information on the application is correct, the student can receive a Pell award. If the original computation of eligibility needs to be corrected, in light of the additional evidence, the student must resubmit a corrected form to the processor and await recalculation. ED does allow schools the option of issuing half the potential Pell award to a student while corrections are pending, but if the award turns out to be wrong, the school is liable for any funds awarded in error.

No other student aid program in Title IV requires documentary proof of application data, though ED has recently proposed this, and many schools have their own validation requirements for all students or students applying for certain aid.

Chronology of policy events concerning validation

Year	Month	Event
1981	(fall)	Draft of contractor report on Pell award errors in 1980-81
		ED proposes increasing validation from 300,000 to 700,000, and amending FY 82 continuing resolution to fund validation activity
1982	January	ED requests authority from Congress to reprogram FY 82 Pell Grant funds to support validation costs; all applicants to submit 1040 tax form to schools
	February	House testimony on validation by aid administrators
	March	ED requests supplemental appropriation to cover added processor costs of recomputing eligibility of those validated; processor begins requiring all Pell applicants to bring documents to schools; all ED funding requests rejected

Year	Month	Event
1982	June	ED directs processor to stop 100 percent validation selection; narrower criteria to be used to select next 300,000
	September	ED contractor begins replication of earlier quality control study of accuracy of Pell Grant awards
1983	March	ED directs processor to use new validation selection criteria for 1983-84 award year processing, including cross-year and error-prone modeling
	August	ED proposes expansion of validation to campus-based aid programs in Title IV, and to Guaranteed Student Loan applicants
	September	ED awards new processor contract, to begin 1984
	October	ED solicits proposals for quality control study of campus-based aid programs and GSL

II. DESIGN PAPER FOR AN EVALUATION OF THE DEPARTMENT OF DEFENSE'S IMPLEMENTATION OF TECHNICAL RISK ASSESSMENT IN THE AREA OF WEAPON SYSTEMS ACQUISITION

Luis Gonzalez

Marcia Gilbert

Joan McLaughlin

July 1984

Table of Contents

I. INTRODUCTION

There are many contributing factors to cost growth in weapon systems acquisition, of which technical problems are a primary contributor. In recent years the Defense Department has recognized the need to address the technical risk issue, as exemplified by Deputy Secretary Carlucci's initiatives. Of particular interest is Initiative 11 which recommends the quantification of technical risk.

CONTRIBUTING FACTORS TO COST GROWTH

A study of six major programs conducted at the University of Southern California,[1] identified twenty six factors which specifically contributed to cost growth. Some of the factors include: over-optimism in design, design changes, inflation, changes in political and customer influences, low bidding, and improper budget and cost control.

Technological uncertainty, however, has been identified as a primary cause of overruns by other studies.[2] After 20 years of examining complex development programs, the Rand Corporation concluded that cost growth appears to arise primarily from efforts to subdue difficult technologies on highly compressed schedules and under optimistic cost assumptions.

It is important to recognize that mastering difficult technology is an integral and necessary part of the weapons acquisition process. This is so because much vital military research and development (R&D) depends on increments of performance improvement that are difficult to achieve. Such programs, consequently, almost always contain critical elements of risk and uncertainty. It is equally important to realize, however, that difficult technology cannot always be subdued fast enough to ensure program "success"--cancellation is a possible outcome of R&D projects.

Risk and its relationship to cost growth have been recognized within the Department of Defense (DoD) for some time. As early as 1969 in two memoranda, the Deputy Secretary of Defense directed the service secretaries to: identify areas of high technical risk; accomplish formal risk analysis; and include explicit consideration of risk assessment, reduction and avoidance in the management of weapon systems acquisition.[3]

The Office of Management and Budget (OMB) is also aware of the importance of risk assessment. An OMB circular dated April 5, 1976, requires the consideration of methods of analyzing and evaluating contractor and government risks as part of the acquisition strategy for major systems.[4]

GAO has also stated the need to address the risk issue in major acquisitions. The following examples are representative.

Testifying in front of the House Committee on Government Operations in November, 1979, a GAO witness said that the Committee would have a much better idea of the possibilities of the cost of weapon systems if DoD were "forced to discuss the risks that are involved" in the programs.[5] In February of this year, a GAO report stated that a major shortcoming of the Army's evaluation of on-going programs has been the lack of comprehensive risk assessments.[6]

CARLUCCI INITIATIVE 11

On April 30, 1981, former Deputy Secretary of Defense Carlucci issued a memorandum[7] on the defense acquisition system, aimed at reducing costs, making the process more efficient, increasing the stability of programs, and decreasing the acquisition time of military hardware. The memo outlined 32 recommendations or "initiatives." One of these, number 11, called for DoD to "incorporate the use of budgeted funds for technological risk."

In this Initiative 11, Deputy Secretary Carlucci states that program managers who had explicitly been requesting funds in contemplation of uncertainties were having these funds deleted in the DoD budget process, by OMB, or by Congress. Thus when uncertainties occurred there were delays in the program or undesirable funding adjustments. To prevent these problems, the Deputy Secretary recommended that DoD increase its efforts to "quantify risk" and to budget "funds to deal with uncertainty." The action required for this recommendation was for the Secretary to emphasize the need to "evaluate, quantify, and plan for risk." In addition, the services were required to adopt a concept the Army had developed to budget for risk--Total Risk Assessing Cost Estimate (TRACE), to be described below--or propose an alternative within 60 days.

Reporting on the status of the initiatives in 1983, Deputy Secretary Thayer stated that

> (t)he specific action to develop procedures to budget for technological risk has been implemented by the Services, and this initiative is now considered completed.[8]

DEFINITION OF RISK

There appears to be no standard definition of the term "risk" that is generally used in DoD or service regulations or documents. In some instances the term is used to refer to the acquisition cycle itself, for example, "development risk" for risk that occurs as a system is being developed, or "production risk" for risk that occurs in the later phase of acquisition. In other instances risk refers to an aspect of the cycle, such as "cost," "schedule," or "performance" risk. Other terms, such as "critical issues," are used as a substitute for "risk" as well.

Even within a single regulation or document the use of the term "risk" is ambiguous. Within Deputy Secretary Carlucci's Memo, for example, Initiative 11 is entitled "Incorporate the Use of Budgeted Funds for Technological Risk." Later, it states that the services should "evaluate, quantify, and plan for risk." It should be assumed that since the title addresses technological risk that this is the kind of risk being referred to, not schedule or cost risks. This was the interpretation made by Deputy Secretary Thayer two years later when he said that the services have implemented procedures to budget for technological risk. Yet in the model the Deputy Secretary recommended for the services' use (i.e., TRACE), technical risk does not necessarily have to be included. Schedule risk, for example, can be included instead.

The definition of risk that will be used for our proposed effort is taken from a recently published defense handbook on risk assessment.[9] Risk in the acquisition of major weapons systems is:

> the probability and consequence of not achieving some defined program goal - such as cost, schedule or technical performance.

In terms of risk assessment techniques, risk is usually broken down in terms of cost, schedule, and/or technical performance. There are techniques which deal exclusively with each of these types of risks and those which allow more than one type of risk to be identified within the same model. We have defined each of these three types of risk, based on DoD documents. For the purpose of this effort, cost, schedule, and technical risk will be defined as follows:

Cost risk: the probability of not achieving program goals within the amount budgeted.

Schedule risk: the probability of not achieving program goals within the time allotted.

Technical risk: the probability of not achieving program goals due to failure of the technology to meet the necessary performance requirements.

These risks are not mutually exclusive. As discussed previously, technical problems appear to have been a major component in the cost overruns experienced in weapon system acquisition: therefore, technical risk is related to cost risk and, in the same way, to schedule risk as well.

RISK ASSESSMENT TECHNIQUES

There are a number of techniques available for risk assessment. The techniques can be quantitative or non-quantitative, depending on whether statistical probabilities are assigned to each risk element identified. The techniques may be "formal,"

involving an explicit breakdown of program elements, or "informal," consisting of a more intuitive assessment of risk.

There is some subjectivity in all risk assessment techniques. In formal and informal, quantitative or non-quantitative assessments, the techniques involve asking experts for their subjective judgments of what the risk elements are as well as the probability of their occurrence. What distinguishes the different kinds of techniques is the information that goes into the subjective judgments--test results, expertise of individuals making the judgments, how the information is obtained (through the use of a Delphi method, for example)--and the kinds of information requested--judgments of high, medium, or low risk as opposed to judgments of statistical probabilities.

As discussed earlier, a variety of techniques may be used to assess cost, schedule, and/or technical risk. The emphasis here is on the assessment of technical risk. Two of the most frequently used techniques for technical risk are the network and risk factor methods.

Briefly, the network technique involves modeling the acquisition process for a weapon system as a network. In such a network the nodes or endpoints represent a milestone point in the program and the links connecting the nodes represent the activities that must be carried out to achieve the endpoint. The probability of successfully carrying out an activity is usually added to the model. Numerous computer simulations are then performed to evaluate the probability of achieving the goal represented by the network. Examples of network models are the Venture Evaluation and Review Techniques (VERT) and Risk Information System and Network Evaluation Techniques (RISNET).

The risk factor method was developed for use in budgeting for technical risk. In this technique, all elements of a weapon system and their associated costs are identified in a Baseline Cost Estimate (BCE). A "risk factor" is then determined for each element associated with risk in the weapon system. This factor is a number by which the BCE should be increased to account for a technical problem if it arises. The BCE and risk factor are determined by individuals with expertise with the technology involved in the weapon system. The risk factor method is most widely used in developing TRACE estimates in the Army.

Risk assessment techniques such as the network and risk factor methods can be used in different aspects of the acquisition process. In the program office it can be used for budgeting, as in the TRACE programs' attempts to budget for risk, and in day-to-day program management, as when decisions about program alternatives have to be made. The assessments could also conceivably be used in decisions made at levels above the program office, for both budgeting and for making realistic decisions about the technology involved in the weapon system. Assessments

of risk could also help determine if program milestones are scheduled appropriately.

Once the decision to do a risk assessment is made, the decision regarding which technique to use appears to be based upon the training of the staff. Other considerations may also be involved, such as time and funding available for the assessment, as well as the availability of a computer.

TRACE

In order to deal with costs associated with program uncertainties, the Army developed the Total Risk Assessing Cost Estimate (TRACE) program. Essentially, TRACE involves adding an incremental dollar figure to the BCE of the program to account for uncertain events. The amount added to the BCE for risk is calculated by combining all the uncertain events for the program and identifying the funds that would be required to account for risk at the .05 probability level. The way in which this amount is calculated differs, depending on the chosen technique, but the result is always the estimate having a 50/50 chance of an occurrence of the uncertain events. TRACE thus represents a compromise between funding for only those aspects of the program that can be identified and costed with certainty, and funding for all possible risks.

Three techniques of risk assessment are typically recommended in the TRACE calculations: network analysis, risk factor analysis, and probabilistic event analysis. Network and risk factor analyses have been described above. Probabilistic event analysis involves breaking the program down into elements, assessing the probability of a problem occurring for each element, calculating the cost of the possible problems, and identifying the probability of a problem's impact on other elements.

As previously mentioned, in Deputy Secretary Carlucci's 1981 initiative on budgeting for technological risk, the recommendation was made for the services to adopt the TRACE concept. Each service's response to this recommendation will be addresssed below in the section on the results of scoping.

Originally, TRACE funds were calculated for the Research, Development, Test and Evaluation (RDT&E) phase of the acquisition cycle. This was because much of the risk associated with weapon system development is identified in this early phase. The Army is now extending TRACE to the production phase as well.

Congressional approval was obtained for TRACE in RDT&E when this program was initiated. The 1983 Memorandum from Deputy Secretary Thayer, however, recommended efforts to improve congressional acceptance of TRACE-RDT&E. In addition, the approval of Congress is still being sought for the more recent extension of TRACE to production (TRACE-P).

RISK AND CONTRACT TYPE

Deputy Secretary Carlucci succinctly stated DoD policy on the issue of contract type and risk in a memorandum to the service secretaries on June 18, 1982.[10] He wrote:

> The principal distinction between various contract types lies in the degree of risk assumed by the parties and in the apportionment of responsibility. To the extent that the selected contract type reflects a fair and reasonable apportionment of risk and responsibility between the government and the contractor, the contract is more likely to facilitate the efficient conduct of a program.

It should be noted that in the context of this memo risk is used in its broadest sense--including overall business/financial risks, as well as cost, schedule and technical risks, as previously defined. It is also important to realize that when technical risk is present, the magnitude of the overall risk will be directly dependent on that technical risk; consequently, the relationship between risk and contract type, as described below, holds when addressing technical risk specifically.

Contract types

Basically, there are two type of contracts: fixed price and cost reimbursement. The major distinction between the two is in the nature of the contractor's obligation and risk. Under a fixed price contract, the contractor must produce the required items or perform the specific service for the fixed price (or within the ceiling price of an incentive contract) or be subject to the penalties provided for in a default clause. There are various types of fixed price contracts--Firm Fixed Price (FFP), Fixed Price with Redetermination (FPR), Fixed Price Incentive Fee (FPIF), and Fixed Price Incentive-Successive targets (FPIS), to name a few.

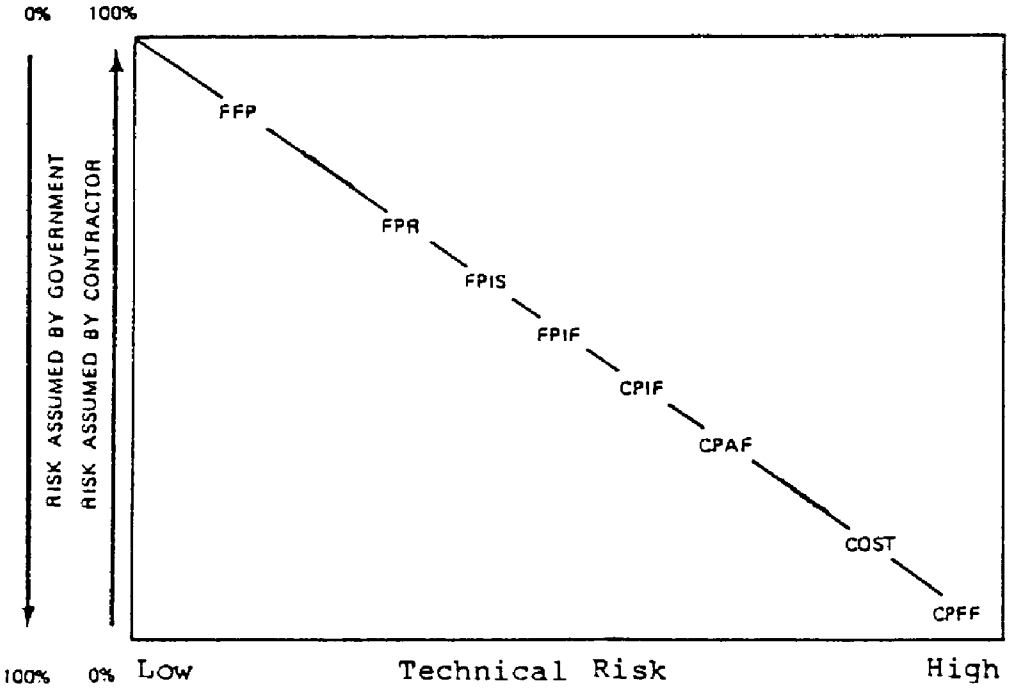
Under a cost reimbursement contract, the product is not paid for on the basis of an invoice price; rather, the Government pays the contractor's costs for material and labor and a portion of his overhead cost in accordance with appropriate clauses in the contract. The principal cost-type contracts include Cost, Cost Plus Fixed Fee (CPFF), Cost Plus Incentive Fee (CPIF), and Cost Plus Award Fee (CPAF).

Relationship between risk and contract type

During the earliest phase of the acquisition process when the actual end-product has not been specifically defined and technical risks are higher, a cost-reimbursement contract such as Cost Plus Fixed Fee (CPFF) appears to be most suitable. Then, as the

product becomes better defined and the technological risks have been reduced, the contract type can be shifted to Cost Plus Award Fee (CPAF) or Cost Plus Incentive Fee (CPIF), and ultimately to a fixed price contract during early production. The relative risk assumed by the Government and the contractor--as well as the degree of technical risk--as a function of the type of contract is shown in Figure 1.

Figure 1
Degree of Risk as a Function
of Contract Type



II. THE DOD ACQUISITION PROCESS

The DoD acquisition process is complex, and at least some familiarity with stages of development, major decision points, and documentation required is necessary to understand the issues involved in technical risk assessments. A brief description of the process is therefore given here, following an outline of the policy directives which guide the process and address risk assessment.

POLICY DIRECTIVES

There are several directives concerning major systems acquisition. Three of these issued by OMB and DoD are particularly important in defining the components and outlining the process of acquisition. The consideration given to risk in each of these directives will be briefly described in this section. First, however, the way in which a system is designated as a major system will be outlined.

According to DoD Directive 5000.1, the Secretary of Defense designates the systems that are to be managed as "major systems." This decision may be based upon:

1) "development risk," "urgency of need" or other interests of the Secretary of Defense;

2) joint acquisition by two or more of the services or by the U.S. and another nation;

3) cost estimates that exceed $200 million (FY80 dollars) in Research, Development, Test & Evaluation or $1 billion (FY80 dollars) in procurement; and/or

4) "significant congressional interest."

OMB issued a circular (A-109) in 1976 which outlined the policies to be followed by each executive branch agency acquiring major systems. The circular is not specific with regard to technical risk, stating only that the acquisition strategy for a program "could typically include . . . methods for analyzing and evaluating contractor and Government risks."

In DoD Directive 5000.1, the ony mention of risk concerns the first criterion given above for designating a system as "major" (that is, designation may be made if "development risk" is involved in a weapon system). DoD Directive 5000.2 makes a more explicit reference to technical or technological risk. In the required documentation for the weapon system review process--to be discussed in more detail below--this type of risk must be "identified" or "addressed."

Each service has its own regulations for the acquisition process. "Risk" is mentioned in several of these regulations, although it is often not clear whether or not the reference is to technical risk. In addition, while some regulations call specifically for "assessment," others merely state that "technical risk" or "risk" should be considered or addressed.

THE PROCESS

While it is difficult to describe all the steps in DoD's major systems acquisition process, a reasonable understanding may be achieved by gaining familiarity with the major milestones and the resulting acquisition phases. It should be realized that for each milestone there are several levels of review within each service culminating with the Defense Systems Acquisition Review Council (DSARC). The Council provides advisory support to the Secretary, who is the decision authority. DSARC membership includes:

Chairman: Defense Acquisition Executive - usually USDR&E (see below)

Members:
- Under Secretary of Defense, Research and Engineering (USDR&E)
- Under Secretary of Defense, Policy
- Assistant Secretary of Defense, Manpower, Reserve Affairs and Logistics
- Assistant Secretary of Defense, Comptroller
- Director, Program Analysis and Evaluation
- Chairman, Joint Chiefs of Staff

Major milestones and acquisition phases

For each major acquisition there are four decision points: program initiation and three DSARC milestones. Following each decision point there is a distinct acquisition phase. A description of each of the decision points and phases follows. The flow of the process is presented in figure 2.

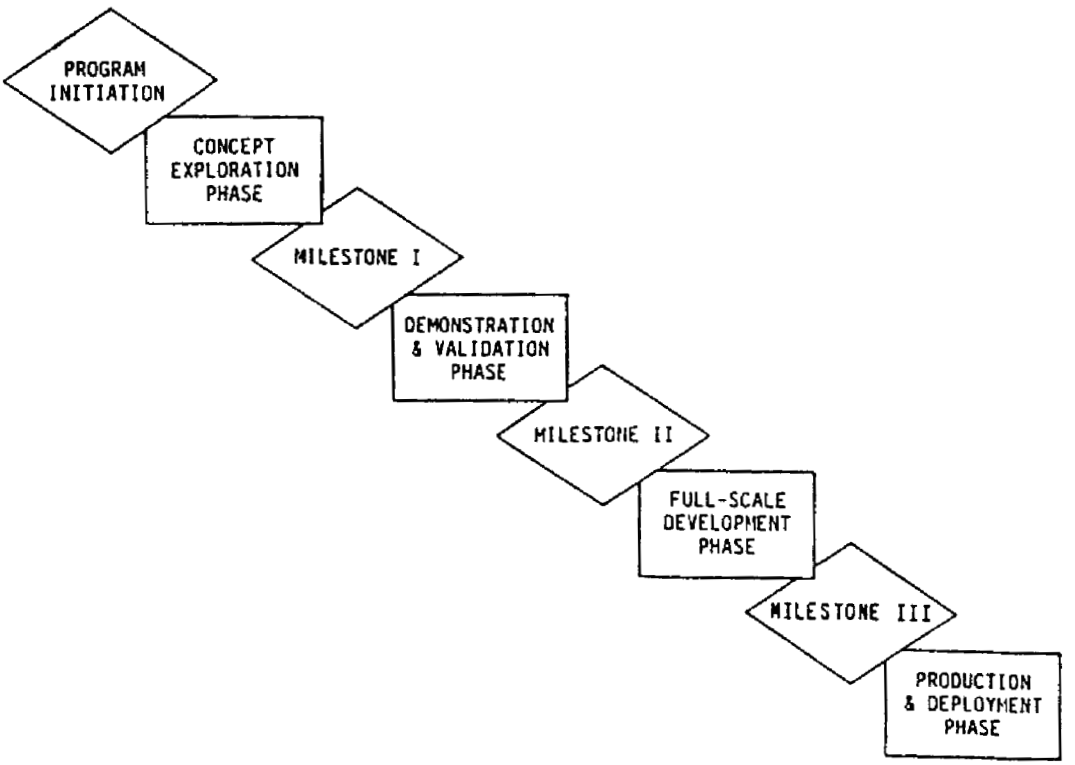
Program initiation decision

The need determination, in the Planning, Programming and Budgeting System process, provides the justification for a new system start. This is normally submitted during the Program Objectives Memorandum (POM) preparation, review and approval process in which funds for the budget year of the POM are requested. The Secretary provides appropriate program guidance after such review. This action provides official sanction for a new program start and authorizes, when funds are available, the initiation of the acquisition.

Concept exploration phase

The initial phase of the acquisition process identifies and explores alternatives and acquires the information necessary to select the best alternatives for system concepts and hardware/computer software development. The technical specifications and economic basis for proposed systems are established in this phase. Critical technical, training, logistic, operational, cost and manpower issues are identified for resolution in subsequent phases in order to minimize future problems. Investigations must also analyze support and readiness criteria of current systems, establish targets for new system(s), develop alternative operational and support concepts, and evaluate manpower and logistic support resource implications of each alternative. In addition, in preparation for Milestone I, a statement must be made of the objectives, responsibilities, resources, and schedule for all test and evaluation efforts.

Figure 2
Acquisition Process



Milestone I

The first major milestone decision is concept selection for entry into the Demonstration and Validation phase. This decision is a validation of the requirement for the program, based upon preliminary evaluation of concepts, costs, schedule, readiness objectives and affordability. It provides authority to proceed with the Demonstration and Validation phase and to develop the system sufficiently to support the next milestone decision. The Milestone I decision establishes thresholds and objectives to be met and reviewed at the next milestone, the acquisition strategy for the recommended concepts--including the nature and timing of the next decision point--and a dollar threshold that cannot be exceeded to carry the program through the next milestone.

Demonstration and validation phase

This second phase consists of steps necessary to verify preliminary design and engineering, accomplish necessary planning, analyze trade-off proposals, resolve or minimize logistic problems identified during the Concept Exploration phase, prepare a formal requirement document, and validate the concept for Full-scale Development. Normally, two or more competitors are used. Prototypes should be used to demonstrate feasibility of the system, subsystem, or components and system-peculiar test measuring and diagnostic equipment and support equipment. An update of the test and evaluation plans must be made in this phase as well.

Milestone II

The second major decision is program go-ahead and approval to proceed with Full-scale Development. The timing of the Milestone II decision is flexible and depends upon the tailored acquisition strategy approved by the Secretary at Milestone I.

Full-scale development phase

In this third phase, the system--including all items necessary for its support, to include training devices and computer resources--is fully developed, engineered, fabricated, and tested. A decision is then rendered on its acceptability for entering the service's inventory. Concurrently, nonmateriel aspects required to field an integrated system are developed, refined and finalized.

Milestone III

The third major decision point is for Production and subsequent deployment. Normally the Milestone III decision for a DoD major program is delegated by the Secretary to the service secretary unless the thresholds established at Milestone II are breached, or there is major public or congressional concern.

Production and deployment phase

During this last phase, operational units are trained, equipment is procured and distributed, and logistic support provided. Product improvements that have been preplanned are applied to the equipment as required.

SUPPORTING DOCUMENTS

The need for a discussion of technical risks in program documentation is recognized from the very outset of the acquisition cycle. The Justification for Major System New Start, the first document required for system approval, must contain a discussion of the maturity of the technology planned for the system, with "particular emphasis on remaining areas of risk."

The documentation required for milestone review also must address technical risks. The System Concept Paper (SCP) required for Milestone I must identify "key areas of technological risk which must be reduced by R&D and validated by T&E [Test and Evaluation] before Milestone II." The Decision Coordinating Paper (DCP), to be prepared for Milestones II and III, must contain a discussion of the continuing technological risks of the selected alternative. The DCP for Milestone II must "discuss the T&E results that show all significant risk areas have been resolved," and that the technology requires only engineering (not experimental) efforts.

Along with the SCP or DCP submitted for each milestone review, a Test and Evaluation Master Plan (TEMP) must also be submitted. The TEMP must contain a description of key areas of technological risk that must be addressed by testing. This complements the requirements for testing called for in discussions of technical risks in the SCP and DCP.

These documents are an integral part of the review process for every major weapon system. The requirement that they all address technical risks demonstrates DoD's concern that such information be considered in making decisions about the project at levels of command above the program office.

Each of the three services also requires that an acquisition plan be developed for any major weapon system. In the Army, this Acquisition Plan is part of the overall acquisition strategy developed by the program office in conjunction with the Development and Readiness Command. The equivalent information in the Navy becomes part of the Acquisition Strategy Paper which is prepared exclusively by the program office and routed through command channels for review. The Air Force convenes a Business Strategy Panel to review the acquisition plan which the program office has prepared. This panel has an "advisory" function, but its recommendations are influential in determining the final form of the

plan. In all instances, the acquisition plan is incorporated in the documentation prepared by each service for every milestone review.

III. RESULTS OF THE SCOPING PHASE

The scoping phase of this job included interviews with DoD personnel at the Office of the Secretary of Defense (OSD) and service headquarters levels, the collection of data from six weapon system program offices (two from each service), a preliminary review of relevant literature, informal discussions with representatives of two major defense contractors, and an examination of the DD-350 database. The DD-350 is a form containing DoD contract information which we felt might be useful for the project.

OVERVIEW OF SCOPING

The interviews at OSD and headquarters levels provided an understanding of DoD policies and of each service's response to Deputy Secretary Carlucci's Initiative 11. We then requested that each service provide us with two programs to examine in greater depth, to determine how Initiative 11 had been implemented, what sorts of technical risk assessment techniques were being used, and how, or if, this was related to a decision on contract type.

We asked that each service provide one program which was close to a Milestone I review and one program which was close to a Milestone III review. We thought the former would allow an examination of recent programs, begun under some of the new policies and at a stage in development associated with many unresolved technical problems. We felt the latter would allow an examination of programs with a more complete file of information from the review process, and at a stage where most of the technical problems are supposed to have been resolved. We examined programs at these two points in the Army and the Air Force. The Navy programs examined were at Milestones II and III as no appropriate Milestone I program was available.

The interviews conducted in the program office and the review of the program documention were aimed at discovering the kinds of risk assessments that are being performed and the use of these assessments. More specifically, we were interested in the techniques used for risk assessment, the documentation of the results, and the way the technical risk assessments fed into the decision-making process. We were particularly interested in whether or not the assessment of technical risk was used in decisions about contract type, and, if so, how the information was considered.

Based on the review of the six programs, it appears that little is being done in the area of formal assessments of technical risks. None of the six programs had performed a quantitative technical risk assessment. Further, the level of technical risk did not enter into the decisions on contract types.

Our preliminary review of relevant literature focused primarily on techniques and applications of technical risk assessments for weapons systems, but also covered such topics as

behavior under uncertainty, business strategies, and contract types. The review showed that virtually all of the relevant work has been performed either directly by the military or under the aegis of DoD. The services have thus spent time developing the tools of technical risk assessments, but the preliminary indications are that they are not being widely used.

Interviews with defense contractors indicated that they play a limited role in performing technical risk assessments, but they remain the primary source of technical information on a system.

ARMY

It was the Army that developed TRACE, the model recommended in Initiative 11. A brief summary of the Army's implementation of TRACE is given below. Both of the programs examined during scoping are TRACE programs. The Multiple Launch Rocket System/ Terminal Guidance Warhead (MLRS/TGW) has TRACE for RDT&E; the M1E1 Tank has TRACE for production (TRACE-P). These programs also are discussed below.

TRACE program

The TRACE program was initiated by the Assistant Secretary of the Army for Research and Development, Norman Augustine, in the early 1970's. His purpose was to offset the effects of cost growth in the R&D phase of weapon acquisition. More recently, however, the Army has begun to implement TRACE for production as well. This extension of the program suggests that the Army believes that TRACE has been successful in RDT&E, and thus may be of use in the Production phase of the acquisition cycle. This "success" may be measured in terms of TRACE's ability to control cost growth, which was the original intention of Mr. Augustine. It may also be that the use of TRACE has proved beneficial in other ways. For example, in calculating TRACE, there could have been an identification of a risk area that was not formerly considered. This area could then be attended to in order to avoid any serious problem. Individuals involved in the Army programs examined during scoping made reference to this unintended benefit of TRACE.

As described above, the TRACE calculation allows risk funds to be budgeted in addition to the baseline cost of program elements. In the Army, only the dollars for baseline costs are given to the program office once funding has been approved. The amount that is budgeted for risk is held at the headquarters level, in the Office of the Deputy Chief of Staff for Research, Development and Acquisition. If a program manager has a need for the risk funds, headquarters' approval must be obtained before the funds are released.

According to TRACE guidelines, risks that may be included in TRACE-RDT&E calculations are: (1) technical design changes; (2) rescheduling because of technical and budgetary problems;

(3) additional testing of design corrections; (4) additional hardware to support design corrections; (5) schedule slippages due to late delivery of components or materials; (6) non-negligent human error; (7) program termination. The following risks are not allowed to be considered in TRACE-RDT&E calculations: (1) cost for modifications that result from changes in the statement of requirements; (2) effects of inflation; and (3) additional costs due to pay increases.

As in TRACE-RDT&E, more than technical risk is intended to be included in TRACE-P. The allowed risk areas are: (1) threat uncertainty; (2) management; (3) materials/purchased parts; (4) facilities/equipment; (5) labor; (6) design changes; (7) producibility; and (8) performance. The list of risk areas not allowed in TRACE-P is too lengthy to be outlined here. It includes such areas as quantity changes and inadequate funding in early years.

For TRACE-RDT&E, unused risk funds can be carried over from the first to the second year of the program. After the second year the unused funds can be reprogrammed to other weapon systems. For TRACE-P, unused risk funds cannot be carried over to the next year. They may be reprogrammed after only one year. In both TRACE programs the unused risk funds can be used for any program that is having budget problems. They do not have to be used as risk funds by these other programs.

Currently, 12 Army programs have been awarded TRACE-RDT&E funds. Since the initiation of TRACE in the late 1970's about 90 percent of the designated programs have used their total allotment of risk funds for TRACE-RDT&E each year. Only one or two programs have received TRACE-P funds for 1984. It remains to be determined whether or not these programs will have used their TRACE-funds by the end of this fiscal year.

MLRS/TGW

The Multiple Launch Rocket System (MLRS) is a free-flight, area fire, artillery rocket system. Its purpose is to provide a large volume of firepower in a short time against time-sensitive targets. The Terminal Guidance Warhead (TGW) for the MLRS was particularly developed to defeat armor.

The Army began developing the MLRS in 1977. It was pursued as an international program, as directed by the Secretary of Defense. The TGW was included as an option to the MLRS at that time, as required by the House Armed Services Committee. The prime contractor for the MLRS is the Vought Corporation.

The nations involved in the MLRS/TGW program are France, Germany, the United Kingdom, and the United States. The multinational group working on the TGW has completed the concept definition phase of the acquisition cycle and has agreed on the best technical approach. At this time, the program is scheduled for a

Milestone I review by the DSARC so that it may proceed into the next phases.

According to an assistant program manager in the MLRS/TGW office one way they are dealing with technical risk is by dividing the Demonstration and Validation phase into two subphases: Component Demonstration and System Demonstration. A great deal of technical risk occurs when hardware is used before it is fully developed. In the first subphase of the MLRS/TGW program the hardware is forced to be developed before the go-ahead is given to enter the second subphase.

Technical risk assessment

The MLRS/TGW program office does a formal quantitative assessment of schedule risk, not technical risk. Two months were spent doing a network analysis, Program Evaluation and Review Technique (PERT). The Assistant Program Manager said that the risk assessment was useful because (1) it got the program manager thinking about the risks and (2) it justified decisions made in the program office to the Pentagon.

The MLRS/TGW has also been identified as a TRACE program for RDT&E. In calculating their risk funds for TRACE, individuals in the program office make use of the PERT model. They assign dollar figures to the schedule risks that have been identified.

Contract decision

The contracts for the MLRS/TGW will be Cost Plus Incentive Fee for development and Fixed Price for production. Since the risk assessment was not focused on technical risk, a quantitative assessment of technical risk could not have been used in determining contract type. The individuals interviewed in the program office believe that there is a relationship between the two, but they implied the relationship was of a general sort, and not one in which a specific assessment fit into the decision.

M1E1 Abrams tank

The M1 Abrams tank is the Army's main ground combat weapon system in its fifth year of production. The M1E1 presents a number of improvements over its predecessor, including a 120 mm gun, improved armor protection, a protective system with microclimate control against nuclear, biological, and chemical warfare, a suspension system upgrade, and weight reduction. The M1E1 program is currently preparing for its Milestone III reviews scheduled for this summer. The goal for production is 60 tanks per month. The program plan is to gradually increase the number of M1E1 tanks while simultaneously decreasing the number of M1 tanks, until only the M1E1 is produced in 1985. The prime contractor for the M1 and M1E1 is General Dynamics.

Technical risk assessment

No quantitative technical risk assessment was performed for the program manager's use. The program manager said that a "subjective," but not necessarily "unstructured," review of risk had been done. The contractors are required to address areas of risk. In doing so, they make subjective judgments but their judgments are backed by more objective data, e.g., cost figures, logistic implications, reliability data. In addition, the contractors present the risks to experts from the program office and the laboratories, who may argue with the contractors and/or call for more tests.

The program manager stated that a more formal quantitative assessment of risk might be more helpful for other programs, but the more subjective assessment was sufficient for the M1E1. Because it was just an improvement to the M1, it already had many of the technical risks worked out.

The M1E1 is a TRACE-P program. In the calculation of the TRACE figures they used a Venture Evaluation and Review Technique (VERT) assessment, the technique used in all TRACE-P calculations by the Army. In the assessment, all major subcomponents of the system are listed. For each subcomponent, categories of risk are then identified for the first three years of production, although TRACE-P funds are approved for use on a yearly basis. The risk categories, enumerated above, include such items as management, threat uncertainty, and materials. Dollar figures are associated with the risks once they are identified.

For the M1E1, some of the elements in the TRACE-P calculation are clearly related to technical risk. For example, one aspect of TRACE-P covers the risk that an auxilliary power unit would be required to obtain desired performance of the microclimate cooling system. Other elements do not address technical risk. For example, "management" of the M1E1 armament is identified as a risk element, which is clearly not a technical risk.

Contract decisions

The M1E1 will go into production with a Firm Fixed Price contract. When asked about the relationship between technical risk assessment and contract decisions, the program manager said that he thought a general assessment of risk was made when a contract type was decided on. He did not believe a more specific relationship could be found, in which a formal assessment of risk was used by those deciding on contracts.

An individual involved in contracting for the Abrams tank said that there are many factors that go into the determination of contract type, and risk was only one of them. Whether a formal or more subjective assessment is done does not really make a difference. He said that even when a more formal assessment is done, as

in the case of the M1, it is typically completed too late to influence the decision anyway.

NAVY

In compliance with Deputy Secretary Carlucci's Initiative 11, the Navy established a pilot program to evaluate the TRACE concept. One of the programs we examined, and describe below, is part of this pilot effort. The other program we examined is taking part in a different pilot effort to implement the recommendations of the Defense Science Board's report on "Solving the Risk Equation in Transitioning from Development to Production."

Pilot TRACE Effort

The Navy's response to Initiative 11 was to set up a pilot program within the Naval Air Systems Command (NAVAIR) to evaluate the use of the TRACE concept developed by the Army. The Navy's concept differs from the TRACE concept used by the Army. The Navy plans to hold the deferral funds at the system command level. Thus, control of the Navy TRACE funds occurs two levels lower in the hierarchy than in the Army, which holds the TRACE funds at the secretariat level.

In the opinion of the NAVAIR TRACE coordinator, existing methods for calculating risk funds are so complicated and time intensive that when affordable, are done by outside experts. Consequently, the outsiders become the risk experts and the program managers gain little knowledge--the transfer of knowledge is away from the decisionmakers.

In fiscal year 1984, three of the four programs with TRACE deferrals used the funds. For fiscal year 1985, four programs are expected to include TRACE funds. One of the two programs we examined, and discuss below, is one of the programs with TRACE funds in both fiscal years, and one of those that absorbed 1984 TRACE funds.

Harrier II

The Harrier II is a vertical/short take-off and landing light attack aircraft, designated the AV-8B, to be used by the Marine Corps and the British Air Force. It can carry general purpose bombs, cluster munitions, laser guided weapons, Maverick and Sidewinder missiles, and a 25mm--fuselage mounted--cannon. For the Marine aircraft, McDonnell Douglas is the prime contractor with British Aerospace as principal subcontractor. For the British Air Force aircraft the two companies reverse roles. In both cases, Rolls Royce, in concert with Pratt & Whitney, will provide the Pegasus Engine.

Although the program is at Milestone III, due to technical problems a decision to go ahead with full production has not been

made. Instead, as it is frequently done with programs with technical difficulties in the transition to production stage, the final decision has been delayed by adding additional intermediate decisions. In this case a Milestone III-A decision, for limited production, was made in August 1983, and a Milestone III-B was scheduled for March 1985. Additionally, an Office of the Secretary of Defense review was held in January 1984.

Technical risk assessment

The deputy program manager stated, "off-the-record," that no formal risk assessment had been made. The TRACE fund was calculated by a "back-of-an-envelope judgment call." He said that the program did not have the time or resources to conduct a formal assessment. He also felt that technical risk assessments are more appropriate around Milestone I.

As to the release of TRACE funds ($7 million) to the program in 1984, it was done on the basis of "unexpected technical problems and schedule difficulties." In fact, the schedule difficulties were those caused by a Congressional decision to reduce their budget by $13 million.

Contract decision

Given that no risk assessments were conducted, the contract type decision could not be related to technical risk assessed. The available evidence indicated that the Secretary of the Navy imposed the contract type at Milestone III-A. At the time, the program office wanted to go with a Fixed Price Incentive contract for the 21 aircraft to be produced in 1985. The Secretary, however, approved this limited production with the clear understanding that a Firm Fixed Price contract would be obtained. But in the opinion of the program office, based on their knowledge of the outstanding technical difficulties, such a contract transfers an unfair share of risk to the contractor.

ALWT

The Advanced Light Weight Torpedo (ALWT) system, designated the MK-50, was a considerable technological push in every component, including warhead, command and control, and propulsion. It is meant to replace the MK-46, the Navy's standard surface torpedo, which is carried by every type of antisubmarine warfare aircraft.

The Milestone II decision, for Full-scale Development and the production of four prototypes by Honeywell, slipped from April, 1983, through two additional dates to January, 1984. A limited production decision, Milestone III-A, is scheduled for the end of 1986, with technical and operational evaluations to take place in 1987 and 1988, respectively. A decision for full production go-ahead, Milestone III-B, is not expected until the end of 1988.

Technical risk assessment

Risk assessment for the ALWT system was totally informal, based on the experience of the technical staff and prior involvement in similar systems. Program officials did not use any quantification of risk; the Deputy Program Manager stated that they do not trust such techniques. What was done was to set up a PERT-like system which included very realistic time estimates.

The program officers were not familiar with TRACE funds and, once they understood what such funds involved, felt added funds would not be of any use to them, since what they needed was additional time. More funds couldn't help them to buy added expertise since such expertise is not available. This is because industry had never developed torpedos on their own. Up to now all torpedos have been developed in government owned and operated facilities.

Contract decision

Again, as no quantitative technical risk assessments were conducted, there was no assessment basis for the decision on contract type. The present contract is a Cost Plus Award Fee, with a maximum of 15 percent, which reflects the concern with the contractor's lack of prior experience in developing torpedos, in addition to the ambitious technical goals of the program. They expect the full production contract to be a Fixed Price with Incentive Fee type.

Risk management program

The MK-50 program was selected by the Navy for a pilot program to develop risk assessment guidelines to be used to manage risk. This pilot program is a direct result of the Defense Science Board's report on "Transition of Weapon Systems from Development to Production." The goal is to use existing management information systems to obtain measures, such as engineering change orders, scrap and rework rates, and engineering staff hours which would indicate transitional technical problem/risk areas.

The pilot program has been providing monthly charts to the MK-50 program office for two months. It still is in its early stages and both the program office and the Navy's sponsoring group are reluctant to provide much information on the effort. It appears, however, to be a promising approach. It is one that relies on measures such as scrap and rework rates and engineering staff hours, rather than financial, historical or statistical measures to assess critical risk areas.

AIR FORCE

The Air Force refused to adopt the TRACE approach toward dealing with risks. Therefore none of their programs has any

TRACE funding. The two programs we examined were the Advanced Tactical Fighter and the Advanced Medium Range Air-to-Air Missile.

Response to Initiative 11

The Air Force's response to Deputy Secretary Carlucci's Initiative 11 was to state that they were satisfied with the procedures they were already using to quantify risks and saw no advantage to the Army's TRACE approach. Further, the Air Force disagreed with the TRACE concept of withholding separate risk funds at headquarters level. The Air Force normally distributes all funds to the program manager, and would not agree to adopt the deferral concept.

ATF

The Advanced Tactical Fighter (ATF) is to be the Air Force's next combat plane, intended to replace the current F-15's and F-16's as the Air Force's front-line fighter. It is intended to provide air superiority with limited air-to-surface capability, but with growth potential for eventual air-to-surface mission dedication. The ATF is planned for deployment in the mid-1990's.

The program is in the Concept Exploration stage, with a Milestone I review scheduled for late fall 1984. Seven major aircraft manufacturers--Boeing, General Dynamics, Grumman, Lockheed, McDonnell Douglas, Northrop, and Rockwell International --were awarded contracts for this phase of the project. The seven contractors have proposed a total of 12 possible airframe designs, and the program office is currently in the process of narrowing the possibilities. Three contractors, from among the seven, will be awarded contracts to continue work in the Demonstration and Validation phase.

Technical risk assessment

The program office is using a PERT system to identify schedule risks, but has not yet performed any assessments of technical risks beyond discussions of potential technical problems. The deputy program manager stated that they plan to assess technical risks in the future, as they move into the next phase, using quantitative approaches taught by the Defense Systems Management College. The program office has not yet had to prepare the paperwork for a DSARC review, where technical risks must be discussed.

The ATF program is incorporating a number of different development programs from the Air Force's laboratories, particularly in the area of avionics. This has meant that much of the program office's knowledge of the technical risks comes from briefings by the laboratories. The laboratories have identified technical risks for various components as high, medium, or low, tied to the likelihood of entering Full-Scale Development at particular dates.

As a part of their contracts, the seven airframe manufacturers must each submit a draft System Concept Paper (SCP) as one of their deliverables. The SCP must include a discussion of technical risks, so the contractors will all be identifying technical risks for the program office.

Contract decision

The issue of technical risks did not enter into the decision on contract type, according to the contracting officer. He stated that he made the decision as to contract type on the basis of the Air Force's desire to involve multiple contractors at this stage, but with limited funding. All seven contractors have Firm Fixed Price contracts.

These contractors participated in a preliminary concept exploration phase at no cost to the Government, providing the Air Force with technical information from their laboratories and independent research and development efforts to aid in the feasibility studies for the ATF. The Air Force expects that their current contracts will not fully compensate the contractors for all their expenditures for the present phase of the project, but acknowledges that the future prospects make it worthwhile to the companies to invest some of their own funds to remain in competition. All seven contractors have signed no-cost extensions to their current contracts, to allow them continued access to Government test facilities until the Demonstration and Validation contracts have been awarded to the three winning firms.

AMRAAM

The Advanced Medium Range Air-to-Air Missile (AMRAAM) is a joint Air Force and Navy program, with the Air Force having a greater administrative role in the development of the weapon. This missile is intended to meet the needs of the Air Force and Navy for an air-to-air missile beginning in the mid-1980's. AMRAAM is to be compatible with the latest Air Force and Navy fighter aircraft.

Hughes Aircraft Company has been the prime contractor for the development of the AMRAAM. A leader-follower approach is to be used for full scale production of the missile, with Hughes remaining as the leader, and Raytheon Corporation taking a follower role. A leader-follower approach involves an initial (leader) company which develops the design and technology and then supplies these to another (follower) company which becomes a second source for production.

The program is at the Milestone III stage, making the transition from full scale development to production. Because of design difficulties, this transition has been stretched out with the use of a III-A review to approve only limited production until some problems are resolved.

At the Milestone II review, the DSARC approval included a requirement that the program use projected cost figures from the Cost Analysis Improvement Group (CAIG)--a group located in OSD which provides analysis to the DSARC--which were higher than those projected by the program office. The Program Element Monitor from the AMRAAM stated that the program then projected the difference between its original figures and the CAIG figures as allowance for risk.

Technical risk assessment

No quantitative assessment of technical risks has been performed for the AMRAAM. The program office has performed formal assessments of risks, but these assessments have been of schedule risks and cost risks, not of technical risks. In addition to the program office's own assessments, both Independent Schedule Analyses and Independent Cost Analyses have been performed.

The program office feels that it has an adequate system for identifying technical risks. The director of engineering stated that estimates of technical risks are based on the "collective judgment of their experts" as to whether the risks are "high, medium, or low."

Contract decision

Since the Validation phase of this project, the contract type has been Fixed Price. This was mandated at a higher level of command in the Air Force. The current contract is a Fixed Price contract with a 140 percent ceiling. The ceiling was moved to 140 percent from 135 percent to accommodate the Navy, which wanted a Cost-Plus type contract. The program is budgeted for the entire 140 percent.

The contracting officer stated that he learns about the technical risks by listening to the discussions and disagreements of the engineers at project meetings, but that technical risks did not play a role in the choice of contract type since the selection was dictated from above.

CONTRACTORS' INPUT

In order to gain an understanding of the contractor's role in the process of identifying and assessing technical risks, we arranged interviews with defense contractors. This provided us with a preliminary understanding of their inputs to the process of technical risk assessments and their perception of the need for quantitative risk assessment approaches.

Interviews were conducted with several individuals at two defense contractors' offices. Representatives of Hughes Aircraft Company and General Dynamics discussed the use of technical risk assessments in their work. There was a great deal of similarity in the statements from the two contractors on the topic.

Both contractors stated that their primary efforts on quantitative technical risk assessments were for proposals in bidding on contracts. In soliciting contractors to bid on projects, the Department of Defense puts out Requests for Proposals (RFP's). These RFP's usually call for a discussion in the proposals of any areas of technical risk in the project, so contractors must devote some attention to the issue in order to win the contract. Once the contract is won, they are unlikely to make continued quantitative assessments of technical risk.

The techniques they use are not the most sophisticated ones available. Both contractors make use of information from their own laboratories and internal R&D efforts in addition to information supplied by the Government. They call on their engineers to give expert opinions about technical risks for various components or subsystems and often use Delphi techniques to arrive at a distribution of expected technical risks for each. The assessments seem to lack sophistication in the integration of the information about components and subsystems into assessments of the whole system. The contractors do not, in general, use modeling techniques which would allow for interdependencies between the technical risks of the subsystems. Rather, they use summing or averaging approaches to combine the estimates, or they discuss the subsystems separately in their proposal presentations.

After a contract is awarded, neither contractor would be likely to continue performing quantitative technical risk assessments. Both contractors stated that during the contract they are the primary source of technical information to the Government, however. They brief the Government on the progress of the work, including presentations on technical risks, but they do not usually provide estimates of these risks other than to state that they are "high," "medium," or "low." The Government is given information about specific technical problems that occur, but this is not part of any unified technical risk assessment process.

The information provided by these two contractors, while in general agreement with one another, may not be representative of defense contractors as a whole. For example, in the 1983 Defense Risk and Uncertainty Workshop, held at the Air Force Academy, Martin Marietta Corporation presented information on a computer program they had developed for performing quantitative technical risk assessments. More information is therefore needed before any generalizations can be made about the usual role of defense contractors in examining technical risks.

DD-350

An additional effort in our scoping phase was an examination of another type of supporting document, the DD-350 form. A DD-350 form is filed for every contract action in DoD, whether for weapon system procurement, support services or supplies. Any DoD contract action should be reported on a form DD-350, providing

such information as the contractor's name, the contracting office, the amount of that action, the type of contract, the type of item being procured, and other similar categories.

The DD-350 data base contains information on all the contract actions by DoD for a given period of time. It can be used to provide background information on the types of contracts awarded, the types of weapon systems procured, and differences among services in procurements.

There are several difficulties with the DD-350 data base, however, which make it less than ideal for use in the current effort. Foremost among these is that the DD-350 contains no information on risks. If information on risks must be collected individually from each program office, then it is simpler and more direct to obtain contract information from the program office as well. Another important problem is the validity of the information contained in the DD-350 data base. There are questions as to the completeness, accuracy, and reliability of the data entered in the system. This problem seems to have been compounded for FY84 data, particularly the first quarter, when a new format and new data system were introduced.

Because of the problems with the data base, particularly the lack of any mechanism for associating risk information with the contract action data, we do not plan to use the DD-350s for the proposed effort. It does provide the potential for use in a possible later effort geared more specifically to contract types.

PRELIMINARY INDICATIONS

We did not find any cases of quantified technical risk assessments in any of the six weapon systems examined. In every case technical risks were assessed non-quantitatively on the basis of engineering expertise or prior experience of program office officials. We did find, however, three cases of quantitative cost risk analysis and two cases of quantitative schedule risk analysis. These results are shown in figure 3.

The decision on contract type is a complex one with technical risk as one of many factors considered. In our evaluation of the contract type decisions for the six systems, we verified its complexity. In fact, other than the two cases where the decision was external to the program office--as discussed for the AMRAAM and AV-8B--it was not clear how the decision came about or who the decisionmakers were, not to mention what factors were considered.

CONCLUSIONS

Given the significance and implications of the information obtained during the scoping phase, we are of the opinion that

Figure 3

Type of Quantitative Risk Assessments

QUANTITATIVE RISK ASSESSMENT:	ARMY		NAVY		AIR FORCE	
	M1E1	MLRS/TGW	AV-8B	ALWT	AMRAAM	ATF
♦ Technical	NO	NO	NO	NO	NO	NO
● Schedule	NO	YES	NO	NO	YES	NO
● Cost	YES	YES	NO	NO	YES	NO

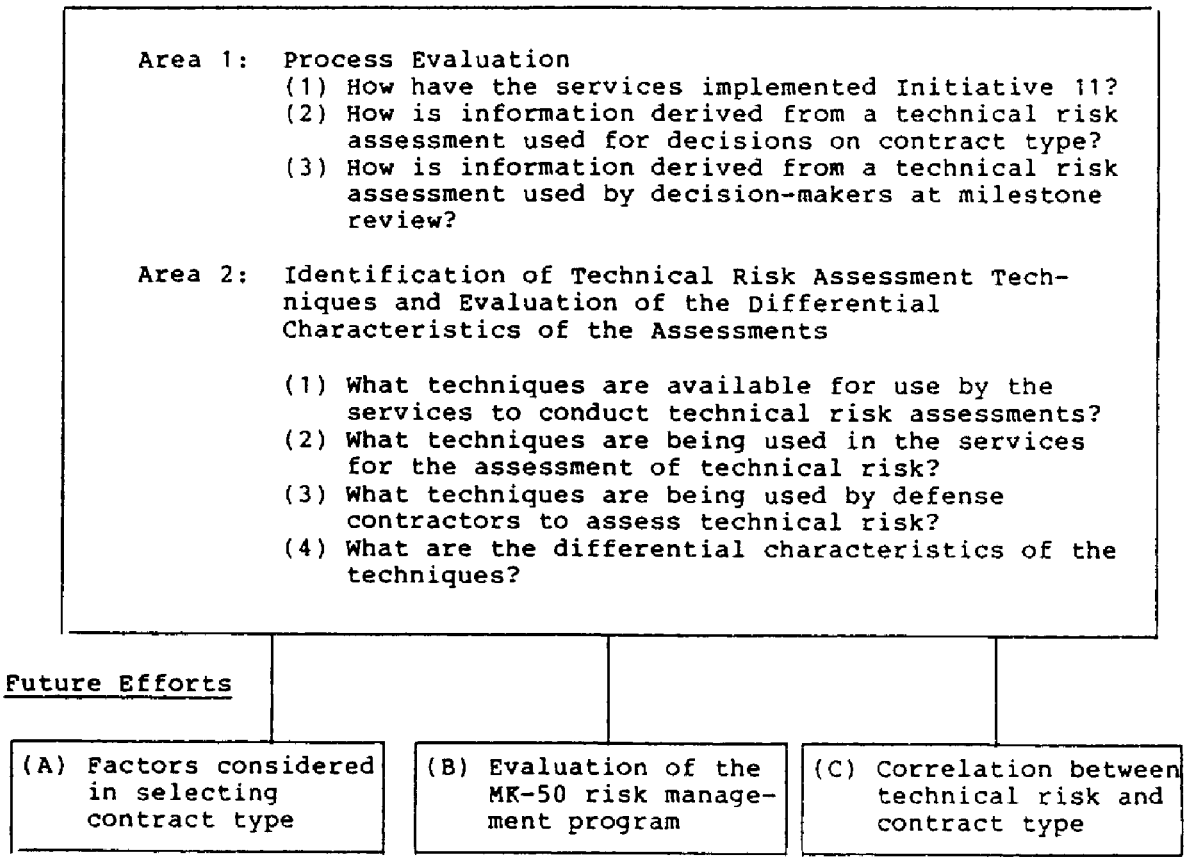
a necessary first effort is to proceed to implementation - in the form of a process evaluation--in order to assess the validity of the preliminary indications in a larger sample. Concurrent with the process evaluation we propose to identify available technical risk assessment techniques and to evaluate the differential characteristics of the assessments; this will include an evaluation of the TRACE programs, given that TRACE is the approach suggested by Deputy Secretary Carlucci. The design, end-product, users' needs and preliminary report outline are provided below; also, the associated questions are summarized under "Proposed Effort", figure 4.

Additionally, if the results of the proposed effort do not provide a better picture of the contract type decisionmaking process, we believe a second effort--figure 4, Future Efforts (A) --should be initiated then. Given that a causal relationship is not likely to have been firmly established, this effort would include an examination of the factors that are considered in the decision on contract type.

Also, in 12 to 18 months time--when the MK-50 risk management pilot effort has matured sufficiently to permit an evaluation of its results--it would be desirable to implement a fourth effort--figure 4, Future Efforts (B)--to address the following areas:

Figure 4

Proposed and Future Efforts



- validity of the "common wisdom" that transition to production problems are truly transitional, rather than the accumulation of previously unresolved problems which must be solved prior to production;

- results of the pilot program, including its applicability to other acquisition phases - other than transition to production; and,

- impact of the program on quality control and producibility issues.

Finally, an effort to examine the correlation between contract type and technical risk is presented in figure 4, Future Efforts (C).

IV. PROPOSED DESIGN

DoD and each of the services have issued directives and regulations requiring technical risk to be addressed. None of these documents directs the use of any particular assessment technique. The question thus becomes not only whether risk assessment is being performed as specified, but also whether the techniques that are used are providing reliable and accurate information about technical risk. Both of these questions are normative. The first requires a comparison of current practices with stated policies, while the second requires an examination of the empirical basis used for risk estimation, the reliability of the approaches used, and possible biases in the process.

In order to answer these larger questions, information is first needed with regard to several descriptive questions. For example, are the statements about "high," "medium," or "low" risk reported to the DSARC members supported by quantitative risk assessments or are they the personal opinion of an individual in the program office? What decisions are made on the basis of these statements about technical risk?

A number of alternative approaches were considered in deciding how best to proceed with the project. One of the alternatives considered was to select only one type of weapon system for in-depth examination. A weapon system which cut across all three services, such as missiles, could be selected. Information on technological problems encountered in the development of such systems could then be learned. All current missile systems, regardless of their stage in the acquisition cycle, from Concept Exploration to Production and Deployment could be included, to provide information on systems across a number of years. This would allow for the development of a matrix to examine technical risks by time of program initiation, by level of technological sophistication, etc., and to see how these factors were related to the kinds of technical risk assessments performed and the contract types used. This approach has several problems, however. It would not allow for generalizations about the relationship between technical risks and contract types; even if conclusions could be drawn for missile systems, there would be no way of knowing if the same factors were relevant for other types of weapon systems. The sample also might be unrepresentative since missile systems might be unique in many of the relevant characteristics under examination. For example, one type of technical risk assessment technique might be especially suited to missiles but not to other systems.

Another alternative approach considered was to use the DD-350 database, discussed above, to select a sample of weapon systems for study. Selection could be made on the basis of contract type and an examination of assessed technical risks could be made within categories. Any relationships found using this approach might be spurious, however, as the selection would not control for

other factors. Also, as previously mentioned, the database does not contain information on technical risks.

The approach which offers the broadest range of information is to perform a process evaluation of services' approaches to technical risk assessments. This process evaluation will lay the groundwork for a later outcome evaluation of the relationship with contract type. The proposed project is a necessary first step to enable the later correlative work to be placed in the proper context.

There are two broad areas which we propose to examine in this project. Area 1 is a process evaluation of how the services have implemented and used technical risk assessment methodologies. The results will then be compared with DoD's specifications to answer the normative question of how well DoD is implementing technical risk assessment. Area 2 is an identification and evaluation of the risk assessment techniques that are available and that are used. The results here will provide information on the normative question of reliability and adequacy of the techniques being used. The evaluation questions, scope, and methodology for implementing a project to examine these areas are provided below.

EVALUATION QUESTIONS

There are three major objectives of the project which relate to the first area. The first objective is to assess how the services have implemented Initiative 11, in accordance with Deputy Secretary Thayer's statements. The second objective is to determine how, or if, information derived from a technical risk assessment is used for decisions on contract type. The third objective is to determine how the information derived from a quantitative technical risk assessment is used by decisionmakers in reviewing the project.

The questions to be addressed in Area 1 are the following:

(1) How have the services implemented Initiative 11?

An examination will be made of how the services have implemented policies calling for quantitative technical risk assessments and how widely the policies have been implemented.

(2) How is information derived from a technical risk assessment used for decisions on contract type?

In selecting a contract type there are many kinds of information which may be considered. One possible factor is the anticipated risks of technical problems. The information derived from a technical risk assessment could thus provide important data for setting contract terms. The results of our scoping efforts, however, indicate that other factors have been more important in the decision on contract type. The objective here will therefore

be to determine if any information from a technical risk assessment is ever used in the selection of contract type, and, if so, how this information is considered in conjunction with other factors.

(3) How is information derived from a technical risk assessment used by decisionmakers at milestone reviews?

There are several layers of review within each service, as well as reviews by the Office of the Secretary of Defense, at which the issue of technical risk is discussed. An objective of the effort will be to determine how the results of a quantitative technical risk assessment enter into this process. Such a process evaluation will also include an examination of what information is used by decisionmakers when no quantitative assessment is performed.

For Area 2, there are four objectives of the project. First, this area will focus on the techniques which are available for use. A second objective is to determine what techniques are actually used by the services to assess technical risks. The third objective is to examine the techniques used by defense contractors when they prepare information on risks for the program offices. Finally, the differential characteristics of these technical risk assessment techniques will be evaluated. More specifically, the questions to be addresssed are the following:

(1) What techniques are available for use by the services to conduct technical risk assessments?

Each of the services provides guidance and training in technical risk assessment techniques. The different techniques which are available for use will be identified. This will include techniques developed by the services and taught to program personnel in various courses, as well as techniques developed by academicians or by industry. The particulars of each techniques will be examined, to learn what information and resources are required for application.

(2) What techniques are being used in the services' assessment of technical risk?

This question focuses on the reality of the risk assessments, i.e., what actually is being done in the program offices. Out of those that are available, the focus here is to identify the techniques that are used.

As part of the evaluation of the techniques being used by the services, particular attention will be given to the TRACE program. The Army has extended TRACE from RDT&E to Production and it has been recommended for use by all the services. It is thus important to consider whether or not the emphasis on TRACE is appropriate.

It is not possible to make definitive statements about the effectiveness of TRACE, since no control group exists for the TRACE programs. Nonetheless, issues can be addressed which either support or question the implementation of TRACE. For example, a study of TRACE could include the techniques used in the calculations and the use of the data resulting from the calculations. Questions could also be aimed at learning if TRACE funds are being used for the purpose for which they were originally intended.

(3) What techniques are being used by defense contractors to assess technical risk?

During the scoping phase it was learned that some program offices relied on the contractor for technical risk assessments. In the case of the ATF, for example, the contractors are being required to draft the System Concept Paper (SCP), which documents technical risk for the service SARC and DSARC. If the program office and those in the review process are relying on information from the contractors, then it seems important to study the techniques used by these industries in addressing technical risk. Their techniques will be identified and considered as described in (2).

(4) What are the differential characteristics of the techniques?

After the techniques have been identified, information will be gathered to allow for comparison of the technical risk assessment techniques. For example, questions will be asked concerning the level of detail the technique assumes, the degree of quantification involved, and whether or not a computer is used for the calculations. The time involved in the assessment and the available training will also be considered. An additional factor to be considered in evaluating an assessment technique is its appropriateness for the circumstances of the program.

SCOPE

Since all three services are said to have implemented programs to budget for technical risk, it is important to determine if there are differences in their approaches to risk assessment, their results, or in the information they supply to higher levels of review. Therefore, weapon system programs will be examined in the Army, the Navy, and the Air Force.

In addition to collecting information from weapon system program offices at the three services, data will also be gathered from defense contractors. Prime contractors for the weapon systems examined will be interviewed if identified as providing supporting data on technical risks to the Government. It may also be necessary to talk with a subcontractor or lower tier manufacturer if the prime contractor does not make the critical subsystems or components. Other defense contractors that have notable

techniques for technical risk assessment, as identified in scoping, will also be studied.

METHODOLOGY

The approach to be used to collect and analyze the necessary information will be discussed in the sections to follow, specifying the weapon systems selection criteria, data to be obtained, data collection methods, analysis plans, and use of experts.

Weapon systems selection criteria

Across the three services there are potentially 120-140 weapon system programs that could be classified as "major acquisitions" at a given time. We propose to take these major acquisition programs as the initial universe from which the programs to be examined will be selected. These programs usually include the most technologically sophisticated and complex systems. The major acquisitions have the greatest visibility and more congressional review than do other programs. They are therefore the most likely programs to have had emphasis placed on implementing the acquisition improvement initiatives.

Within this universe there will be some programs which are not appropriate for this project. We therefore propose to judgmentally exclude certain programs. Based on what we have learned thus far, programs which are either very early in the acquisition cycle or very late in the acquisition cycle are not appropriate for us to examine. Very early programs will not yet have prepared the documentation we will be requesting (discussed below), and will not yet have progressed through the levels of review we are interested in examining. Very late programs will be in full scale production, the approval for which requires that all technical risks be resolved. In addition, the relevant documentation for these programs is likely to have been prepared several years ago, due to the length of the acquisition process. We therefore propose to exclude from consideration any programs that have not passed the Milestone I review and any programs that have already passed a full Milestone III review. Programs which have passed a limited Milestone III review (III-A, III-B, etc.) will be retained as this indicates unresolved problems requiring postponement of full production. Thus, the programs to be examined will be in the Demonstration and Validation phase, the Full-Scale Development phase, or in limited production.

We further propose to exclude ship hull programs from consideration, but not ship systems such as electronics. Ship hulls are to be excluded because of the long periods of time involved in building them (up to 10 years), and the low level of technical risks for any but experimental deep-sea craft.

After these stated exclusions are made, the programs available for examination should include approximately 30 weapon

systems. We propose to examine all programs which meet our criteria.

The proposed sample of major acquisitions will allow us to look at specific types of weapon systems across all three services. An examination may thus be made of the consistency in assessing a particular set of technical risks. For example, all tactical missile programs may be examined, to provide a comparison of programs using similar technologies. It will also allow us to examine important emerging technologies across programs.

The programs to be examined will be the source of data for the process evaluation and will provide the information on what technical risk assessment techniques are used.

Data collection

Area 1

For all the selected programs, the documentation required to address the topic of technical risks will be collected. This will include the following:

- SCP - System Concept Paper.
- DCP - Decision Coordinating Papers.
- TEMP - Test and Evaluation Master Plans.
- Briefing charts and information prepared for DSARC and/or service SARC presentations.
- Any independent program assessments, such as Independent Cost Assessments, Independent Schedule Assessments, or Design Reviews.
- Any documentation of a risk assessment performed by, or for, the program office, including back-up information.
- The statements of work for the contracts.
- The Acquisition Plan (usually incorporated as an appendix to the SCP or DCP).

A content analysis will be used to evaluate the documentation of technical risk. A standard set of questions will address whether or not technical risk is discussed, and, if so, how it is discussed. We are interested specifically in learning whether or not quantitative terms are used to describe the degree of risk and/or whether the technique used in assessing risk is mentioned.

In addition, interviews will be conducted at each program office with the program manager (or deputy), the contracting

officer, the chief engineer, and other personnel as necessary (e.g., if someone else performed the technical risk assessment).

The prime contractor(s) for each program will be requested to supply any technical risk assessments which they performed for their proposal and/or as a part of their ongoing project work. Relevant contractor personnel will be interviewed. Subcontractors or lower tier manufacturers will be interviewed as necessary to obtain information on the critical subsystems or components.

Area 2

The data required for each question concerning technical risk assessment techniques are as follows:

(1) For the study of techniques available to the services, the data will consist of the documentation disseminated by each service, such as letters of instruction and handbooks. In addition, individuals at the training facilities, such as the Defense Systems Management College, will be interviewed regarding instruction in risk assessment techniques.

(2) Technical risk assessment techniques used by the services will be studied using the information gathered from the program offices. Risk assessments and any supporting documentation will be requested. In addition data will be obtained from the individuals involved in the risk assessments, including the staff that collected the information for the assessment and those that provided the information, for example engineers and systems analysts.

(3) Data obtained from defense contractors will be similar to those obtained from each of the services, that is, risk assessments, supporting documents, and interviews.

In addition to the documents gathered from program offices and contractors, much of the information needed will come from program personnel. These data will be collected through the use of a structured interview technique, using a combination of closed-ended and open-ended questions in individual interviews. A set of relevant closed-ended questions will be prepared for use, to ensure that basic information is gathered consistently at all program offices. Additional probing by interviewers will also be important, however, to increase the understanding of the unique program characteristics, so follow-up open-ended questions will be used.

The content of the questions centers on the evaluation questions outlined above. The interviews will focus on how technical risk assessments were performed and what use was made of the results. Individuals will be asked what techniques they used to perform technical risk assessments, including the particulars of how they chose their methods, where they obtained the information used in the assessments, the amount and kind of information used

(such as test results, expert judgment) and how often the assessments are updated. They will also be asked how the information derived through technical risk assessments has been used, as for managing the program, briefing decisionmakers, choosing the contract type, etc. Their perceptions of the utility of quantitative assessments will be addressed. If a quantitative assessment was performed, they will be asked if, and how, they have translated that into other terms, such as "high," "medium," or "low." If technical risks have not been assessed, they will be asked why not, what they are doing instead, and how they are providing information to decisionmakers on risks without such an assessment.

Analysis

The analyses of the information collected for both areas will be primarily qualitative rather than quantitative. The only quantitative analyses planned are simple cross-tabulations. The number of programs performing quantitative technical risk assessments by service will be calculated, and may be further broken down by technique, weapon system type, or command. In addition, the number of techniques used in each of the services will be reported, as well as statements concerning how many of these are formal, quantified, etc.

For Area 1 qualitative analyses will focus on the services' implementation of Initiative 11 and on technical risk information made available to decisionmakers. The analyses will therefore include a consideration of such things as the consistency of data found, the subjectivity of the information given to decisionmakers, the understanding of the requirements among program personnel, and the perceptions of the utility of quantitative technical risk assessment approaches.

The analyses for Area 2 will focus on comparisons of the techniques. Comparisons will focus on characteristics of the techniques which may serve to strengthen statements about risk to be made from the assessment as well as those characteristics that appear to be basic to any assessment of risk. It seems that the program managers and those involved in the review process treat risk assessments in a similar manner no matter what technique is used. Comparisons of the methods will offer evidence to determine whether or not this approach is warranted.

END-PRODUCT

Report to the Congress.

USERS' NEEDS

The results of the assignment will be of use to Congress in its oversight and appropriation roles. It will provide information to the Senate Committee on Governmental Affairs not only on DoD's needs for "Technical Risk Budget Deferrals" but also on DoD's efforts to assess risks. In addition, it will allow the

Congress to decide whether increased resources are needed in this area and whether they should be contingent on DoD's demonstration that quantitative technical risk assessments are being conducted.

The results of this assignment will also provide the Secretary of Defense with information on the services' efforts to implement Initiative 11. It will bring a clearer understanding of the ways in which technical risks are assessed, and of the actual use of technical risk information in the acquisition review process and in the program offices. This information will assist the Secretary in determining whether any new guidance, or clarification of old guidance, is needed, whether training in technical risk assessment techniques is adequate, and whether additional resources to conduct quantitative technical risk assessments should be requested.

PRELIMINARY REPORT OUTLINE

DIGEST

1. Introduction.

   A. Cost Growth of Major Acquisitions

   B. Technology Risk Impact

   C. Design

      1) Objectives
      2) Scope
      3) Methodology

2. Major Systems Development Perspectives.

   A. Policies

   B. Acquisition process

   C. Technical Risk Considerations

3. Performing the Analysis.

   A. Process Evaluation

      1) Implementation of Initiative 11
      2) Technical risk assessments and contract types

   B. Available Technical Risk Assessment

      1) Available to the services
      2) Used by the services
      3) Used by defense contractors
      4) Differential characteristics

4. Assessing the Results - Conclusions, Agency Comments and Our Recommendations.

   A. Conclusions

   B. Agency Comments

   C. Recommendations

      1) Secretary of Defense
      2) Congress

FOOTNOTES

[1]Rowe, A.J. (principal investigator) Disruption Research Project. Unpublished Paper, University of Southern California, April, 1977.

[2]Perry, R. et al., Systems Acquisition Strategies. R-733-PR/ARPA, RAND Corp., Santa Monica, CA, June, 1971.

[3]Deputy Secretary of Defense (David Packard) Memorandum, Subject: "Improvement in Weapon Systems Acquisition." Pentagon, Washington, D.C., July 31, 1969.

Deputy Secretary of Defense (David Packard) Memorandum, Subject: "Policy Guidance on Major Weapon System Acquisition." Pentagon, Washington, D.C., May 28, 1970.

[4]Office of Management and Budget Circular A-109, Subject: "Major System Acquisition." Washington, D.C., April 5, 1976.

[5]U.S. Congress, Committee on Government Operations, House Report No. 96-656. Inaccuracy of Department of Defense Weapons Acquisition Cost Estimates. November 16, 1979, p. 12.

[6]U.S. General Accounting Office Report, The Army Needs More Comprehensive Evaluations to Make Effective Use of Its Weapon System Testing. NSIAD-84-40, February 24, 1984.

[7]Deputy Secretary of Defense (Frank C. Carlucci) Memorandum, Subject: "Improving the Acquisition Process." Pentagon, Washington, D.C., April 30, 1981.

[8]Deputy Secretary of Defense (Paul Thayer) Memorandum, Subject: "Guidance on the Acquisition Improvement Program." Pentagon, Washington, D.C., June 8, 1983.

[9]Defense Systems Management College, Risk Assessment Techniques: A Handbook for Program Management Personnel. Fort Belvoir, VA, July 1983.

[10]Deputy Secretary of Defense (Frank C. Carlucci) Memorandum. Pentagon, Washington, D.C., June 18, 1982.

BIBLIOGRAPHY

Amdor, Stephen L., and Roy R. Kilgore. Quantitative Risk Assessment: A Test Case. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, School of Systems and Logistics, 1974. Master's thesis.

Anderson, Richard M. Handling risk in defense contracting. Harvard Business Review, July 1969, Volume 47, pp. 90-98.

Army Logistics Management Center, A Course of Instruction in Risk Analysis, Fort Lee, Va.: 1971.

Babiarz, Anthony S., and Peter W. Giedras. A Model to Predict Final Cost Growth in a Weapon System Development Program. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, School of Systems and Logistics, 1975. Master's thesis.

Baillie, Allan S. Management of risk and uncertainty. Research Management, March 1980, Volume 23(2), pp. 20-24.

Banash, Robert C., and James B. Beeson. Cost/Schedule Uncertainty Analysis of the XM1/Alternative Programs. Rock Island, Ill.: Army Armament Command, Systems Analysis Directorate, 1976.

Barnett, Paul J., and Harman K. Wales. An Assessment of the Applicability of Production Readiness Reviews to Multinational Coproduction Programs. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, School of Systems and Logistics, 1981. Master's thesis.

Beeckler, C. Eugene, and Kimrey D. Newlin. Economic Price Adjustment (EPA) Provisions. Fort Lee, Va.: Army Procurement Research Office, U.S. Army Logistics Management Center, 1977.

Bell, Chauncey F. Cost-Effectiveness Analysis as a Management Tool. Santa Monica, Calif.: RAND Corporation, October 1964.

Bevelhymer, Herbert L. A Proposed Methodology for Weapon System Development Risk Assessment. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, School of Engineering, 1973. Master's thesis.

Blanning, Robert W., Paul R. Kleindorfer, Steven Dana, and Ellen J. Goldman. Research Opportunities in the Management of Weapons Systems Acquisition: Executive Summary. Philadelphia, Pa: Wharton School of Finance and Commerce, Department of Decision Sciences, September 1975.

Blanning, Robert W., Paul R. Kleindorfer, Steven Dana, and Ellen J. Goldman, Research Opportunities in the Management of Weapons Systems Acquisition. Philadelphia, Pa.: Wharton School of Finance and Commerce, Department of Decision Sciences, November 1975.

Blue Ribbon Defense Panel. Report to the President and the Secretary of Defense on the Department of Defense, Washington, D.C.: July 1970 (Appendix E. Staff Report on Major Weapon Systems Acquisition Process).

Bruno, O.P., and Raymond Bell. Analysis of testing risks for an air defense system. Proceedings 1975 Annual Reliability and Maintainability Symposium, January 28-30, 1975, Washington, D.C. New York: Institute of Electrical and Electronics Engineers, Inc., pp. 427-431.

Carter, R.D. A Survey of Techniques for Improving Cost Estimates of Future Weapon Systems. Falls Church, Va.: Analytic Services, Inc., March 1965.

Colvin, Charles E. Radar hardware second buy decision risk analysis. Proceedings of the Twelfth Annual U.S. Army Operations Research Symposium, Volume II, October 2-5, 1973. Washington, D.C.: Department of the Army, Office of the Assistant Chief of Staff for Force Development, pp. 745-757.

Congressional Budget Office. A Review of the Department of Defense December 31, 1982 Selected Acquisition Reports (SARs), August 1983.

Corey, James M. The inefficiency of sealed-bid competition. Defense Systems Management Review, Spring 1980, Vol. 3(2), pp. 42-46.

Cox, Larry, and Michael Bohn. Acquisition Strategy Comparison Model (ASCM), Volume I, Executive Summary and Report. Arlington, Va.: Analytic Sciences Corp., May 1982.

-----. Acquisition Strategy Comparison Model (ASCM), Volume II, Appendices. Arlington, Va.: Analytic Sciences Corporation, 1982.

Crawford, Leslie P. A Case Study in Risk/Decision Analysis. Fort Belvoir, Va.: Defense Systems Management College, 1973. Study project report.

Cuff, James D. Risk-decision analysis in weapons system acquisitions. Long Range Planning, March 1973, Vol. 6(1), pp. 49-55.

Cummins, J. Michael. Incentive contracting for national defense: a problem of optimal risk sharing. Bell Journal of Economics, Spring 1977, Vol. 8(1), pp. 168-185.

Davis, Guy W. The Dilemma of Uncertainties Associated with Cost Estimating in the Project Management Office. Fort Belvoir, Va.: Defense Systems Management School, 1976. Student Study Report.

DeMong, Richard F. The effectiveness of incentive contracts: what research tells us. National Contract Management Quarterly Journal, Fourth Quarter, December 1978, Volume 12 (4), pp. 12-22.

Dixon, Max Wayne. A Statistical Analysis of Deviations from Target Cost in NAVAIRSYSCOMHQ Fixed-Price Incentive Contracts During the 1949-1965 Time Frame. Monterey, Calif.: Naval Postgraduate School, March 1973.

DOD tests contract incentives (in News Summary). Defense Management Journal, Second Quarter 1983, Volume 19(2), p. 43.

Eberth, Robert William. Escalation Provisions in DOD Procurement: A Review of the Problem and a Framework for Analysis. Monterey, Ca.: Naval Postgraduate School, 1974. Master's thesis.

Edgar, Lt. Col. John D., USAF. Controlling Murphy: how to budget for program risk. Concepts: The Journal of Defense Systems Acquisition Management, Summer 1982, Volume 5(3), pp. 60-73.

Emmelhainz, Margaret A. Innovative contractual approaches to controlling life cycle costs. Defense Management Journal, Second Quarter 1983, Volume 19(2), 36-42.

Fallows, James. National Defense. New York: Random House, 1981.

Fisher, G. H. The Nature of Uncertainty. Santa Monica, Ca.: RAND Corp., 1973.

Fox, J. Ronald. Arming America: How the U.S. Buys Weapons. Cambridge, Mass.: Harvard University Press, 1974.

-----. Note on government contracting and methods of government procurement. Harvard Business School, mimeo. Boston, Mass.: Intercollegiate Case Clearing House, 1980.

Gansler, Jacques S. A new dimension in the acquisition process. Defense Systems Management Review, Autumn 1977, Volume 1(4), pp. 6-12.

-----. The Defense Industry. Cambridge, Mass.: MIT Press, 1980.

Gates, Robert K., Robert S. Bicknell, and John E. Bortz. Quantitative models used in the RIW decision process. Proceedings 1977 Annual Reliability and Maintainability Symposium, January 18-20, 1977. Washington, D.C.: AR&M Symposium, pp. 229-236.

Gerber, Hans U. An Introduction to Mathematical Risk Theory. Philadelphia, Pa.: University of Pennsylvania, Wharton School, S.S. Huebner Foundation for Insurance Education, 1979.

Gilby, Howard M. Decision risk analysis of the impact of the heavy lift helicopter advanced technology component (ATC) program of alternative methods of powering the ATC dynamic system test rig. Proceeding of the Twelfth Annual U.S. Army Operations Research Symposium, Volume 1, October 2-5, 1973. Washington, D.C.: Department of the Army, Office of the Assistant Chief of Staff for Force Development, pp. 572-582.

Glover, William L., and John O. Lenz. A Cost Growth Model for Weapon System Development Programs. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, School of Systems and Logistics, 1974. Master's thesis.

Gordon, Harvey J. The role of the contract in systems acquisition. Defense Systems Management Review, Winter 1980, Volume 3(1), pp.30-42.

Guide for Transitioning Army Missile Systems From Development to Production. Army Missile Command, Redstone Arsenal, Alabama: Systems Engineering Directorate, July 1981. Technical Report RS-81-6.

Hanrahan, John D. Government by Contract. New York: Norton, 1983.

Hersh, Michael H. Risk Aversion vs. Technology Implementation. Fort Belvoir, Va.: Defense Systems Management College, 1977. Study project report.

Hlavinka, Duane K. Lessons Learned: Production Restart of a Major Weapons System. Fort Belvoir, Va.: Defense Systems Management School, May 1976.

Hoivik, Thomas Harry. The Navy Test and Evaluation Process in Major Systems Acquisition. Fort Belvoir, Va.: Defense Systems Management College, 1976.

Hunt, Raymond G. Contractor Responses to Award Fee Contracts. National Contract Management Jounral, Winter 1982, Volume 15(2) pp. 84-90.

Hwang, John D., David Chappell, and Howard M. Gilby. Risk analysis of the improved cobra armament program. Proceedings of the Twelfth Annual U.S. Army Operations Research Symposium, Volume 2, October 2-5, 1973. Washington, D.C.: Department of the Army, Office of the Assistant Chief of Staff for Force Development, pp. 736-744.

-----. An impact assessment algorithm for R&D project risk analysis. Proceedings of the Twelfth Annual U.S. Army Operations Research Symposium, Volume 2, October 2-5, 1973. Washington, D.C.: Department of the Army, Office of the Assistant Chief of Staff for Force Development, pp. 559-571.

Ingalls, Edward G., and Peter R. Schoeffel. Risk assessment for defense acquisition managers. Program Manager, September - October 1983, pp. 27-33.

Insley, Patricia A., William P. Hutzler, Gerald R. McNichols, A. Stuart Atkinson, and Ronald E. James. Shortening the Acquisition Cycle: Research on Concurrency (Phase I Report). Falls Church, Va.: Management Consulting and Research, Inc., 1982.

Jones, Julius E. An Analysis of Incentive Contracts with Respect to Risk. Fort Leavenworth, Kan.: Army Command and General Staff College, 1971.

Kaufman, Richard F. The War Profiteers. New York: Bobbs-Merrill, 1970.

Kost, John D., Jr. Defense Management Simulation (1973 Version). Washington, D.C.: Industrial College of the Armed Forces, Simulation and Computer Directorate, 1973.

Lave, Lester R., ed. Quantiative Risk Assessment in Regulation. Washington, D.C.: The Brookings Institution, 1982.

Lenk, Barry R. Government Procurement Policy: A Survey of Strategies and Techniques. Washington, D.C.: George Washington University, Program in Logistics, 1977.

Lewark, William H., Jr. The Technical Assessment Annex - A Formal Technical Risk Analysis Role for the Air Force Laboratories in the DSARC Process. Fort Belvoir, Va.: Defense Systems Management School, 1975. Student Project Report 75-2.

Lewis, Warfield M., Jr. A Simple Statistical Method of Presenting the Uncertainty Associated with Life Cycle Cost Estimates, Fort Belvoir, Va.: Defense Systems Management school, 1973. Study project report.

Lilge, Ralph W. Total Risk Assessing Cost Estimate (TRACE), An Evaluation. U.S. Army Aviation Research and Development Command, Systems and Cost Analysis Division (DRDAV-BCD), St. Louis, Mo., February 1979.

Lochry, Col. Robert R., Lt. Col. R.D. Hensley, Lt. Col. Philip Flammer, Lt. Col. Donald R. Smith, Maj. Richard G. Head, Maj. Eugene M. Henry, Maj. William R. Hodson, III, Maj. Eric B. Nelson, Capt. Gerald G. Carson and Capt. John F. Cuilmartin. Final Report of the USAF Academy Risk Analysis Study Team. U.S. Air Force Academy, Colo., 1971.

Long, John Amos. Life Cycle Costing in a Dynamic Environment. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, 1983. Doctoral dissertation.

Lorette, Richard J. Do we really want research on the acquisition of major weapon systems? National Contract Management Journal, Winter 1976-77, Vol. 10(2), pp. 64-70.

Losquadro, Joseph P. Management Control in Weapons Systems Acquisition. Monterey, Calif.: Naval Postgraduate School, September 1978.

Lowrance, William W. Of Acceptable Risk: Science and the Determination of Safety. Los Altos, CA: William Kaufman, Inc., 1976.

Mann, Maj. Greg A., USAF. VERT: A risk analysis tool for program management. Defense Management Journal, May/June 1979, Vol. 15(3), pp. 32-36.

Martin, Martin Dean. A Conceptual Cost Model for Uncertainty Parameters Affecting Negotiated, Sole-Source Development Contracts. Norman, Okla.: University of Oklahoma, 1971. Doctoral dissertation.

Martin, Lt. Col. Martin D., Maj. Gerald R.J. Heuer, Capt. John C. Kingston, and Capt. Eddie L. Williams. A proposed definition and taxonomy for procurement research in the DOD. National Contract Management Journal, Winter 1977-1978, Vol. 11(2), pp. 89-105.

-----. An evaluation of the definition, classification and structure of procurement research in the DOD. National Contract Management Quarterly Journal, Fourth Quarter, December 1978, Vol. 12(4), pp. 35-59.

Martin, M.D., Alan J. Rowe and Herold A. Sherman (eds.). Proceedings of Management of Risk and Uncertainty in the Acquisition of Major Programs, U.S. Air Force Academy, Colorado Springs, February 9-11, 1981. Los Angeles, CA: University of Southern California, 1981.

Mazza, Thomas N., and Robert C. Banash. Decision risk analysis for XM204, 105mm howitzer, towed reliability/durability requirements. Proceedings of the Twelfth Annual U.S. Army Operations Research Symposium, Volume 2, October 2-5, 1973. Washington, D.C.: Department of the Army, Office of the Assistant Chief of Staff Development, pp. 445-460.

McNichols, Gerald R. Cost-risk procedures for weapon system risk analysis. Proceedings, Annual Reliability and Maintainability Symposium, January 27-29, 1981. New York: Institute of Electrical and Electronics Engineers, pp. 86-94.

Meehan, John D., and Thomas O. Millett. Major Weapon System Acquisition: An Analysis of DOD Management Arrangements. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, School of Engineering, September 1968.

Meneely, Commander Frank T., SC USN. Determining the appropriate contract type. Concepts: The Journal of Defense Systems Acquisition Management, Summer 1982, Vol. 5(3), pp. 44-49.

Meridian Corp. Preliminary Analysis of Technical Risk and Cost Uncertainty in Selected DARPA Programs (Final Report). Falls Church, Va.: Meridian Corp., 1981.

-----. Preliminary Analysis of Technical Risk and Cost Uncertainty in Selected DARPA Programs (Interim progress report). Falls Church, Va., 1981.

Moore, William F., and John M. Cozzolino. More effective cost-incentive contracts through risk reduction. Defense Management Journal, July 1978, Vol. 14(4), pp. 12-17.

Morehouse, W. Progress in Resources Planning Through PERT. Utica, NY: General Electric Co., Light Military Electronics Department, 1960.

Morrow, Garcia E., David D. Acker, Eugene Beeckler, and Elmer H. Birdseye. Lessons Learned: Multiple Launch Rocket System. Arlington, Va.: Information Spectrum, Inc.; and Ft. Belvoir, Va.: Defense Systems Management College, 1980.

Navy Department. Justification of estimates for fiscal year 1984. Submitted to Congress January 1983, Procurement, Weapons Procurement, Navy.

Neuman, Frederick. How DCAA uses risk analysis in planning and programming audits. Internal Auditor, June 1979, Vol. 36(3), pp. 32-39.

Neuman, Gary. Platoon early warning device (PEWD) decision risk analysis (DRA). Proceedings of the Twelfth Annual U.S. Army Operations Research Symposium, Volume 2, October 2-5 1973. Washington, D.C.: Department of the Army, Office of the Assistant Chief of Staff for Force Development, pp. 690-701.

Niemeyer, William A. James C. Emanuel, Edward T. Kusterer, and Ronald J. Vegoda. Technical Risk Assessment of Extended Configurations of the M113A1E1. Aberdeen Proving Ground, Md.: Army Materiel Systems Analysis Activity, 1978.

Norton, Monte G., Richard D. Abeyta, and Paul E. Grover. Production Risk Assessing Methodology (PRAM). Fort Lee, Va.: Army Procurement Research Office, 1982.

Office of the Assistant Secretary of Defense (Comptroller). Acquisition Management, Selected Acquisition Report.

O'Flaherty, J. Identification and Estimation of High Cost or High Risk Elements. McLean Va.: Research Analysis Corp., 1970.

Pace, Dean Francis. Negotiation and Management of Defense Contracts. New York: Wiley-Interscience, 1970.

Palmer, Daniel L. Evaluation of automatic transmissions for use in military wheeled vehicles (decision risk analysis). Proceedings of the Twelfth Annual U.S. Army Operations Research Symposium, Volume 2, October 2-5, 1973. Washington, D.C.: Department of the Army, Office of the Assistant Chief of Staff for Force Development, pp. 668-679.

Pearcy, Stephan R. Decision analysis for XM578 APFSDS cartridge development program. Proceedings of the Twelfth Annual U.S. Army Operations Research Symposium, Volume 2, October 2-5, 1973. Washington, D.C.: Department of the Army, Office of the Assistant Chief of Staff for Force Development, pp. 426-436.

Perry, Robert. American Styles of Military Research and Development. Santa Monica, CA: Rand Corporation, 1979.

Perry, Robert, Giles K. Smith, Alvin J. Harman, and Susan Henrichsen. System Acquisition Strategies. Santa Monica, CA: RAND Corp., June 1971.

Petruschell, R.L. Project Cost Estimating. Santa Monica, Calif.: RAND Corporation, September 1967.

Proceedings of Department of Defense Cost Research Symposium, March 2-3, 1966. Washington, D.C.: Assistant Secretary of Defense (Systems Analysis).

Proceedings of the Twelfth Annual U.S. Army Operations Research Symposium, Volume 2, October 2-5, 1973. Washington, D.C.: Office of the Assistant Chief of Staff for Force Development, Department of the Army.

Raiffa, Howard. Decision Analysis: Introductory Lectures on Choices Under Uncertainty. Reading, MA: Addison-Wesley Publishing Co., 1968.

Reid, Seton M. Decision risk analysis of the An/TSQ-73. Proceedings of the Twelfth Annual U.S. Army Operations Research Symposium, Volume 2, October 2-5, 1973. Washington, D.C.: Department of the Army, Office of the Assistant Chief of Staff for Force Development, pp. 718-724.

Rippy, D., and P. Sweeney. Penetration Study: Behavioral Aspects of Decisions Under Uncertainty During Weapon System Acquisition. Dayton, Ohio: University of Dayton, School of Engineering, 1980.

Risk Assessment Techniques: A Handbook for Program Management Personnel (First Edition). Fort Belvoir, Va.: Defense Systems Management College, July 1973.

Scott, Eugene L. The cost growth phenomenon. National Contract Management Journal, Winter 1983, Vol. 16(2), pp. 37-45.

Shea, Joseph F. Observations on defense acquisition. Defense Systems Management Review, Vol. 1(4), Autumn 1977, pp. 29-36.

Sherrer, Charles W. Achieving a higher and more competitive state-of-the-art in DOD procurement procedures. National Contract Management Journal, Winter 1982, Vol. 15(2), pp. 71-83.

Simonson, G.R. Misconceptions of profit in defense policy. National Contract Management Journal, Winter 1982, Vol. 15(2), pp. 15-20.

Sizelove, J. Douglas. Remotely monitored battlefield sensor system (REMBASS) program decision risk analysis. Proceedings of the Twelfth Annual U.S. Army Operations Research Symposium, Volume 2, October 2-5, 1973. Washington, D.C.: Department of the Army, Office of the Assistant Chief of Staff for Force Development, pp. 712-717.

Smith, Charles H., and Charles M. Lowe, Jr. Sole source and competitive price trends in spare parts acquisition. National Contract Management Journal, Winter 1982, Vol. 15(2), pp. 51-56.

Smith, Giles K. Air Force Acqusition Options for the 1980s: A Briefing on Study Plans. Santa Monica, CA: Rand Corp., 1979.

Solinsky, Kenneth S. Controlled competition for optimal acquisition. Defense Systems Management Review, Spring 1980, Vol. 3(2), pp. 47-55.

Springer, Robert M., Jr. Controlling risk in reliability incentive contracting. National Contract Management Journal, Winter 1975-76, Vol. 9(2), pp. 1-9.

Stimson, Richard A., and A. Douglas Reeves. Improving defense contractor productivity. Defense Management Journal, Third Quarter 1983, Vol. 19(3), pp. 41-44.

Suver, Lt. Col. James D., and Maj. F. Theodore Helmer. A Bibliography of Selected Studies in the Weapons Acquisition Area. United States Air Force Academy, Colo.: Department of Economics and Management, 1972.

Svetlich, William G. The Systems Acquisition Process and Its Limitations in the Department of Defense. Washington, D.C.: National Defense University, 1979. Monograph, National Security Management series.

Thompson, William E., III. Risk implications for cost growth in weapon system acquisition programs. Concepts: The Journal of Defense Systems Acquisition Management, Spring 1982, Vol. 5(2), pp. 116-128.

Timson, F.S. Technical Uncertainty, Expected Contract Payoff, and Engineerng Decisionmaking in a System Development Project. Santa Monica, Calif.: Rand Corp. 1970.

Trueman, Richard E. An Introduction to Quantitative Methods for Decision Making. New York: Holt, Rinehart and Winston, Inc., 1974.

U.S. General Accounting Office. A Range of Cost Measuring Risk and Uncertainty In Major Programs--An Aid to Decision-making, PSAD-78-12. Washington, D.C.: February 2, 1978.

-----. Delays in Definitizing Letter Contracts Can Be Costly to the Government, PSAD-80-10. Washington, D.C.: November 16, 1979.

-----. Better Navy Management of Shipbuilding Contracts Could Save Millions of Dollars, PSAD-80-18. Washington, D.C.: January 10, 1980.

-----. Financial Status of Major Federal Acquisitions September 30, 1979, PSAD-80-25. Washington, D.C.: Feberuary 12, 1980.

-----. Issues Identified in 21 Recently Published Major Weapon System Reports, PSAD-80-43. Washington, D.C.: June 12, 1980.

U.S. General Accounting Office. Use of Cost-Deferred-Fee Contracts Can be Costly to the Government, MASAD-81-10. Washington, D.C.: March 11, 1981.

-----. Financial Status of Major Federal Acquisitions September 30, 1980, MASAD-81-13. Washington, D.C.: March 20, 1981.

-----. Acquiring Weapon Systems In A Period of Rising Expenditures: Implications For Defense Management, MASAD-81-26. Washington, D.C.: May 14, 1981.

-----. Improving the Weapon Systems Acquisition Process, MASAD-81-29. Washington, D.C.: May 15, 1981.

-----. The Navy Can Reduce the Cost of Ship Construction If It Enforces Provisions of the Contract Escalation Clause, PLRD-81-57. Washington, D.C.: August 24, 1981.

-----. Status of Major Acquisition As Of September 30, 1981: Better Reporting Essential to Controlling Cost Growth, MASAD-82-24. Washington, D.C.: Aprill 22, 1982.

-----. Improving the Effectiveness and Acquisition Management of Selected Weapon Systems: A Summary of Major Issues and Recommended Actions, MASAD-82-34. Washington, D.C.: May 14, 1982.

-----. Assessment of Admiral Rickover's Recommendations to Improve Defense Procurement, PLRD-83-37. Washington, D.C: January 27, 1983.

-----. Status of Major Acquisitions as of September 30, 1982, NSIAD-83-32. Washington, D.C.: September 7, 1983.

-----. Weapon Systems Overview: A Summary of Recent GAO Reports, Observations and Recommendations on Major Weapon Systems, NSIAD-83-7. Washington, D.C.: September 30, 1983.

-----. The Army Needs More Comprehensive Evaluations to Make Effective Use of Its Weapon System Testing, NSIAD-84-40. Washington, D.C.: February 24, 1984.

Wall, William C., Jr. The prudent use of engineers in program management. Defense Management Journal, March/April 1979, Vol. 15(2), pp. 14-18.

Wendt, Robert L. Practical implications of acquisition policy reform. Defense Management Journal, Fourth Quarter 1982, Vol. 18(4), pp. 15-19.

Wentz, William. Understanding how the defense department allocates resources. GAO Review, Summer 1983, pp. 27-29. 49.

Williams, Robert F., and Richard D. Abeyta (eds.). Management of risk and uncertainty in systems acquisition. Proceedings of the 1983 Defense Risk and Uncertainty Workshop, July 13-15, 1983, Defense Systems Management College, Ft. Belvoir, Va. Fort Lee, Va.: U.S. Army Materiel Systems Analysis Activity, Army Procurement Research Office.

Woody, Col James R., USAF. DOD procurement policy: the effect on aerospace financial risk. Defense Management Journal, Third Quarter 1982, Vol. 18, pp. 22-27.

Worm, George H. Application of Risk Analysis in the Acquisition of Major Weapon Systems. Clemson, SC: Clemson University, Department of Industrial Management, 1980.

III. DESIGN PAPER FOR THE WOMEN'S RETIREMENT PROJECT: AN EVALUATION OF THE FACTORS THAT INFLUENCE THE TIMING OF RETIREMENT FOR WOMEN

Joanne Frankel

Scott Crosse

January 1985

Table of Contents

I. INTRODUCTION

This paper describes the results of the scoping and planning phase of the women's retirement project. The purpose of this phase was to plan a study of the factors that influence the timing of retirement for women, contrasting them with the factors important for men. Analysis of data from retired and working women (55 and over) would focus on finances, health, demographics, labor force attachment, and, especially, attitudes, such as commitment to work, job satisfaction, and perception of peer and family pressures. We would try to describe retired women on each of these factors as they were just before their retirement, and compare their characteristics to those of women of the same age who had not yet retired. By examining these factors for both groups and relating them to already taken decisions (or to retirement plans), we would attempt to identify those factors that seemed to influence the retirement decision. Information from younger women (40 to 54) would be included to test for age-related differences in characteristics which might have implications for future retirement patterns. Information about men would be used for comparative purposes.

Given the aging of America's population, the increased labor force participation of women, and the increased national concern about retirement income programs, this sort of information is timely and important. However, after carefully considering alternative study designs, we conclude that the women's retirement project should not be implemented by PEMD. We base this conclusion on two arguments.

First, our analysis of existing data sources shows that none would provide the information needed in this study. Therefore, a new, costly, data collection effort would be necessary. Moreover, carrying out this effort would involve considerable risk, because of measurement problems and procedural hurdles in each of the data collection options we considered.

Second, even if we could do it successfully, the product of this work would be descriptive, not predictive. That is, the study would result in a report documenting differences in finances, health, employment, and attitudes between women who decide to continue working and those who decide to retire at particular ages. It would also document the retirement plans of younger groups of working women, as well as differences among them in finances, labor force attachment, and so on. It could not, however, in the time available for implementation (one to two years) develop predictive information about likely retirement patterns for future cohorts.

Predicting future retirement patterns would require a longitudinal study and, according to PEMD staff working in this area, additional model-building analytic work, which together could require about four to six years. Without this additional

work, which we do not now propose, it is not clear that the descriptive information is of enough interest to justify the cost of the study. That is, such information would not allow us to quantitatively evaluate specific policies under consideration by the Congress, nor to contribute to the accuracy of predictions about the future.

In sum, therefore, it seems that the likely benefits of the study do not justify its costs. The remainder of this paper describes how we came to this conclusion. It begins by explaining why we chose women's retirement as an area worthy of investigation. It then describes the objectives of the planned study, the information we wanted to collect from specific groups, the major design issues we addressed, and the alternatives we considered. Finally, we summarize the findings that led to our final recommendation.

II. WOMEN'S RETIREMENT: AN IMPORTANT BUT RELATIVELY NEGLECTED AREA

Importance of the Topic

Why is it useful to know when various groups of women now in the labor force are likely to retire, and the factors that are likely to influence the timing of that decision? Over the long run, such information is important because the future status of retirement income programs, such as Social Security, Civil Service retirement, and private pension plans, is very sensitive to aggregate trends, not only in demographic factors like fertility and longevity, but also in retirement age (Light, 1983).

Being able to predict the retirement ages of various cohorts of women becomes increasingly important as more and more women become eligible for retirement benefits in their own right. Consider the Social Security system, for example. The Office of the Actuary predicts that, in 2000, nearly 70 percent of retirement age women will be eligible for Social Security benefits based on their own work histories. An economist with the National Commission on Employment Policy sets the figure even higher, at 80 percent (Sandell, 1983).

Small changes in retirement ages of these women can mean millions of dollars more or less than anticipated to pay for not only Social Security retirement benefits, but also federal retirement and private pensions, all of which are "big ticket items." According to the 1984 Statistical Abstract, in 1982 (the latest date for which these figures are presented in the Abstract), outlays for Social Security retirement benefits amounted to $156 billion or about 21 percent of all federal outlays, and Civil Service retirement benefits consumed another $19.6 billion, or about an additional 3 percent of outlays. During that same year, private pensions controlled over $573 billion in assets.

Making the predictions about women's retirement trends that will affect these pension systems involves taking into account the potential effect of changes in female labor force participation. For all women 16 and older, labor force participation jumped to 53 percent in 1982 from only 38 percent in 1960. In 1982, nearly 70 percent of women in the 20-44 year old age group, 62 percent of 45-54 year old women, and 42 percent of 55-64 year old women were in the labor force. In contrast, labor force participation for men dropped from 83 percent in 1960 to 77 percent in 1982, with the most extreme decline among 55-64 year olds, reflecting the male trend toward early retirement (Statistical Abstract, 1984).

The figures for women suggest that more are entering the labor force relatively early (participation rates for 20-24 year olds went from 46 percent in 1960 to 70 percent in 1980) and

staying in more continuously. Moreover, important attitudinal changes, paralleling the changes in labor force participation, seem to have taken place. According to one analysis of poll data (Oppenheimer, 1970), while employment of single women was generally accepted in the 1930s, employment of married women was not. In contrast, analysis of data from the National Longitudinal Survey showed that by the mid-seventies, the "coming of age" for women who will reach retirement after the turn of the century, a significant change toward more egalitarian attitudes about the appropriateness of work for women had occurred (Waite, 1978).

Will these changes in labor force participation and work attitudes affect the timing of retirement for women? One way of approaching this question is to look at the cross-sectional data for a group of older women whose work experience parallels some of today's younger women, that is, for a group that entered relatively early and worked more or less continously.

Never-married women seem to meet these requirements. Therefore, labor force participation trends for them are informative. Interestingly, although rates for middle-aged married women have increased, those for middle-aged and older never-married women have dropped (Rix, 1984). For 45-54 year old never-married women, participation rates went from about 81 percent in 1960 to about 70 percent in 1980. For 55-64 year old never-married women, the figures were 67 percent in 1960, but only 55 percent in 1980.

Rix (1984), a researcher for the Congressional Caucus for Women's Issues, summarized the situation like this:

> Undoubtedly, unlike most of their married counterparts, these [never-married] women have had a long, perhaps life-long attachment to the labor force. Eligibility for Social Security, retired worker benefits and, in some cases for private pensions, coupled with a desire for leisure, may explain labor force trends that are similar to those of middle-aged and older men (p. 16).

If, as these data suggest, labor force attachment and associated factors (such as work attitudes and pension coverage) do make a difference in the timing of retirement for older women, then gathering descriptive information from younger women on these factors could be an important prerequisite to improved predictions about future retirement patterns. Thus, collecting information from younger, as well as older women, was an important part of the planned study.

What Is Known About The Topic

Despite the potential usefulness of information on women's retirement, our work during the early stages of scoping clearly indicated that information in this area is inadequate. Our

review, which appears as Appendix 1, focused on two topics: what is known about the similarities and differences between the sexes in their retirement decisions; and the impact of attitudinal factors (including not only commitment to work, but also job satisfaction, peer pressure, fear of inflation, and so on) on the retirement decision for both men and women.

As we describe in Appendix 1, we limited the literature review to studies that: 1. reported findings from primary research (rather than literature reviews or syntheses); 2. used the retirement decision or some approximation of it as a dependent variable; and 3. either compared the sexes or looked exclusively at women. (For the review of attitudinal factors influencing retirement, we also included studies of men only.)

Studies of Women's Retirement

Our literature review dramatically illustrated the point that current knowledge of women's retirement rests on very few, often inadequate, original data sources:

1. We found only 26 studies of women's retirement (or women's vs. men's retirement) that met the three criteria listed above. Eleven of the studies examined factors influencing women's retirement decisions, or approximations of that variable, such as planned retirement age or attitudes toward retirement. Fifteen additional studies examined differences in retirement decisions of men and women. These studies are listed in Exhibits 1 and 2, and are described more fully in Appendix 1.

2. Twelve of the these 26 studies reanalyzed the Retirement History Survey (George et al., 1984; Henretta and O'Rand, 1980; Honig, 1983a; O'Rand and Henretta, 1982; Sherman, 1974; Anderson et al., 1980; Clark and Johsnon, 1980; Clark et al., 1980; Hall and Johnson, 1980; Hanoch and Honig, 1983; Honig, 1983b; and Quinn, 1978). This is a longitudinal study, begun in 1969 and ended ten years later, which studied retirement for men and single women. Information from married women was collected only on a limited basis as part of their husbands' interviews, or later on, from the women who became widowed. But, because the initial RHS sample did not include married women as respondents in their own right, conclusions drawn from studies using the RHS can not be generalized to all women of retirement age. Thus, nearly half of the studies that we found on women's retirement excluded an important group: married women. (One of the studies in this group, George et al, 1984, also examined information from another data base which, while it included married women, used a small sample drawn from a limited population.)

3. Three of the studies reanalyzed the National Longitudinal Survey (Chirikos and Nestel, 1983; and Shaw, 1983 and 1984). This project began in 1966 with a group of 5000 women who were 30 to 40 years old and a group of 5000 men who were 45 to 59 years old. These people were interviewed every two to three years. Data for men were last collected in 1983 and for women in 1984. However, the National Longitudinal Survey includes only women younger than 62, and the currently available public use tapes include only women as old as 59. Therefore, the studies based on it used proxy measures of retirement, such as planned retirement age and exits from the labor force, which may not be good predictors of actual retirement status.

4. The remaining projects were a mixed lot. Some (for example, Johnson and Price-Bonham, 1980; Price-Bonham and Johnson, 1983) used very small samples. Some (such as, Atchley, 1983; and Barfield and Morgan, 1978) analyzed proxy variables such as retirement attitudes. Some (for example, Schmitt et al., 1979, and Chartock, undated) were based on limited populations. Others (such as Palmore, 1965, and Streib and Schneider, 1971) were dated and, therefore, possibly not relevant to present retirement-age women or those who may retire over the next decade or so.

In general, the research suggested that the retirement decision for men and women may be subject to some of the same influences. Demographic factors such as education and occupational status seem to figure in the decisions of both sexes, and both seem to be influenced by health, employment and financial factors. In contrast, the sexes seemed to differ on a variety of factors that may affect retirement decisions, especially work history (which tends to be more intermittent for women) and the effects of spousal characteristics (which work differently for women than for men). Again, however, these findings were based on a small set of studies, many of which used the same two data sets and, consequently, presented only limited information on married women (Retirement History studies) or analyzed planned retirement or labor force participation for middle-aged groups, rather than actual retirement status (National Longitudinal Survey studies).

Research on Attitudes and Retirement Decisions

We divided the research on attitudes and the retirement decision into four categories based on the dependent variable used. Some studies looked at attitudinal influences on retirement status (i.e., retired or not). Others looked at attitudinal relationships with planned retirement age (usually before, at, or after age 65). A third category correlated views about retirement (that is, is retirement anticipated as a negative or positive experience) with measures of job

Exhibit 1

Studies of Women's Retirement Status or Related Variables

Reference	Data Base	Population	N	Dependent Variable
Chirikos & Nestel (1983)	NLS	Women, 30–44 in 1967[a], nationwide	3167	Labor force participation
Shaw (1983)	NLS	Continously married women, 30–44 in 1967[a], nationwide	1131	Exit from the labor force
Shaw (1984)	NLS	Married women, working, seeking work, or intending to seek work, with retirement plans, 45-56 in 1979[a], nationwide	800+	Planned retirement age (<62, 62–64, 65+)
Henretta & O'Rand (1980)	RHS	Married women, employed, 58–63 in 1969[b], nationwide	5833[c]	Labor force participation
Honig (1983a)	RHS	Unmarried women, white, with previous SS earnings, 62–67 in 1973[a], nationwide	1270	Retirement status (full employment, partial retirement, full retirement)
O'Rand & Henretta (1982)	RHS	Women, unmarried in 1969, with sustained work attachment 1964–1968, 58–63 in 1969[b], nationwide	1399	Retirement status (retire <62 vs. later, retire 62–64 vs. later)
Sherman (1974)	RHS	Unmarried women, 58–63 in 1969[a], nationwide	Not specified	Labor force participation
Dowdall (1974)	Original	Married women, with ≥ 1 children, 15–64 in 1968–1969[a], Rhode Island	592	Employment status
Jaslow (1976)	Not specified	Women, 65 + in 1968[a], nationwide	2398	Employment status (currently employed, retired, never employed)

[a] Cross-sectional survey design; year is when data were collected.

[b] Longitudinal survey design; year is when data were first collected.

[c] The Retirement History Survey (RHS) excluded married women as respondents in the initial 1969 panel. Information about married women referred to in RHS studies comes from limited data about wives of respondents collected during the husbands' interviews, from respondents' widows, or from single women respondents who married in subsequent years.

Exhibit 1 (continued)

Reference	Data Base	Population	N	Dependent Variable
Johnson & Price-Bonham (1980)	Original	Married women, employed, 50+ (year of data collection not specified), one community	59	Acceptance of retirement stereotypes, resistance to retirement
Price-Bonham & Johnson (1983)	Original	Married women, employed, professionals and nonprofessionals, 55-63 (year of data collection not specified), one community	100	Attitude towards retirement

Exhibit 2

Studies Comparing Men and Women's Retirement Decisions or Related Variables

Reference	Data Base	Population	N	Dependent Variable
Anderson et al. (1980)	RHS	Husbands and wives, 58–63 in 1969[b], nationwide	600 couples (multiple observations per respondent)	Labor force participation
Clark & Johnson (1980)	RHS	Husbands and wives, 58–63 in 1969[a], nationwide	2657 couples	Labor force participation
Clark et al. (1980)	RHS	Husbands and wives, non-self-employed, 58–63 in 1969[b], nationwide	3312 couples	Labor force participation
George et al. (1984)	RHS, Duke Second Longitudinal study (DSLS)	Men, unmarried 58–63 in 1969[b] (RHS) Men, women 46–70 in 1969[b], in one community (DSLS)	1468 men, 377 women (RHS) 156 men, 79 women (DSLS)	Retirement status
Hall & Johnson (1980)	RHS	Married men and unmarried women, employed, 58–63 in 1969[a], nationwide	3557 men, 1054 women	Planned retirement age
Hanoch & Honig (1983)	RHS	Married men and unmarried women, 58–63 in 1969[b], nationwide	3130 men, 1359 women (multiple observations per respondent)	Labor force participation
Honig (1983b)	RHS	Married men and unmarried women, 58–63 in 1969[b], nationwide	Not specified (multiple observations per respondent)	Labor force participation
Quinn (1978)	RHS	Married men, and unmarried men and women, 58–63 in 1969[a], nationwide	5623 men, 2224 women	Labor force participation

[a] Cross-sectional survey design; year is when data were collected.

[b] Longitudinal survey design; year is when data were first collected.

Exhibit 2 (continued)

Reference	Data Base	Population	N	Dependent Variable
Atchley (1983)	Original	Men and women, employed, 50+ in 1975[a], one community	214 men, 142 women	Attitude toward retirement, plans to retire, planned retirement age
Barfield & Morgan (1978)	Original	Husbands and wives, unmarried men and women, 35-64 in 1976[a], nationwide	394 couples, 32 unmarried men, 60 unmarried women	Plans to retire early (<65 for husbands, unmarried men and women; <62 for wives)
Chartock (no date)	Original	Men and women, retired from the retail trades (ages, year of data collection not specified)	400	Retirement decision
Haug et al. (1982)	Original, in progress	Men and women, enrolled in HMO, 60-64 in 1983[a], Cleveland area	300 men, 500 women	Retirement status, intention to retire, adaption to retirement
Palmore (1965)	1963 Social Security Administration Survey of the Aged	Men and women, 62+ in 1963[a], nationwide	7701 men, 9660 women	Retirement status
Schmitt et al. (1979)	Original	Men and Women, retired and working Michigan civil servants, 55-65 (year of data collection not specified)	642	Retirement status
Streib & Schneider (1971)	Cornell Study of Occupational Retirement	Men and women, >63 in 1952[b], nationwide	1486 men, 483 women	Retirement timing

Exhibit 3

Studies of Attitudes and Retirement Status or Related Variables

Reference	Data Base	Population	N	Dependent Variable
		Studies Examining Retirement Status		
George et al. (1984)	RHS, Duke Second Longitudinal study (DSLS)	Men, unmarried 58-63 in 1969[b] (RHS) Men, women 46-70 in 1969[b], in one community (DSLS)	1468 men, 377 women (RHS) 156 men, 79 women (DSLS)	Retirement status
Jaslow (1976)	Not specified	Women, 65+ in 1968[a], nationwide	2398	Retirement status
Katona et al. (1969)	1966 Survey of Consumer Finances	Auto and agricultural implement workers, retired and working, 58-61 in 1969[a]	1123	Retirement status, planned retirement age
Parnes & Nestel (1975)	NLS	Men, 50-60 in 1971[a], nationwide	5020	Retirement status, expected retirement age
Pollman & Johnson (1974)	Original	Men, UAW machine operators, eligible for retirement, 60-65 (year of data collection not specified), midwest	700	Retirement status
Schmitt et al. (1979)	Original	Male and female Michigan civil servants, retired and working, 55-65 (year of data collection not specified)	642	Retirement status
Schmitt & McCune (1981)	Original	Male and female Michigan civil servants, retired and working, 55-70 (year of data collection not specified)	892	Retirement status

[a] Cross-sectional survey design; year is when data were collected.

[b] Longitudinal survey design; year is when data were first collected.

Exhibit 3 (continued)

Reference	Data Base	Population	N	Dependent Variable
		Studies Examining Planned Retirement Age		
Atchley (1983)	Original	Men and women, employed, 50+ in 1975[a], one community	214 men, 142 women	Attitude toward retirement, planned retirement age
Barfield & Morgan (1970)	Original	Family heads, 35-59 in 1966-67[a], non-retired auto workers, nationwide workers	1652 family heads, 646 auto workers	Planned retirement age (before, at or after 65)
McPherson & Guppy (1979)	Original	Men, employed, 55-64 (year of data collection not specified), one community	269	Planned retirement age (before, at or after 65)
Rose & Mogey (1972)	VA Normative Aging Study	Men, employed, veterans, Boston area (ages, year of data collection not specified)	2000	Preferred retirement age
		Studies Examining Attitudes Towards Retirement		
Fillenbaum (1971)	Original	Nonacademic university employees, 25+ (year of data collection not specified)	200	Attitudes toward retirement
Harris (1981)	Original	Men and women, 18+ in 1981[a], nationwide	3427	Attitudes toward retirement
Goudy et al, (1975)	Original	Men, employed, 50+ (year of data collection not specified), midwestern city	1922	Attitudes toward retirement
Johnson & Price-Bonham (1980)	Original	Married women, employed, 50+ (year of data collection not specified), one community	59	Attitudes toward retirement
Price-Bonham & Johnson (1983)	Original	Married women, employed, professional and non-professional, 55-63 (year of data collection not specified), one community	100	Attitudes toward retirement

Exhibit 3 (continued)

Reference	Data Base	Population	N	Dependent Variable
		Studies Examining Reasons for Retirement		
Evans & Ekerdt (no date)	VA Normative Aging Study	Men, recently retired, veterans, 47-76 in 1981[a], Boston area	70	Reasons for retirement
Messer (1969)	Original	U.S. Civil Service early retirees, <65 (year of data collection not specified)	3299	Reasons for retirement
Palmore (1965)	1963 Social Security Administration Survey of the Aged	Men and women, 62+ in 1963[a], nationwide	7701 men, 9660 women	Reasons for retirement
TIAA-CREF (1983)	Original	TIAA-CREF annuitants, 60-90 in 1982[a]	Not specified	Reasons for retirement
U.S. GAO (1982)	Original	Newly entitled SS beneficiaries in 1980[a], nationwide	1709	Reasons for applying for SS benefits (before vs. at or after 65)

satisfaction, life satisfaction, and commitment to the worker role. A fourth type of study, which asked people why they retired, was examined to see if attitudinal reasons were given.

As listed in Exhibit 3, we found 21 studies (some of which are the same as the women's studies referenced earlier) that fit one or more of these categories. In general, this work suggested that some attitudes (particularly job satisfaction, which had been studied most frequently) do influence the timing of retirement, and that others (specifically, approval of women's employment in general and the saliency of the work role for oneself) influence women's labor force participation patterns, and, consequently, their subsequent Social Security and pension coverage. Thus, the research lent support to our notion that it is important to include such attitudes in studies of the retirement decision.

But the research also hinted at the possibility that financial and employment factors may moderate the influence of attitudes and that attitudinal factors may be most important for people with options, that is, with sufficient income and in good enough health to make choices about their futures. For example, one study of female labor force participation (Dowdall, 1974)* found that attitudinal variables accounted for the greatest percentage of variance in equations for upper income groups for whom work is not an economic necessity.

Like studies of women's retirement, studies of the attitudinal influences on retirement decisions suffered from a number of important limitations:

1. Only seven of the 21 studies examined the retirement decision more or less directly, by examining attitudinal differences between retirees and workers. (These are listed in the first part of Exhibit 3.) The others used proxy variables, such as planned retirement age or views of what retirement is likely to be like, which may or may not reflect actual decisions.

2. Even those studies analyzing differences between retirees and workers relied on retrospective information for some attitudinal measures. This is because, with three exceptions (George et al., 1984; Parnes and Nestel, 1975; and Schmitt and McCune, 1981), all used cross-sectional designs, requiring retirees to remember how they felt about, for example, their jobs, just before retirement. The problem is that subsequent adjustment to retirement might have colored these memories.

---

*This study is not listed in Exhibit 3, since it does not focus on retirement per se.

3. The studies in general were based on limited populations, such as state civil servants, male auto workers, university employees, or veterans. Moreover, with some exceptions (Katona et al., 1969; Parnes and Nestel, 1975; Barfield and Morgan, 1970; Louis Harris and Associates, 1981; and U.S. GAO, 1982) many relied on non-probability samples.

4. Three of the twenty one studies (Schmitt and McCune, 1981, McPherson and Guppy, 1979; and George et al., 1984) included demographic, health, financial, and attitudinal variables together in a single explanatory equation. All were based on select populations, Michigan civil servants in one case, residents of a single urban community in another, and participants in a local health insurance plan in the third. Two of the studies examined actual retirement status, while the other used retirement plans as a proxy. In each case, adding attitudinal variables significantly improved the explanatory ability of the researchers' equations for at least some of the subgroups analyzed. Therefore, although the evidence was slim, it did suggest that attitudinal variables might well be included in analyses of the retirement decision.

In short, then, the field was ripe for new work documenting the types of factors that influenced retirement decisions for women as compared to men, with special emphasis on attitudinal factors that had been so inadequately studied even for men.

III. OBJECTIVES OF THE PLANNED STUDY

In this project, we did not intend to model the retirement decision in a formal sense by developing predictive or explanatory equations, but, rather, aimed to produce descriptive information, interesting in its own right, and appropriate for subsequent use in developing a model in a second PEMD project. We intended to address three questions:

1. What factors influence women in different age groups to retire? How are the factors influencing the decision for women in each age group similar or different from the factors influencing the decision for men in those age groups?

The factors that we were interested in are listed in Exhibit 4. We developed this list through the literature search described in Appendix 1 and additional searches done for the retirement issue area planning paper. Using these sources, we identified factors that had been demonstrated, in other studies, to be related to timing of retirement or some proxy measure, such as retirement plans.

We supplemented this activity through open-ended interviews with retirement counselors, recent retirees, and people facing retirement. These interviews, which are summarized in Appendix 2, seemed to confirm the importance of some of the factors on the list, particularly attitudinal and financial ones. Respondents told us that they decided to retire because of family pressures, the desire to leave a job "while still on top," and because of custom and peer pressures ("everyone seemed to retiring at that age"). We also found that pension and Social Security eligibility were important considerations for these respondents. Thus, the interviews lent some "real-life" credence to the discussion of factors that appeared in the literature.

Operationally, we planned to address the first study question by looking at each of the factors in Exhibit 4 for retired and working women and retired and working men in specific age groups. If we found differences between retired and working groups on these factors, we would infer that they were associated with the retirement decision. For example, if a larger proportion of retired than of working women reported that they had been eligible for a pension in their most recent year of work, we would infer that pension eligibility was one factor influencing the timing of retirement.

The age groups for which these analyses would be done were: 55-61, 62-64, 65, and 66-70. We chose these ranges to coincide with Social Security eligibility levels (that is, not eligible, eligible for reduced benefits, eligible for full benefits, and eligible for delayed retirement credits).

Exhibit 4

Items for Data Collection

Demographic

1. Sex
2. Age
3. Education
4. Race
5. Marital status
6. Change in marital status last 5 years

Financial

1. Pensions
   - Social Security eligibility: respondent, spouse
   - Social Security benefits[a] (actual or expected): respondent, spouse
   - Pension eligibility: respondent, spouse
   - Pension amounts: respondent, spouse
2. Other Income
   - Total family income
   - Respondent's wage
   - Spouse's wage
   - Nonwage income
3. Assets
   - Home Equity
   - Other (savings, stocks, bonds, IRAs, KEOGHs)
4. Liabilities
   - Amount owed: mortgage
   - Amount owed: other
   - Support of dependents (parents) or children)

Work related

1. Employment status: respondent, spouse
2. Labor force history
3. Occupational category
4. Early retirement, mandatory retirement in employer pension plan, flexible work options

Health

1. Functional health
2. Self-reported health

Attitudinal

1. Work role attitudes
   - Appropriateness of work role for women
   - Saliency of work, primary role perception
   - Job satisfaction
2. Subjective well being
   - Life satisfaction, morale
   - Locus of control
3. Attitude toward retirement[b]
   - Respondent
   - Spouse
4. Perception of economic market-place conditions
   - Inflation
   - Employer attitudes toward older workers

[a] Information to be calculated from wage and labor force participation data.

[b] Will include questions on planned retirement age.

2. What factors influence women's decisions to retire early? To retire late? How do the factors important in early/late retirement for women compare to the factors important in early/late retirement for men?

Operationally, answering these questions would involve comparing women who had retired early (before age 62) to those who had retired late (after age 65) on factors such as health at retirement, lifetime labor force participation patterns, Social Security and pension eligibility and other financial factors in our list, as well as on attitudinal factors such as appropriateness of work roles for women. Again, we would infer that differences between early and late retirees would mean that the factors were associated with the timing of retirement. Information from men would again be used for comparative purposes.

3. What are the characteristics of younger women and men on selected factors that seem to be important in the retirement decisions of older groups?

The key here is the term "selected factors." It made sense to get information only on a subset of items in the list, that is, on those items where we suspected that differences between younger and older women might exist which could influence future retirement patterns. A case in point, labor force participation, was described earlier. Based not only on the national labor force statistics but also on our own interviews with women facing retirement and recent retirees, our initial hypothesis was that women who enter the labor force relatively early and stay in continuously will behave more like men in their retirement decisions--that is, will retire earlier than women who enter the workforce later in life or whose participation is not continuous. If many women were showing high labor force attachment, it would suggest, we would argue, that in the future more women will retire early than at present.

Other factors where we expected intercohort differences that could alter future retirement patterns are:

a. Pension and Social Security coverage;

b. Saliency of work; views about appropriateness of work role for women;

c. Attitude toward retirement, including planned retirement age; and

d. (for analytic purposes) Demographic information, such as age, race, marital status, education, and occupation.

Much of the information we wanted for these younger groups is already available elsewhere. We had data on pension and Social Security coverage for women, on labor force attachment,

and on demographic characteristics. Less readily available is the attitudinal information listed in categories "b" and "c." Our need for this information, along with the desire to be able to relate these variables by collecting all the data from one group, justified including the younger age groups in the study we wanted to design.

IV. DESIGN ISSUES AND DECISIONS

When designing a project to address the three study questions, we faced the following major design issues:

1. In general, what kind of design would be necessary to permit inferences to be made about male/female differences in retirement decisions, and, more importantly, to permit inferences about older/younger women differences in factors affecting the retirement decision?

2. What sample sizes were needed for each of the age and sex groups of interest in the study to detect these differences with reasonable power?

3. What data collection alternatives were available which met our general design and sampling requirements?

4. What measurement problems would we need to solve in designing our data collection instruments?

Our attempts to address these issues and the decisions we made in each area are described below.

General Design Required to Detect Differences in Retirement Patterns

To make inferences about male/female and younger/older differences in factors affecting the retirement decision, we had several major design requirements. First, we wanted to generalize to the national population. Therefore, it was apparent to us that the study should use a survey, or an existing data base based on a survey, rather than case studies. Only a survey would allow us to make generalizations to the populations (women and men, 40-54 and 55 and older) in which we were interested.

Secondly, we needed, preferably, a longitudinal panel design, that is, one in which information was collected from the same respondents at several points in time. This design was preferable because it would allow us to disentangle cohort vs. age-related differences in attitudes, to determine if attitudes expressed by, say, a 40-49 year old woman were a function of when she was born (and thus should remain relatively stable), or a function of her age, and therefore likely to change with time. If attitudes affected the timing of retirement for older women, knowing their stability for younger groups would be important to make even descriptive statements about likely future retirement patterns.

A longitudinal design also would allow us to address the retrospective data problem facing the project. This problem refers to the fact that some of the information needed in this study (specifically data to address study questions 1 and 2) would be based on respondent memory. For example, we wanted to compare the health at time of retirement of women who retired at, say, age 62 to the health of 62-year-old women who chose to continue working. But, to get this information through a cross-sectional survey, we would be faced with asking a 70-year-old who retired when she was 62 to describe her health at retirement--that is, eight years earlier. A longitudinal design, with several data collection points, would be more likely to capture respondents who had recently retired, and it could collect, for example, health data, before retirement.

A third design requirement, particularly important if new data were to be collected, involved the method of data collection. We eliminated mail surveys because of the probability of low response rates, because we did not want respondents to consult with others for attitudinal items, and because of the difficulty of using open-ended questions. Unless we could find an on-going personal interview survey to which we could add items, we decided to use the phone method chiefly because of cost. We realized, however, that certain questions, particularly financial ones, would be difficult to ask over the phone unless an advance call to respondents were made. These preliminary contacts therefore became part of our general design requirements if we needed new data collection.

In short, then, the optimum, but still feasible, design necessary to detect differences in retirement decisions involved using an existing national longitudinal data base. If we could not find a suitable data base, and a new survey were necessary, we would collect data by phone, unless we could find an ongoing personal interview survey to which we could add questions. In the event of new data collection, we would need to precede the actual interviews with introductory contacts. Our analyses of factors where memory was likely to color answers (health and attitudes) would be limited to recent retirees, unfortunately decreasing the power of our tests for those variables. But the possibility of a longitudinal follow-up, increasing the eventual number of recent retirees in the data base, and allowing us to disentangle age and cohort effects, would be built into the project by including a few questions that would allow us to track respondents.

Sampling Requirements

Having decided the major parameters shaping the study, we then needed to determine minimum sample sizes needed to detect existing population differences with a given degree of certainty. As shown by the description of study questions, we wanted to include in the study information on all (or most) of the factors for women and men 55 and above. We would analyze the information by retirement status for narrow age groups in this population. For example, our analysis might involve comparing women who had or had not retired by age 55-61 on job attitudes, labor force patterns, pension eligibility, and the like. We also would compare women who had retired before age 62 to those retiring after age 65. For comparative purposes, all such analyses also would be done separately for men.

To meet these requirements, we needed the minimum sample sizes shown in the top part of Exhibit 5. With these sample sizes, we would have an 80 percent chance of detecting a difference in proportions of 10-20 percent or more between two groups, provided that we were also willing to take a 5 percent chance of falsely finding a non-existent difference. Or, in other words, the power of our test would be 80 percent at an alpha level of .05.

For younger groups, the sample sizes shown in the bottom portion of Exhibit 5 would result in a sampling error of 7 percent (at the 95 percent level of confidence) for women in each of the younger age groups, and a sampling error of 10 percent for men in each group. We felt that these errors were tolerable.

Adding the numbers for all subgroups, we determined that the minimum sample size needed for this survey would be 2200 cases.

Data Collection Alternatives

With our general design and sampling requirements in mind, we considered a number of data collection alternatives at various stages of scoping.

Early Decisions

The first was to reanalyze an existing longitudinal data base, an option that would have avoided all the costs of new data collection. We examined six likely candidates, evaluating each against standards for content and population coverage, design, and recency of data collection. However, as summarized in Exhibit 6 and described in Appendix 3, we found that no longitudinal data base could stand "as is" for secondary analysis. (Appendix 3 also describes a similar analysis we did for cross-sectional data bases, where we reached the same conclusion.)

Exhibit 5

Target Sample Sizes

Stratum I Ages 55-70	Women		Men	
	Retired	Working	Retired	Working
55-61	100	100	100	100
62-64	100	100	100	100
65	100	100	100	100
66-70	100	100	100	100
Subtotal	400	400	400	400

Stratum II Ages 40-54	Women	Men
40-49	200	100
50-54	200	100
Subtotal	400	200

Total N - 2200

Note: Sample sizes shown for stratum 1 will allow a difference in proportions of at least 10 percent (for more extreme proportions) or at least 20 percent (for proportions closer to one half) to be detected with 80 percent power. This is for a two-tailed test with alpha at .05. Source: Cohen, J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press, 1977.

Exhibit 6

Assessment of Extant Longitudinal Data Bases
for Use in Secondary Analysis

Data Base	Year of last Data Collection	Population Coverage	Content Coverage	Sample Size and Sampling
Retirement History Survey (RHS)	1979	Few married women	Adequate	11,200
National Longitudinal Survey (NLS)	1984	No women over 61	Adequate	47,300
Michigan Panel Survey of Income Dynamics (MPSID)	1984	Few married women	Adequate	4,900
Survey of Income and Program Participation (SIPP)	1984	Adequate	Omits attitudes	40,000
Current Population Survey (CPS)	1984	Adequate	Omits attitudes and health	60,000
Survey of Consumer Finances (SCF)	1983	Focus on household	Omits attitudes and health	3,800

In order of preference, our next option was to supplement an existing longitudinal data base, by collecting data from age groups not covered in the original. The National Longitudinal Survey looked good--at first. It covered the content we were interested in and apparently required only that we add to it survey data from older women. However, a closer look showed some other significant population gaps. We soon realized that the original panel design included only narrow age groups (30 to 40 year old women and 45 to 59 year old men). While we could watch these groups as they aged, we would not have current data on 50 to 59 year old men or 40 to 44 year old women, since panel members had reached that age many years ago. Forty to 44 year old men were also missing, because the young men's panel added in the late 70s had not yet reached middle age.

Having rejected the NLS, we looked for other alternatives. One option, which we quickly rejected, was the idea of using GAO staff as data collectors. Our experience on other projects suggested that GAO is not set up to do large scale national probability telephone surveys of the type envisioned for this project. We have no central interviewing facility (with a bank of phones and space for interviewing and coding operations). Further, while we did not cost the GAO alternative for this project, comparative cost estimates made for another PEMD project strongly suggested that, apart from the logistical problems involved, using GAO staff would be an expensive alternative to the options described below.

Remaining Options

These decisions left us with four remaining alternatives, which we considered in the later stages of scoping:

1. Adding questions to the Survey of Income and Program Participation (SIPP).

The SIPP is a personal interview panel study, carried out by the Census, that focuses on household finances and income support programs. But it also collects information on labor force participation, health (limited items) and demographics. The SIPP uses two panels of 20,000 households each. Each household is kept in the survey for 2-1/2 years during which it is interviewed nine times. What this option offered was the possibility of adding a 10 minute "topical module" to one wave of the survey which we would use to ask respondents the attitudinal questions that the SIPP lacks. However, adding questions required successful negotiations with the SIPP advisory board; our success was not guaranteed. Another drawback is that, although the SIPP uses a longitudinal design, its time frame is very short, and Census confidentiality procedures would preclude our reinterviewing respondents. We estimated that the SIPP option would cost about $150,000 for one wave of data collection from the approximately 8,000 respondents who fall in the 40-70 year old age range.

2. Adding questions to the Survey of Consumer Finances (SCF).

This personal interview survey of 3800 households, which represents a 71 percent response rate for the households initially contacted, was first done in 1983. A phone resurvey of the original respondents had been planned for this spring, but funding was not available. GAO's Human Resources Division was seriously considering using the survey for its pension study. Like the SIPP, the survey offered the possibility of adding questions to an ongoing effort. HRD would have added health items; we would have needed to add attitudinal items. Again, however, our success was not guaranteed, and a longitudinal design would have been difficult to carry out. Further, the sample had only about 1500 respondents in the desired age ranges, a number somewhat below our requirement. The SCF option was estimated to cost about $100,000 for a one-time data collection from these respondents.

3. Using an expired rotation group of the Current Population Survey (CPS).

This option offered the opportunity to buy, as a sampling frame, addresses from one or more groups (each with about 7500 respondents) that had been out of the CPS for at least two years. Since the frame consisted of addresses, going back several years meant that people originally interviewed for the CPS were not necessarily the ones to be contacted for our survey. However, Census staff estimated that about 75 percent of original respondents could be recontacted. In addition to the sampling frame, using an expired CPS rotation group also involved buying Census staff time to help with sampling specifications and questionnaire formatting, and Census interviewing and data reduction and cleaning facilities. Using Census staff for data collection was mandatory because confidentiality procedures prohibit the Census from releasing its sampling frame outside the agency. Census recommended that the survey be done by phone with personal interview follow-up of non-respondents and people without telephones. This option would require an introductory contact to prepare respondents for the type of information we would be requesting. Longitudinal follow-up would have been possible. We estimated that costs for the first wave of data collection from 2200 respondents, exclusive of pre-interview contacts, would be about $150,000.

4. Using the services of a private contractor to field a random-digit-dialing (RDD) telephone survey.

With this approach, the initial sampling frame consists of clusters of phone numbers which are first screened to identify those clusters likely to include large numbers of residences. Screening questions are then used to identify major target groups: that is, employed/retired women and men 55-70 and men and women 40-54. Using a private contractor to do an RDD survey

would require going through the RFP process, but is the only alternative (other than using GAO staff) that would not involve OMB clearance. This option also would require some pre-interview contacts with respondents. Interviewing costs would be about $280,000 for the sample of 3300 respondents necessary with this technique. (The reason for the larger sampling requirement is explained below.) Again, a longitudinal design would be possible with this option.

We judged each of the alternatives just described (the SIPP, the SCF, the CPS rotation group, and the private contractor) against a number of criteria. These are shown in Exhibit 7. Essentially, both the SIPP and the SCF could be eliminated from further consideration because, for both, the risk of not having our questionnaire items approved by the surveys screening committees was unacceptably high. Moreover, the SCF recently lost its funding, as we noted above.

On balance, using an expired CPS rotation group would seem to be the best choice. Census has demographic information associated with each address in the sample, and, as mentioned earlier, it estimated that in 75 percent of the cases the original respondents would still be living at these addresses. This demographic information would eliminate many of the screening phone calls necessary for random-digit-dialing. Moreover, as shown in the exhibit, Census staff indicated that it might be possible to merge work history information from Social Security Administration files with the CPS sampling frame data. However, using the CPS would require submitting questionnaire items for clearance by the Office of Management and Budget. Although, according to GAO's Office of the General Counsel, technically the Census Bureau would be asking for the clearance, GAO participation in the clearance process would be unprecedented.

If submitting to OMB clearance were judged to be infeasible, the only remaining data collection option would be to use the services of a private contactor for RDD. But this process is very labor intensive. Westat, a survey research firm experienced in random-digit-dialing, estimated that to reach one working woman in the 55-70 year old group would require screening calls to 15.4 households.

Since the process is so labor intensive, it precludes screening by finer age groups. Rather, the respondents in the major subgroups would be distributed by age according to their distribution in the population. For example, only about 3 percent of all 55-70 year old women in the labor force are 65. Therefore, using random-digit-dialing, we would expect that only about 12 of the 400 respondents planned for the group of 55-70 year old working women would be 65. Similarly, only about 13 percent of the working women in the 55-70 age group are 66-70, so that, using random-digit-dialing, we could expect only about

Exhibit 7

Comparison of Data Collection Alternatives

	Survey of Income and Program Participation (SIPP)	Survey of Consumer Finances (SCF)	Expired Current Population Survey Panel	Random-Digit-Dialing with Contractor
Risk of not getting questions accepted	Very High	High	Low	None
OMB clearance requirement	Yes	Yes	Yes	No
Content	-No attitudinal items -Social Security and pension amounts in one variable -Retirement status of some group is unclear -With exception of above, other data elements seem adequately covered	-No attitudinal items -No health items -Most demographic financial, and employment/work-related items are covered -Some data--e.g., participation in IRA's, assets, liabilities--only at household level	-Limited only only by what can be asked in phone interview (30-45 minutes) -Some possibility that CPS sample could be matched to SSA file for work history information	-Same as CPS, with exception of possibility of merging with SSA file.

Exhibit 7 Continued

	Survey of Income and Program Participation (SIPP)	Survey of Consumer Finances (SCF)	Expired Current Population Survey Panel	Random-Digit-Dialing with Contractor
Sampling considerations	-Estimate that 8,000 (40% of 20,000) would fall in 40-70 year old age range. -Since respondent provides data for all household members, self-reporting of attitudinal data would be a problem in some cases.	-Estimate that 1,500 would fall in 40-70 year old age range. -Therefore, sample size somewhat less than our target.	-Demographic information available for estimated 75% of addresses. -Would allow target N's for particular age X sex cells to be reached more easily.	-Screening for particular age X sex cells would greatly add to cost. -But, without such screening, sample of 400, 55-70 year old working women would yield only 12 65 year olds, and only 52 66-70 year olds. -Therefore, requires an estimated 3,300 interviews to reach desired targets in "small" cells.
Cost (very rough ball-park estimates)	$18.75 per case or $150,000 for 8,000 interviews. Includes cleaning.	$100,000 for 10 minutes with 500 respondents in desired age range.	$60-70 per case or $132,000 - $154,000 for 2,200 cases. Includes cleaning.	$85 per case or $280,000 for 3,300 interviews. Does not include cleaning.
Scheduling	Wave 7 to be fielded in Sept., 1985. Clean tape by Fall, 1986.	No exact date set. Overall funding uncertain.	If rough q're draft ready in March, clean tape by July, 1986. If we clean, few months earlier.	If rough q're draft ready in March, clean tape by April-May, 1986. Includes GAO cleaning and weighting.

52 respondents in that category. Therefore, using random-digit-dialing would require both that we collapse age groups (putting 65 year olds in with the 66-70 year old category), and that we increase sample sizes to obtain sufficient numbers of respondents in the smaller age categories. We calculated that, under random-digit-dialing, obtaining 100 cases in each of the fine age categories shown earlier in Exhibit 5 would require a sample of 675 each of retired and working women and retired and working men in the elderly stratum. Adding these numbers to the 600 for the younger age groups, means that the random-digit-dialing sample size requirement would be 3300. Using Westat's figure of $85 per case, a random-digit-dialing survey would cost more than $280,000 for data collection alone. This figure is an underestimate, since it does not include the pre-interview contacts that would be necessary with this option, nor does it include funds for data cleaning. Including these costs would raise the estimate to perhaps $350,000 for one wave of data collection.

Issues In Instrument Design

While exploring data collection options, we also examined some of the problems inherent in developing a data collection instrument that would yield reliable and sensitive measures of the factors potentially affecting the retirement decision. Our first problem was how to collect accurate financial and employment history information through a phone survey. The solution was threefold. We planned for pre-interview contacts (as mentioned in the discussion above); we planned to explore the possibility of merging SSA information on work history with the survey data, which we might have been able to do with the CPS data collection option; and we culled the initial list of data collection factors to only those items which we thought reasonable to collect through a survey. To illustrate this last point, although our list of factors had included both pension eligibility and pension amount, we intended to rely more heavily on the former, since we doubted that working elderly respondents would be able to report expected pension amounts accurately, This was just a temporary solution, however. If we intended to use the data to model the retirement decision, we would eventually need to calculate estimated pension amounts using labor force and wage data.

A second measurement problem was where to find reliable attitudinal items that would differentiate between men and women and between age groups. (Apart from the problems mentioned above, we were not worried about finding demographic, financial, and health items, since these had been used--some repeatedly--in other retirement studies.) A promising start for attitudinal items was the Roper Center, an archive for public opinion data. The center's computer retrieval system contains banks of items on various topics used in national surveys fielded since 1982. Roper estimated that a search of the base for attitudinal items in our

areas of interest that had been used in national probability phone surveys could be done for $450. The printout would display results by sex, by age, and, where available, by retirement status. Thus, we could have used it as one source of attitudinal items that discriminate between these different groups.

While we would have found that source helpful, it also would have been advisable to conduct some additional open-ended in-depth interviews at the beginning stages of instrument design. These would have been useful to help further identify appropriate attitudinal factors to be explored in the survey, as well as response choices for the closed-ended attitudinal items to be used. Once a structured instrument had been drafted, further pretesting would have been necessary before full-scale implementation. These instrument design activities would have added to the cost and time requirements of the project.

V. SUMMARY AND CONCLUSIONS

We have described four major design issues addressed during the scoping phase of the women's retirement project and summarized the approaches considered and the decisions made in each area. Our recommendation not to go further with this project is based on our findings for these four issues:

1. The optimum design will require longitudinal data collection to disentangle changes in attitudes due to cohort from changes due to age. A cross-sectional design would be a clear-cut compromise based on cost, which we can justify in our own minds only by considering it as the first wave of a longitudinal effort.

2. Sample size requirements are large, ranging, depending on the option, from 2200 to 3300 cases.

3. Each of the possible data collection options presents its own hurdles. No existing data base can be used; GAO does not have in-house capability for doing a survey of this magnitude using the sampling procedure (RDD) that would be required; using Census facilities will involve the unprecedented step of submitting to OMB clearance; and the private contractor route will involve an RFP (not frequently done at GAO) and a very large financial commitment.

4. While we began to address the problem of obtaining financial information over the phone, we do not have confidence that accurate information could be obtained. Further, since attitudinal measures are not generally included in other national retirement studies, these would have to be developed from other sources, including a series of in-depth open-ended interviews. This activity, as well as pretests of the structured instrument, would have added to the costs and time requirements of the study, while not necessarily guaranteeing that our measures would be good ones.

Finally, as noted at the outset, the significant expenditure necessary to carry out "Wave I" of this project would buy essentially descriptive data. Using the information to improve forecasts of retirement trends, the ultimate bottom-line of this effort, would require not only implementing the longitudinal feature of the survey, but also investing additional staff time for model-building analytic work (for example, constructing sets of predictor variables including, but not limited to the measures of Social Security and pension wealth, and testing the necessity of separate models for men and women.)

In sum, although the women's retirement project would address an important problem, it would do so with some risk, at great cost, and with no guarantee of success. In our judgment, therefore, the total picture does not justify going further with this effort.

REFERENCES

Anderson, K., R. Clark, and T. Johnson. "Retirement in Dual-Career Families." In R. Clark (ed.), Retirement Policy in an Aging Society. Durham, N.C.: Duke University Press, 1980.

Atchley R. "The Process of Retirement: Comparing Women and Men." In M. Szinovacz (ed.), Women's Retirement: Policy Implications of Recent Research. Beverly Hills, Calif.: Sage Publications, 1983.

Barfield, R., and J. Morgan. Early Retirement: The Decision and the Experience and a Second Look. Ann Arbor, Mich.: University of Michigan, Institute for Social Research, 1970.

-----. "Trends in Planned Early Retirement". The Gerontologist, 18:1 (1978), 13-18.

Chartock, P. Factors in the Retirement Decision: Comparisons between Men and Women. New York: Hunter-Brookdale Center on Aging, no date.

Chelimsky, E. "Data from the Retirement History Survey." Letter to Director, Bureau of the Census, July 16, 1982.

Chirikos, T., and G. Nestel. "Economic Consequences of Poor Health in Mature Women." In L. Shaw (ed.), Unplanned Careers: Working Lives of Middle-Aged Women. Lexington, Mass.: Lexington Books, 1983.

Clark, R., and D. Barker. Reversing the Trend toward Early Retirement. Washington, D.C.: American Enterprise Institute for Public Policy Research, 1981.

-----. and T. Johnson. Retirement in the Dual Career Family. Final Report, SSA Grant No. 10-P-90543-4-02, 1980.

-----. and T. Johnson, and A. McDermed. "Allocation of Time and Resources by Married Couples Approaching Retirement." Social Security Bulletin, 43:4 (1980), 3-16.

Davidson, J. "Employment Concerns of Older Women," Working Paper, National Policy Center on Women and Aging, University of Maryland, College Park, Md., 1983.

Dowdall, J. "Structural and Attitudinal Factors Associated with Female Labor Force Participation." Social Science Quarterly, 55 (1974), 121-30.

Evans, L., and D. Ekerdt. "Retirees' Accounts of the Retirement Decision." VA Outpatient Clinic, Boston, Mass., no date.

Fillenbaum, G. "On the Relation Between Attitude to Work and Attitude to Retirement." Journal of Gerontology, 26:2 (1971), 244-48.

George, L. Fillenbaum, G. and E. Palmore. "Sex differences in the antecedents and consequences of retirement." Journal of Gerontology, 39-3 (1984), 364-371.

Goudy, W., E. Powers, and P. Keith. "Work and Retirement: A Test of Attitudinal Relationships." Journal of Gerontology, 30:2 (1975), 193-98.

Gratton, B., and M. Haug. "Decision and Adaptation: Research on Female Retirement." Research on Aging, 5:1 (1983), 59-76.

Hall, A., and T. Johnson. "The Determinants of Planned Retirement Age." Industrial and Labor Relations Review, 33:2 (1980), 241-54.

Hanoch, G., and M. Honig. "Retirement, Wages, and Labor Supply of the Elderly." Journal of Labor Economics, 1:2 (1983), 131-51.

Haug, M., B. Gratton, and D. Cogan. Health and Women's Retirement Decisions and Adaptions. Grant proposal to NIA, 1982.

Henretta, J., and A. O'Rand. "Labor Force Participation of Older Married Women." Social Security Bulletin, 43:8 (1980), 40-46.

Honig, M. "Partial Retirement in the Labor Market Behavior of Older Women." Hunter College, New York, N.Y., 1983. (a)

-----. "Retirement, Wages, and Labor Supply of the Elderly: Male-Female Comparisons." Hunter College, New York, N.Y., 1983. (b)

Jaslow, P. "Employment, Retirement, and Morale among Older Women." Journal of Gerontology, 31:2 (1976), 212-18.

Johnson, C., and S. Price-Bonham. "Women and Retirement: A Study and Implications." Family Relations, 29 (1980), 380-85.

Katona, G., J. Morgan, and R. Barfield. "Retirement in Prospect and Retrospect." In W. Donahue et al. (eds.), Trends in Early Retirement. Ann Arbor, Mich.: University of Michigan, 1969.

Light, P. "Social Security and the Politics of Assumptions." Paper presented at the meeting of the American Political Science Association, Chicago, Illinois, August 30-September 3, 1983

Louis Harris and Associates. Aging in the Eighties: America in Transition. Washington, D.C.: National Council on the Aging, 1981.

Macke, A., P. Hudis, and D. Larrick. "Sex-Role Attitudes and Employment Among Women: A Dynamic Model of Change and Continuity." Center for Human Resource Research, Ohio State University, Columbus, Ohio, 1979.

McPherson, B., and N. Guppy. "Pre-Retirement Life-Style and the Degree of Planning for Retirement." Journal of Gerontology, 34:2 (1979), 254-63.

Messer, E. "Thirty-eight Years Is a Plenty." In W. Donahue et al. (eds.), Trends in Early Retirement. Ann Arbor, Mich.: University of Michigan, 1969.

Morrison, M. "The Aging of the U.S. Population: Human Resource Implications." Monthly Labor Review, (1983),

Older Women: The Economics of Aging. Washington, D.C.: The Women's Studies Program and Policy Center, George Washington University, 1980.

Oppenheimer, V. "The Female Labor Force in the United States: Demographic and Economic Factors Governing Its Growth and Changing Composition." Berkeley, California: University of California, 1970.

O'Rand, A., and J. Henretta. "Delayed Career Entry, Industry Pension Structure, and Early Retirement in a Cohort of Unmarried Women." American Sociological Review, 47 (1982), 365-73.

Palmore, E. "Differences in the Retirement Patterns of Men and Women." The Gerontologist, 15:1 (1965), 4-8.

Parnes, H., and G. Nestel. "Early Retirement." In H. Parnes et al. (eds.), The Pre-Retirement Years. Washington, D.C.: Government Printing Office, 1975.

Pollman, A., and A. Johnson. "Resistance to Change, Early Retirement and Managerial Decisions." Industrial Gerontology, 1:1 (1974), 33-41.

Price-Bonham, S., and C. Johnson. "Attitudes Toward Retirement: A Comparison of Professional and Nonprofessional Married Women." In M. Szinovacz (ed.), Women's Retirement: Policy Implications of Recent Research. Beverly Hills, Calif.: Sage Publications, 1983.

Quinn, J. The Early Retirement Decision: Evidence from the 1969 Retirement History Study. Washington, D.C.: Health, Education, and Welfare, 1978.

Rix, S. Older Women: The Economics of Aging. Washington, D.C.: Women's Research and Education Institute of the Congressional Caucus for Women's Issues, 1984.

Rose, C., and J. Mogey. "Aging and Preference for Later Retirement." Aging and Human Development, 3 (1972), 45-62.

Sandell, S. Women's Increased Lifetime Labor Supply: Implications for Sound Security Policy. Washington, D.C.: National Commission for Employment Policy, 1983.

Schmitt, N., et. al. "Comparison of Early Retirees and Nonretirees." Personnel Psychology, 32 (1979), 327-40.

-----, and T. McCune. "The Relationship between Job Attitudes and the Decision to Retire." Academy of Management Journal, 24 (1981), 795-802.

Shaw, L. "Causes of Irregular Employment Patterns." In L. Shaw (ed.), Unplanned Careers: The Working Lives of Middle-Aged Women. Lexington, Mass.: Lexington Books, 1983.

-----. "Retirement Plans of Middle-Aged Married Women." The Gerontologist, 24:2 (1984), 154-59. (a)

-----. "Employment Problems of the Older Woman." Paper presented at the Women's Research and Education Institute Symposium on Low-Wage Work: Its Causes and Consequences, Washington, D.C., 1984. (b)

Sherman, S. "Labor-Force Status of Nonmarried Women on the Threshold of Retirement." Social Security Bulletin, 37 (1974).

Statham, A., and P. Rhoton. "Attitudes toward Women Working: Changes over Time and Implications for the Labor-force Behaviors of Husbands and Wives." In L. Shaw (ed.), Unplanned Careers: The Working Lives of Middle-Aged Women. Lexington, Mass.: Lexington Books, 1983.

Statistical Abstract of the United States: 1984. Washington, D.C.: U.S. Department of Commerce, 1984.

Streib, G., and C. Schneider. Retirement in American Society: Impact and Process. Ithaca, N.Y.: Cornell University Press, 1971.

Szinovacz, M. "Introduction: Research on Women's Retirement." In M. Szinovacz (ed.), Women's Retirement: Policy Implications of Recent Research. Beverly Hills, Calif: Sage Publications, 1983.

TIAA-CREF (Teachers Insurance and Annuity Association-College Retirement Equities Fund). "A New Look at Retired Participants." The Participant, (1983), 1-5.

U.S. GAO (United States General Accounting Office). Demographic and Economic Characteristics of Social Security Retiree Families, HRD-82-131, Washington, D.C., 1982.

Waite, L. "Projecting Female Labor Force Participation From Sex-Role Attitudes." Social Science Research, 7 (1978), 299-317.

APPENDIX I: RELATED RESEARCH

This review covers the two topics of concern in the women's retirement project. It describes what is known about the similarities and differences between the sexes in their retirement decisions, and summarizes the impact of attitudinal factors for both men and women.

What We Know About Women's Retirement

Until recently, women were largely neglected in the whole broad area of retirement research, not to mention the more narrow area of the retirement decision. This neglect is illustrated by Szinovacz's (1983)[1] analysis of the 121 retirement reseach papers presented at annual meetings of the American Gerontological Association between 1970 and 1981. Based on her review, Szinovacz concluded that, prior to 1975, research concerned specifically with female retirement was practically nonexistent. In another recent review, Gratton and Haug (1983) confirmed this conclusion at least partially. They stated that, while research on women's adaptation to retirement has been conducted for more that a decade, studies of the retirement decision of women remain rare.

Our own findings support these statements. We reviewed studies that: 1) reported findings from primary research, rather than literature reviews or syntheses; 2) used the retirement decision or some approximation of it as a dependent variable; and 3) either compared the sexes or else focused exclusively on women. (For the review of attitudinal factors described later in this document, we also included studies that looked only at men.)

The procedures used to locate these references were wide-ranging. GAO librarians conducted searches of computerized data bases, covering the work of private and academic institutions, as well as government organizations (including GAO and its sister agencies). In addition, we carried out a separate search of the files of the library at the American Association of Retired Persons. The bibliographies of sources (including literature reviews) uncovered by these activities also were tapped, as were the work papers from our previous project developing PEMD's retirement issue area. Finally, we contacted individuals with interests in this area, both in and out of government. Because of the comprehensiveness of these procedures, we feel confident that our review includes the major studies done in the area of women's retirement decisions.

Yet, despite this wide net, we located relatively few studies meeting our criteria. The studies are listed in Exhibits A and B, and described in detail in the project descriptions at the end of this paper. Eleven of them look mainly at factors

---

[1]References in this appendix are listed on pages 163-167.

Exhibit A

Studies of Women's Retirement Status or Related Variables

Reference	Data Base	Population	N	Dependent Variable
Chirikos & Nestel (1983)	NLS	Women, 30–44 in 1967[a], nationwide	3167	Labor force participation
Shaw (1983)	NLS	Continously married women, 30–44 in 1967[a], nationwide	1131	Exit from the labor force
Shaw (1984)	NLS	Married women, working, seeking work, or intending to seek work, with retirement plans, 45–56 in 1979[a], nationwide	800+	Planned retirement age (<62, 62–64, 65+)
Henretta & O'Rand (1980)	RHS	Married women, employed, 58–63 in 1969[b], nationwide	5833[c]	Labor force participation
Honig (1983a)	RHS	Unmarried women, white, with previous SS earnings, 62–67 in 1973[a], nationwide	1270	Retirement status (full employment, partial retirement, full retirement)
O'Rand & Henretta (1982)	RHS	Women, unmarried in 1969, with sustained work attachment 1964–1968, 58–63 in 1969[b], nationwide	1399	Retirement status (retire <62 vs. later, retire 62–64 vs. later)
Sherman (1974)	RHS	Unmarried women, 58–63 in 1969[a], nationwide	Not specified	Labor force participation
Dowdall (1974)	Original	Married women, with ≥ 1 children, 15–64 in 1968–1969[a], Rhode Island	592	Employment status
Jaslow (1976)	Not specified	Women, 65 + in 1968[a], nationwide	2398	Employment status (currently employed, retired, never employed)

[a] Cross-sectional survey design; year is when data were collected.

[b] Longitudinal survey design; year is when data were first collected.

[c] The Retirement History Survey (RHS) excluded married women as respondents in the initial 1969 panel. Information about married women referred to in RHS studies comes from limited data about wives of respondents collected during the husbands' interviews, from respondents' widows, or from single women respondents who married in subsequent years.

Exhibit A (continued)

Reference	Data Base	Population	N	Dependent Variable
Johnson & Price-Bonham (1980)	Original	Married women, employed, 50+ (year of data collection not specified), one community	59	Acceptance of retirement stereotypes, resistance to retirement
Price-Bonham & Johnson (1983)	Original	Married women, employed, professionals and nonprofessionals, 55–63 (year of data collection not specified), one community	100	Attitude towards retirement

Exhibit B

Studies Comparing Men and Women's Retirement Decisions or Related Variables

Reference	Data Base	Population	N	Dependent Variable
Anderson et al. (1980)	RHS	Husbands and wives, 58–63 in 1969[b], nationwide	600 couples (multiple observations per respondent)	Labor force participation
Clark & Johnson (1980)	RHS	Husbands and wives, 58–63 in 1969[a], nationwide	2657 couples	Labor force participation
Clark et al. (1980)	RHS	Husbands and wives, non-self-employed, 58–63 in 1969[b], nationwide	3312 couples	Labor force participation
George et al. (1984)	RHS, Duke Second Longitudinal study (DSLS)	Men, unmarried 58–63 in 1969[b] (RHS) Men, women 46–70 in 1969[b], in one community (DSLS)	1468 men, 377 women (RHS) 156 men, 79 women (DSLS)	Retirement status
Hall & Johnson (1980)	RHS	Married men and unmarried women, employed, 58–63 in 1969[a], nationwide	3557 men, 1054 women	Planned retirement age
Hanoch & Honig (1983)	RHS	Married men and unmarried women, 58–63 in 1969[b], nationwide	3130 men, 1359 women (multiple observations per respondent)	Labor force participation
Honig (1983b)	RHS	Married men and unmarried women, 58–63 in 1969[b], nationwide	Not specified (multiple observations per respondent)	Labor force participation
Quinn (1978)	RHS	Married men, and unmarried men and women, 58–63 in 1969[a], nationwide	5623 men, 2224 women	Labor force participation

[a] Cross-sectional survey design; year is when data were collected.

[b] Longitudinal survey design; year is when data were first collected.

Exhibit B (continued)

Reference	Data Base	Population	N	Dependent Variable
Atchley (1983)	Original	Men and women, employed, 50+ in 1975[a], one community	214 men, 142 women	Attitude toward retirement, plans to retire, planned retirement age
Barfield & Morgan (1978)	Original	Husbands and wives, unmarried men and women, 35–64 in 1976[a], nationwide	394 couples, 32 unmarried men, 60 unmarried women	Plans to retire early (<65 for husbands, unmarried men and women; <62 for wives)
Chartock (no date)	Original	Men and women, retired from the retail trades (ages, year of data collection not specified)	400	Retirement decision
Haug et al. (1982)	Original, in progress	Men and women, enrolled in HMO, 60–64 in 1983[a], Cleveland area	300 men, 500 women	Retirement status, intention to retire, adaption to retirement
Palmore (1965)	1963 Social Security Administration Survey of the Aged	Men and women, 62+ in 1963[a], nationwide	7701 men, 9660 women	Retirement status
Schmitt et al. (1979)	Original	Men and Women, retired and working Michigan civil servants, 55–65 (year of data collection not specified)	642	Retirement status
Streib & Schneider (1971)	Cornell Study of Occupational Retirement	Men and women, >63 in 1952[b], nationwide	1486 men, 483 women	Retirement timing

affecting women's retirement decisions (or approximations of that variable, such as planned retirement age or attitudes toward retirement). An additional 15 studies examine differences in retirement decisions of men and women.

These exhibits dramatically illustrate that current knowledge of women's retirement is based on a small number of original data bases. Twelve of these studies (Sherman, 1974; Quinn, 1978; Anderson et al., 1980; Hall and Johnson, 1980; Clark and Johnson, 1980; O'Rand and Henretta, 1982; Henretta and O'Rand, 1980; Clark et al., 1980; Hanoch and Honig, 1983; Honig, 1983a, 1983b; and George et al., 1984) reanalyze the Retirement History Survey. Three studies (Chirikos and Nestel, 1983; Shaw, 1983; and Shaw, 1984a) reanalyze the National Longitudinal Survey. Eleven are based on other sources, most of which are very limited as described below.

Limitations of the Research Done to Date

The fact that this body of knowledge rests on so few data sources has a number of implications:

1. We believe that the Retirement History Survey (RHS) should not be used to study women's retirement, mainly because it excluded married women as respondents in the initial 1969 panel. Information about these women was collected only on a limited basis as part of their husbands' interviews. or, later on, from the women who became widowed. In addition, single women respondents who married in subsequent years were kept in the sample. But, because the initial RHS sample did not include married women as respondents in their own right, conclusions drawn from studies using the RHS cannot be generalized to all women of retirement age. Further, RHS respondents were last surveyed in 1979; thus, the information is becoming dated. Finally, GAO's prior experience with the RHS led it to abandon a project because of the data base had high item non-response, unrealistic extreme values, and internal response inconsistencies (Chelimsky, 1982).

2. Since the National Longitudinal Survey (NLS) includes only women younger than 62, the studies based on it use proxy measures of retirement, such as planned retirement age and exits from the labor force. But planned retirement age is not the same as actual retirement status, and exits from the labor force--particularly for middle-aged women--may be indicative of irregular lifetime employment patterns rather than of retirement.

3. The studies using other data sources are a mixed lot. Some (for example, Johnson and Price-Bonham, 1980; Price-Bonham and Johnson, 1983) use very small samples. Some (for example, Atchley, 1983; Barfield and Morgan, 1978) analyze proxy variables, such as retirement attitudes. Some (for example, Schmitt et. al, 1979; Chartock, no date) are based

on limited populations. Others (such as Palmore, 1965) are dated.

Given all of these limitations, the findings of this research, which are summarized below, must be interpreted as tentative.

Is the Retirement Decision for Women Different than the Decision for Men?

In contrast to the situation for women, a great deal of research about men's retirement decisions already has been done. Clark and Barker's (1981) summary of this literature concluded that among the important factors influencing men's retirement are pension characteristics (eligibility for and size of pension and social security benefits); other financial variables (such as wages and assets); demographic characteristics (for example, age, number of dependents, and spouse's characteristics); macro-economic conditions (such as unemployment); and health. (It is interesting to note that Clark and Barker do not mention attitudinal variables in their review--an omission which, as illustrated later on, is characteristic of much multi-factor research on the retirement decision.)

Despite a recent finding to the contrary (George et al., 1984), many researchers have found that some factors that are important for men, also influence women. For example, the health of a worker, male or female, has generally been found to affect the retirement decision (Hall and Johnson, 1980; Hanoch and Honig, 1983; Sherman, 1974; O'Rand and Henretta, 1982; Quinn, 1978; Chirikos and Nestel, 1983; Honig, 1983a). All other things being equal, persons at or near retirement age tend to retire if they are in poor health or disabled. There is less agreement about the influence of husbands' health limitations on wives' decisions. While Henretta and O'Rand (1980) found that women whose husbands suffer health limitations are more likely to leave the labor force, Shaw (1984a) found that they were less likely to plan early retirements. Some researchers (for example, Gratton and Haug, 1983) question the validity of the health measures typically used in studies. More often than not, the measures are based on respondents' reports of their perceived health as opposed to more objective measures such as medical records and reports of specific functional limitations.

Demographic variables also appear to play a part in both men and women's retirement decision-making. A number of researchers have found that men and women who are more educated tend to retire later than those with less education (Shaw, 1984a; Honig, 1983a; Hall and Johnson, 1980; Hanoch and Honig, 1983; Sherman, 1974; O'Rand and Henretta, 1982). Based on comparisons of male and female retirees and workers, high occupational status is associated with later retirement for both sexes (O'Rand and Henretta, 1982; Streib and Schneider, 1971). However, these findings do not hold for planned retirement age (Atchley, 1983),

again pointing to the fact that retirement plans and actual retirement status are not the same.

Marital status also seems to affect retirement decisions, especially for women. In particular, single women--including those who are divorced, separated, or widowed--tend to remain in the labor force longer than women who are married (Sherman, 1974; Chirikos and Nestel, 1983; Streib and Schneider, 1971). Among married couples, having an employed spouse increases the likelihood for both men and women to remain in the labor force (Clark and Johnson, 1980; Shaw, 1984a). This finding, which suggests that married couples make their retirement decisions jointly, has important implications for any study of women's retirement, as discussed below.

The relationships between the retirement decision and some employment and financial factors also appear similar for men and women, though the evidence is more limited. Workers who are self-employed plan to and actually do remain in the labor force longer than workers who are not (Hall and Johnson, 1980; Hanoch and Honig, 1983), perhaps because they enjoy more flexibility in arranging their work lives. Workers--both male and female--are more likely to plan to retire and to actually retire if they are eligible for Social Security (Quinn, 1978; Hall and Johnson, 1980) and private pension benefits (Hall and Johnson, 1980), although some conflicting findings and differences between men and women (discussed below) exist. Non-wage income and total assets also affect retirement decisions of both sexes, with higher income and assets generally associated with lower labor force participation (Clark et al., 1980; Clark and Johnson, 1980; Quinn, 1978; Hanoch and Honig, 1983).

The relationship for other financial variables is hazier. For example, Clark et al. (1980) and Clark and Johnson (1980) have found that higher wages are associated with increased labor force participation among 58-63 year old married couples. Other studies, using the same data base (the RHS), have produced results that hold for wives, but not their husbands (for example, Anderson et al., 1980; and Quinn, 1978). Finally, Hall and Johnson's (1980) study of planned retirement ages--again using the same data base--seems to indicate that both men and women with high wages plan earlier retirements than others.

Perhaps most significant among the factors that may distinguish men and women in their retirement decision-making is work history (Shaw, 1984b). Women show much greater variability in their lifetime work patterns than do men. While men tend to work continuously from the time they enter the labor force until they retire, women often have interrupted work histories, usually as a result of childbearing and child care responsibilities. Shaw (1984b) claims that by the time they are 45 years old, men and women, on the average, have worked 25 and 11 years, respectively.

These figures are important because work history affects a variety of factors that may influence retirement decisions. One of the most important is eligibility for retirement benefits. Because of their intermittent work patterns, as well as their concentration in industries where pension plans are not offered or do not include flexible vesting and portability provisions, fewer women than men are covered by pension plans (O'Rand and Henretta, 1982). Also because of work patterns, women receive lower monthly benefits from Social Security, which is often their sole source of support (Older Women: The Economics of Aging, 1980).

Since financial considerations seem to be important in the retirement decisions of both sexes, older women's relatively poorer economic status may be one influence keeping them in the labor force. This may be especially true for single women, who are most dependent on themselves for income. In 1980, 47 percent of widowed, 63 percent of divorced, and 43 percent of separated 55-64 year old women were in the labor force; the comparable figure for married women was only 37 percent (Davidson, 1983).

In addition to differences in labor force participation patterns, other potential sources of differences between men and women's retirement decisions center around asymmetrical relationships involving spouses' characteristics. That is, a wife influences her husband differently than she is influenced by her husband. In particular, the effect of workers' eligibility for Social Security or pension benefits on spouses' retirement decisions seems to depend on workers' gender. For instance, Anderson and associates (1980) found that wives' eligibility for Social Security benefits makes it more likely that husbands will retire but husbands' eligibility does not affect their wives' decisions.

Findings of asymmetrical relationships such as these reinforce Gratton and Haug's (1983) point that it is very important to consider gender and marital status in attempts to understand retirement decision-making. They argue that the marital unit should be treated as the unit of analysis rather than the individual, whether worker or spouse. Other researchers (Anderson et al. 1980; Clark and Johnnson, 1980) also stress the importance of considering spousal characteristics when analyzing individual labor force participation.

Final points of difference in men and women's retirement decisions stem from the effects of a few isolated financial and demographic variables. For example, some studies suggest that Social Security and pension eligiblity may affect men and women differently (Clark et al., 1980; Hanoch and Honig, 1983; Quinn, 1978). Also, as mentioned earlier, some findings indicate that women, but not men, with higher wages are more likely to remain in the labor force (Anderson et al., 1980; Quinn, 1978). According to some research (Clark et al., 1980; Clark and Johnson, 1980; Quinn, 1978), having dependent children increases

the likelihood of labor force participation but only for men. However, other research (Henretta and O'Rand, 1980; Sherman, 1974) has produced contradictory findings. Finally, although the findings are again inconsistent with other studies (for example, Clark et al., 1980; Sherman, 1974), some researchers have found that women, but not men, are more likely to remain in the labor force as they age (Anderson et al, 1980; Clark and Johnson, 1980). It may be that, because of their discontinuous work histories, women have to work longer to qualify for pension benefits.

Conclusions Suggested by the Research on Women's Retirement

What can we conclude from the research done so far? A look at all the data suggests that, in many ways, the retirement decision for men and women is subject to the same influences. Demographic factors such as education and occupational status appear to play a part in the decisions of both sexes, and both seem to be influenced by health, employment, and financial factors. On the other hand, the sexes seem to differ on a variety of factors that may affect the retirement decision, especially work history and the effect of spouses' characteristics. Again, however, these conclusions are based on a small set of studies, many of which use the same two data bases and, consequently, present only limited information on married women (RHS studies), or analyze planned retirement or labor force participation for middle-aged groups, rather than actual retirement status (NLS studies).

Further, the bulk of studies done up to now, both those examining differences between the sexes and those looking at them separately, stress demographic, health, and, particularly, financial and economic variables. Relying on these factors may not capture retirement decision-making adequately. At a minimum, we can see that the amount of variance accounted for in many of these studies is modest. For example, for the studies in Exhibits A and B that predict women's retirement decisions, the amounts range from 1 to 36 percent (not including studies where the analysis was inappropriate); the median is roughly 10 percent. (See the project descriptions for the details by study.)

Why do these studies do so poorly in accounting for variance? One possibility is that they emphasize factors that limit an individual's options. For instance, poor health or a severe disability may force a person to retire, just as insufficient assets may force continued employment.

However, these could be extreme cases; it may be that a significant segment of the retirement age population does not experience such strong constraints. For these people, retirement decision-making may be as much or more heavily influenced by attitudes. Attitudinal factors are the topic of the following section.

Attitudes and the Retirement Decision

Research on attitudes and the retirement decision can be divided into five categories based on the dependent variable used. Some studies look at attitudinal influences on retirement status (i.e., retired or not); others look at attitudinal relationships with planned retirement age (usually before, at, or after 65), or with views about retirement (that is, is retirement anticipated as a negative or positive experience). Another category of studies, which asks people why they retired, can be examined to see if attitudinal reasons are given. The final study type--perhaps most removed but still related to our interests in the current project--examines attitudinal influences on female labor force participation. As explained below, while not focusing directly on retirement, this last category was included to examine the indirect influence of sex role attitudes on retirement patterns.

Limitations of the Research Done to Date

Studies in the first four categories--that is, those that examine retirement-related variables--are listed in Exhibit C and described in the attached study summaries. We are reasonably confident that the exhibit includes the major studies done on attitudinal influences on the retirement decision. Thus, some of the limitations apparent in this body of research are important to note:

1. Only seven of the studies examine the retirement decision more or less directly, by examining attitudinal differences between retirees and workers. Again, proxy measures used by studies in the other categories, such as planned retirement age, or views of what retirement is likely to be like, may not reflect future decisions.

2. Even those studies analyzing differences between retirees and workers rely on retrospective information for some attitudinal measures. This is because, with three exceptions (Parnes and Nestel, 1975; Schmitt and McCune, 1981, and George et al., 1984), all used one-shot cross-sectional designs, requiring that retirees remember how they felt about, for example, their jobs just before retirement. However, subsequent adjustment to retirement might have colored these memories. This retrospective problem is one that must be acknowledged by any cross-sectional study of the retirement decision.

3. The studies in general are based on limited populations (for example, state civil servants, male auto workers, university employees, or veterans). Moreover, with some exceptions (e.g, Katona et al., 1969; Parnes and Nestel, 1975; Barfield and Morgan, 1970; Louis Harris and Associates, 1981; and U.S. GAO, 1982), many of the studies rely on non-probability samples, some with very small sample sizes.

Exhibit C

Studies of Attitudes and Retirement Status or Related Variables

Reference	Data Base	Population	N	Dependent Variable
		Studies Examining Retirement Status		
George et al. (1984)	RHS, Duke Second Longitudinal study (DSLS)	Men, unmarried 58-63 in 1969[b] (RHS) Men, women 46-70 in 1969[b], in one community (DSLS)	1468 men, 377 women (RHS) 156 men, 79 women (DSLS)	Retirement status
Jaslow (1976)	Not specified	Women, 65+ in 1968[a], nationwide	2398	Retirement status
Katona et al. (1969)	1966 Survey of Consumer Finances	Auto and agricultural implement workers, retired and working, 58-61 in 1969[a]	1123	Retirement status, planned retirement age
Parnes & Nestel (1975)	NLS	Men, 50-60 in 1971[a], nationwide	5020	Retirement status, expected retirement age
Pollman & Johnson (1974)	Original	Men, UAW machine operators, eligible for retirement, 60-65 (year of data collection not specified) midwest	700	Retirement status
Schmitt et al. (1979)	Original	Male and female Michigan civil servants, retired and working, 55-65 (year of data collection not specified)	642	Retirement status
Schmitt & McCune (1981)	Original	Male and female Michigan civil servants, retired and working, 55-70 (year of data collection not specified)	892	Retirement status

[a] Cross-sectional survey design; year is when data were collected.

[b] Longitudinal survey design; year is when data were first collected.

Exhibit C (continued)

Reference	Data Base	Population	N	Dependent Variable
		Studies Examining Planned Retirement Age		
Atchley (1983)	Original	Men and women, employed, 50+ in 1975[a], one community	214 men, 142 women	Attitude toward retirement, planned retirement age
Barfield & Morgan (1970)	Original	Family heads, 35–59 in 1966–67[a], non-retired auto workers, nationwide workers	1652 family head, 646 auto workers	Planned retirement age (before, at or after 65)
McPherson & Guppy (1979)	Original	Men, employed, 55–64 (year of data collection not specified), one community	269	Planned retirement age (before, at or after 65)
Rose & Mogey (1972)	VA Normative Aging Study	Men, employed, veterans, Boston area (ages, year of data collection not specified)	2000	Preferred retirement age
		Studies Examining Attitudes Towards Retirement		
Fillenbaum (1971)	Original	No academic university employees, 25+ (year of data collection not specified)	200	Attitudes toward retirement
Harris (1981)	Original	Men and women, 18+ in 1981[a], nationwide	3427	Attitudes toward retirement
Goudy et al, (1975)	Original	Men, employed, 50+ (year of data collection not specified), midwestern city	1922	Attitudes toward retirement
Johnson & Price-Bonham (1980)	Original	Married women, employed, 50+ (year of data collection not specified), one community	59	Attitudes toward retirement
Price-Bonham & Johnson (1983)	Original	Married women, employed, professional and non-professional, 55–63 (year of data collection not specified), one community	100	Attitudes toward retirement

Exhibit C (continued)

Reference	Data Base	Population	N	Dependent Variable
		Studies Examining Reasons for Retirement		
Evans & Ekerdt (no date)	VA Normative Aging Study	Men, recently retired, veterans, 47-76 in 1981[a], Boston area	70	Reasons for retirement
Messer (1969)	Original	U.S. Civil Service early retirees, <65 (year of data collection not specified)	3299	Reasons for retirement
Palmore (1965)	1963 Social Security Administration Survey of the Aged	Men and women, 62+ in 1963[a], nationwide	7701 men, 9660 women	Reasons for retirement
TIAA-CREF (1983)	Original	TIAA-CREF annuitants, 60-90 in 1982[a]	Not specified	Reasons for retirement
U.S. GAO (1982)	Original	Newly entitled SS beneficiaries in 1980[a], nationwide	1709	Reasons for applying for SS benefits (before vs. at or after 65)

4. Only five studies analyze women as a separate group. Two of them (Johnson and Price-Bonham, 1980; Price-Bonham and Johnson, 1983) use small samples selected from church groups, universities, clubs, etc. Another (Palmore, 1965) is nearly twenty years old and merely presents univariate descriptions of reasons for retirement. The fourth study in this category (Atchley, 1983) is limited to a single community and examines planned retirement age, not actual retirement status. The fifth study (George et al., 1984) based part of its analysis on a small sample from a limited population (participants in a local health insurance plan) and part on the Retirement History Survey, which excludes married women.

In short, then, many of the same limitations evident in the research on women's retirement also hold for research on attitudinal influences. Therefore, the research findings on this topic must be interpreted with similar caution.

What Attitudes Influence the Retirement Decision?

Findings for particular attitudinal factors are summarized below.

Sex role attitudes. As used in the research reviewed here, sex role attitudes refer to approval of women's employment in general and more personal feelings about the saliency of the worker role for oneself. These types of attitudes may influence the retirement decision both indirectly and directly. For instance, based on their analysis of NLS data for young (18-28) and middle-aged (34-48) women, Macke et al. (1979) found that work-related sex role attitudes predict labor force entry for both groups. Dowdall (1974), who studied 15-64 year old married women in one state, also found an association between labor force status and work approval for women in several age categories. Statham and Rhoton (1983), who analyzed ten year work patterns for 30-44 year old women in the NLS sample, found that wives' perceptions of their husbands' attitudes also affect female labor force participation.

The relationship between sex role attitudes and labor force behavior appears to strengthen over time. The Statham and Rhoton (1983) analysis suggests that work experience increases women's commitment to work and, thus, the continuity of their labor force participation and eventual Social Security and pension eligibilty and benefits.

To the best of our knowledge, no researcher has documented differences in sex-role attitudes between retired vs. employed older workers. Only one study, Price-Bonham and Johnson (1983), has come close, investigating sex-role attitudes and feelings about retirement. This research produced somewhat contradictory findings (negative relationship for professionals, positive for non-professionals), which may have been a function of its small, non-probability sample, or which may indicate real occupational

again suggests class-based differences in decision-making. In George et al.'s (1984) analysis, life satisfaction and related variables were not significant predictors of retirement status, or class-based differences. Because of the general lack of attention to this subject, new research directly relating retirement decisions and sex-role attitudes would indeed be ground-breaking.

Life satisfaction. Overall life satisfaction and morale have been investigated in three studies using retirement-related variables. Johnson and Price-Bonham (1980) reported that working women who were highly satisfied with their present lives had less positive attitudes toward retirement than other women. Using a national probability sample of more than 2,000 women, Jaslow (1976) found that, for women 65 and older, morale generally was higher for employed than for retired groups. However, this relationship was reversed for high income women, a finding that but the small sample sizes call into question the power of the tests these researchers applied.

Job Satisfaction. This variable has been studied in relation to retirement more than any other attitudinal factor. Given the variety of populations and measures used, results are surprisingly consistent. For example, several studies, some using limited and others national probability samples, report that retirees recall having lower job satisfaction than do workers, that people with lower job satisfaction tend to plan for earlier retirements, or to view retirement in a more favorable light (Schmitt et al, 1979; Schmitt and McCune, 1981; Fillenbaum, 1971; Goudy et al., 1975; Barfield and Morgan, 1970; Johnson and Price-Bonham, 1980; Atchley, 1983). Investigators who have asked retirees why they retired report that, for a sizeable share, job satisfaction is a major or contributing factor (Evans and Ekerdt, no date; Messer, 1969; TIAA-CREF, 1983).

On the other hand, two studies (Rose and Mogey, 1972; McPherson and Guppy, 1979) found either no or a negative relationship between retirement and job satisfaction measures, and another study (George et al., 1984) found that the relationship held for men, but not for women. Thus, the total picture suggests that the jury is still out on this issue.

Perception of Market Place and Macro-Economic Conditions.

The availability of jobs for older workers is constrained by employer policies on job retention and concerns about the cost and productivity of older workers (Morrison, 1983). While mandatory retirement has been eliminated for many workers under 70 (by the Age Discrimination and Employment Act passed in 1978), employee perceptions of informal pressures to retire may influence their decisions. Conversely, flexible work opportunities for older workers may encourage them to remain on the job not only by offering real work alternatives, but also by changing their perceptions of their employers' attitudes.

Some support for the notion that older workers' perceptions of these kinds of market place conditions may influence their decisions comes from the research done by Evans and Ekerdt (no date). More than a quarter of the men they interviewed as part of a long-term aging study reported that "administrative pressures," such as encouragement to retire or actual job demotion, were a primary or contributing reason for retirement.

Although we have found no corraborating studies, we also speculate that perceptions of macro-level economic conditions may influence decision-making. For example, fear of future inflation may make a prospective retiree reluctant to live on a fixed retirement income. On the other hand, some experts have argued that actual inflation--as opposed to perceptions of inflation--has less impact on retirement than one would expect. This is because Social Security benefits are adjusted for inflation, which also causes private assets, most notably home equity, to increase in value (Clark and Barker, 1981). Similarly, some researchers (Quinn, 1978; Shaw, 1983) have found that actual high rates of local unemployment are associated with labor force withdrawal, although findings about whether both men and women are affected are inconsistent.

Conclusions Suggested by the Research on Attitudes

This paper is concerned with attitudes because we hypothesize that studying attitudes will improve our ability to understand or explain the retirement decisions for both sexes. The research done to date suggests two conclusions relevant to this hypothesis.

First, although few studies include demographic, health, financial, and attitudinal variables together in an explanatory equation, those that do can be examined to see what difference attitudinal variables make. A case in point is Schmitt and McCune's (1981) analysis of retirement status of Michigan Civil servants. In that study, adding job attitudes to an equation already containing demographic, financial, and health variables significantly improved the equation's predictive ability. Another example is the study done by McPherson and Guppy (1979), who examined plans for early retirement among employed 55-64 year old men in an urban community. They also found that adding attitudinal variables--in this case, perception of job constraints and desire for leisure--significantly improved the amount of variance explained by an equation already containing demographic, health, and limited financial information. In a third study of 156 men (George et al., 1984), increased interaction with friends and reporting that one would not work unless necessary significantly increased the probability of retiring, net of age, income, and health. (However, these attitudinal variables were not significant predictors of retirement for women in this same study.) In short, then,

although the evidence is admittedly very slim, it does suggest that attitudinal variables should be included in analyses of the retirement decision.

Secondly, at the same time, the research hints at the fact that financial and employment variables may moderate the influence of attitudinal factors. For example, one study of labor force participation for women (Dowdall, 1974) found that attitudinal factors accounted for the greatest percentage of variance in equations for upper income groups. Thus, we return to the idea suggested at the beginning of this section that attitudinal factors may be most important for people with options: that is, those with sufficient income and in good enough health to make choices about their future.

I. Summary of Research on Women's Retirement Status or Related Variables

Reference	Study Characteristics: Data Base	Population	N	Design	Variables: Dependent	Independent	Results[a]	R^2
				Studies on Women Only				
Chirikos & Nestel (1983)	NLS	Women, 30-44 in 1967, nationwide	3167	Cross-sectional	Labor force participation	Demographic, financial, work/employment and health	Women who are working are more likely to: not have a child < 6 yrs. old (whites), be divorced (whites), have more than 0-8 years of education (whites), have completed a training program, have fewer health limitations	
Shaw (1983)	NLS	Continously married women with intermediate work attachment (i.e., < 6 mos. working per yr. and >2 wks. in >1 yr. in 1966-1971) 30-44 in 1967, nationwide	1131	Longitudinal (respondents interviewed 8 times 1967-1977)	Exit from labor force	Financial, demographic, work/employment, and health	Women who leave the work force are more likely to: not have children 6-12 (whites), have other family income (whites), have a change in ave. income (blacks), have prior work experience (whites), have fewer % of weeks worked, when unemployment rate higher (blacks)	

[a] Parentheses around a group name indicates that a result applies to that group only; otherwise the result applies to all groups studied. Only results with $p<.05$ are listed (for studies that provide levels of statistical significance of results).

Reference	Study Characteristics: Data Base	Population	N	Design	Variables: Dependent	Independent	Results	R^2
Shaw (1984)	NLS	Married women, working, seeking, or intending to seek work, with retirement plans, 45-56 in 1979, nationwide	800+	Cross-sectional	Planned retirement age (<62, 62-64, 65+)	Demographic, financial, work/employment, and health	Women who plan to retire before 62 are more likely to: be eligible for a pension, not be eligible for pension later, not be eligible for SS, not be self-employed, have retired husband, not have husband with health limitation, consider asset income important, not be black	
Henretta & Rand (1980)	RHS	Married women, employed, 58-63 in 1969, nationwide	5833	Longitudinal (respondents interviewed in 1969, 1971, 1973)	Labor force participation	Demographic, financial, and work/employment	Women in the labor force are more likely to: have husband in poor health, be younger, have more quarters of pension coverage, have higher wages, not be covered by husband's pension, be ineligible for pension, support family members	
Honig (1983a)	RHS	Unmarried women, white, with previous SS earnings, 62-67 in 1973, nationwide	1270	Cross-sectional	Retirement status (full employment, partial retirement, full retirement), defined in terms of ratio of current to potential maximum annual earnings	Demographic, financial, work/employment, and health	Women who are fully retired are more likely to: have a health limitation, have less education, have pension coverage, have fewer years of SS earnings, have a high family income, have low or unknown SS benefit, have more years since started SS earnings, not have interrupted SS earnings sequence, have less labor force experience	

	Study Characteristics				Variables			
Reference	Data Base	Population	N	Design	Dependent	Independent	Results	R^2
O'Rand & Henretta (1982)	RHS	Women, unmarried in 1969, with sustained work attachment 1964-1968, 58-63 in 1969, nationwide	1399	Longitudinal (respondents interviewed in 1969, 1971, 1973)	Retirement Status (retire <62 vs. later, retire 62-64 vs. later)	Demographic, financial, work/ employment, and health	Women who retire are more likely to: not have had child and first job after 35, have low occupational status, have less education, have become widowed or separated/ divorced in 1969, have changed marital status in 1969, have a health impairment, have pension coverage	
Sherman (1974)	RHS	Unmarried women, 58-63 in 1969, nationwide	Not specified	Cross-sectional	Labor force participation	Demographic and financial	Women in the labor force are more likely to: not receive SS, not receive support from children, support children, be white, be younger, be divorced, separated or never married, have more education, be in good health	
Dowdall (1974)	Original	Married women, with $\geq$ 1 children, 15-64 in 1968-1969, Rhode Island	592	Cross-sectional	Employment status	Demographic and attitudinal (see Appendix 2)	Women who are employed are more likely to: be better educated, have lower husband's income, have older youngest child	.04 for non-attitudinal variables

Reference	Study Characteristics: Data Base	Population	N	Design	Variables: Dependent	Independent	Results	R^2
Jaslow (1976)	Not specified	Women, 65+ in 1968, nationwide	2398	Cross-sectional	Employment status (currently employed, retired, never employed)	Demographic, health, and attitudinal (see Appendix 2)	Age, health, physical limitation, and income related to employment status (nature of relationships indeterminate)	
Johnson & Price-Bonham (1980)	Original	Married women, employed, 50+ (year of data collection not specified), one community	59	Cross-sectional —	Acceptance of retirement stereotypes, resistance to retirement	Demographic, financial, work/employment, and attitudinal (see Appendix 2)	Women who resist retirement or have stereotypes of retirement are more likely to: plan to continue business activities after retirement, be at present job shorter time, have more or fewer post retirement social activities planned, not expect to use savings in retirement, have smaller husband's pension, have larger assets	.57 (analysis inappropriate with this N)
Price-Bonham & Johnson (1983)	Original	Married women, employed, professionals and non-professionals, 55-63 (year of data collection not specified), one community	100	Cross-sectional	Attitude toward retirement	Demographic, financial, work/employment, and attitudinal (see Appendix 2)	Women with positive attitudes toward retirement are more likely to: be at work shorter time (prof.), work fewer hours weekly (prof.), have less education (prof.), have higher family income (prof.), plan more for retirement (nonprof.), have fewer sources of retirement income (nonprof.), view husbands pension as income source, plan to continue work activities, have fewer leisure activities planned	

II. Summary of Research Comparing Men and Women's Retirement Decisions or Related Variables

Reference	Study Characteristics: Data Base	Population	N	Design	Variables: Dependent	Independent	Results	R^2
Studies on Men and Women								
Anderson et al. (1980)	RHS	Husbands and wives, 58-63 in 1969, nationwide	600 couples (multiple observations per respondent)	Longitudinal (respondents interviewed in 1969, 1971, 1973)	Labor force participation	Demographic, and financial	Men and women in the labor force are more likely to: be eligible for SS (men), be ineligible for SS (women), have spouse under 65 (men), be older (women), have older spouse (men), have spouse ineligible for SS (men), have higher wages (women), have higher spouse's wages (men), have lower real housing value (women)	.31 for husbands, .10 for wives
Clark & Johnson (1980)	RHS	Married men and women, 58-63 in 1969, nationwide	2657 couples	Cross-sectional	Labor force participation	Demographic, and financial	Men and women in the labor force are more likely to: be ineligible for pension (men), be eligible for pension (women), have lower SS wealth, have lower spouse's wages, have higher wages, have lower welfare income, have lower assets, have spouse in labor force, be ineligible for SS (women), have spouse ineligible for pension (women), have lower pension wealth (women), have lower spouse's SS wealth (women), support children (men), be older (women), not have a family disability (men)	

	Study Characteristics				Variables			
Reference	Data Base	Population	N	Design	Dependent	Independent	Results	R^2
Clark et al. (1980)	RHS	Husbands and wives, non-self-employed, 58-63 in 1969, nationwide	3312 married couples	Longitudinal (respondents interviewed in 1969, 1971, 1973)	Labor force participation	Demographic and financial	Men and women in the labor force are more likely to: have fewer extra pensions (men), have more extra pensions (women), have lower SS wealth (men), have higher SS wealth (women), have higher wages, have lower spouse's wages, have lower assets, have lower home equity, be younger (women), be ineligible for SS (women), have fewer spouse's extra pensions (women), have lower spouse's SS wealth (women), have lower welfare income (men), support more children (men)	
George et al. (1984)	RHS	Men, unmarried women 58-63 in 1969	1948 men 377 women	Longitudinal	Retirement Status	Demographic, finances, health, job attitudes	In both analyses, retirement related to multiple variables for men, but only to age for women	.16 men .09 women
	Duke Second Longitudinal Study (DSLS)	Men, women 46-70 in 1969	156 men 79 women	Longitudinal	Retirement Status	Demographic, finances, health, job attitudes		.25 men .25 women
Hall & Johnson (1980)	RHS	Married men and unmarried women, employed, 58-63 in 1969, nationwide	3557 men, 1054 women	Cross-sectional	Planned retirement age	Demographic, financial, work/employment, and health	Men and women planning to retire earlier are more likely to: be eligible for SS, be eligible for pensions, have higher wages, have lower non-wage income, not be self-employed, have less education, have poorer health, own a home (men); men who plan to never retire before 62 are more likely to have a job with a compulsory retirement age; women who plan to retire after 65 are more likely to have a job with a compulsory retirement age	

	Study Characteristics				Variables			
Reference	Data Base	Population	N	Design	Dependent	Independent	Results	R^2
Hanoch & Honig (1983)	RHS	Married men and unmarried women, 58-68 in 1969, nationwide	3130 men, 1359 women (multiple observations per respondent)	Longitudinal (respondents interviewed in 1969, 1971, 1973, 1975)	Labor force participation	Financial, demographic, work/employment, and health	Men and women in the labor force are more likely to: be self-employed, have more work experience, have less work experience on longest job, have more education, have fewer health limitations and disabilities, have more years of SS covered earnings, have more years since SS first covered earnings, have an interrupted SS earnings sequence, have a positive PIA entitlement, have a lower PIA entitlement, have a lower family non-wage income; men in the labor force are less likely to be covered by pensions	.37 for men, .36 for women
Honig (1983b)	RHS	Married men and unmarried women, 58-69 in 1975, nationwide	Not specified (multiple observations per respondent)	Longitudinal (respondents interviewed 1969, 1971, 1973, 1975)	Labor force participation	Demographic, financial, work/employment, and health	Men and women in the labor force are more likely to: be eligible for SS, have smaller SS benefits, etc. (details of other relationships not specified)	

	Study Characteristics				Variables			
Reference	Data Base	Population	N	Design	Dependent	Independent	Results	R^2
Quinn (1978)	RHS	White married men (WMM), white unmarried men (WUM), "other than white married men" (OWMM), white unmarried women (WUW), 58-63 in 1969, nationwide	4539 WMM, 667 WUM, 417 OWMM, 2224 WUW	Cross-sectional	Labor force participation	Demographic, financial, and work/employment	Men and women in the labor force are more likely to: have fewer health limitations, be ineligible for SS, haver lower income from assets (WMM, WUW), have higher job autonomy (men), be ineligible for pension and SS (WMM), have more dependents (WMM), have worse working conditions (WUW), have higher wages (WUW); men are more likely to be in the labor force when: Unemployment rate lower (WMM), employment rate changes <2% (OWMM), employment rate changes <4% (WMM)	.25 for WMM, .20 for WUM, .37 for OWMM, .18 for WUW
Atchley (1983)	Original	Men and women, employed, 50+ in 1975, one community	214 men, 142 women	Cross-sectional	Attitude towards retirement, plans to retire, planned retirement age	Demographic, health, and attitudinal (see Appendix 2)	Men and women who plan to retire later are more likely to: expect more pensions (women), expect fewer pensions (men), have lower occupational status (women)	
Barfield & Morgan (1978)	Original	Husbands and wives, unmarried men and women, 35-64 in 1976, nationwide	394 married couples, 32 unmarried men, 60 unmarried women	Cross-sectional	Plans to retire early (<65 for husbands, unmarried men and women; <62 for wives)	Demographic, financial, and work/employment	Men and women who plan to retire early are more likely to: be younger (husbands), expect extra pensions (husbands), have higher family income (wives), be younger when free of mortgage payments, be younger when free of responsibility for children (wives)	.06 for husbands, .09 for unmarried men and women, .01 for wives

	Study Characteristics				Variables			
Reference	Data Base	Population	N	Design	Dependent	Independent	Results	R^2
Chartock (no date)	Original	Men and women, retired from the retail trades	400	Cross-sectional	Retirement decision	Demographic, financial, work/employment, and attitudinal (see Appendix 2)	Men and women differ on labor force attachment but not on commitment to work and job involvement	
Haug et al. (1982)	Original, in progress	Men and women, enrolled in HMO, 60-64 in 1984, Cleveland area	300 men, 500 women	Longitudinal (respondents interviewed 1983, 1985)	Retirement status, intention to retire, adaption to retirement	Demographic, financial, work/employment, health, and attitudinal (see Appendix 2)	Not available yet	
Palmore (1969)	1963 Social Security Administration Survey of the Aged	Men and women, 62+ in 1963, nationwide	7701 men, 9660 women	Cross-sectional	Retirement Status	Demographic, and reasons for retirement (see Appendix 2)	Men only are more likely to retire if they are in lower paid occupations	
Schmitt et al. (1979)	Original	Men and women, retired and working Michigan civil servants 55 $\leq$ 65 (year of data collection not specified)	642	Cross-sectional	Retirement Status	Demographic, work/employment, and attitudinal (see Appendix 2)	Women are more likely to be retired	.10 for attitudinal variables, .22 for non-attitudinal variables, .28 for both

	Study Characteristics				Variables			
Reference	Data Base	Population	N	Design	Dependent	Independent	Results	R^2
Streib & Schneider (1971)	Cornell Study of Occupational Retirement	Men and women, > 63 in 1952, nationwide	1486 men, 483 women	Longitudinal (respondents interviewed 5 times 1952-1958)	Retirement timing	Financial and demographic	Men and women who retire later are more likely to: have higher status occupations, have more education, have better health (women), have higher income; women who retire later are more likely to be widowed or separated/divorced	

III. Summary of Research on Attitudes and Retirement Status or Related Variables

Reference	Study Characteristics: Data Base	Population	N	Design	Variables: Dependent	Independent	Results[a]	R^2
Studies Examining Retirement Status								
George et.al., (1984)	RHS	Men, unmarried women 58-63	1845	Longitudinal	-Retirement status	-Job attitudes	-Attitudinal variables predict retirement for men only	
	Duke Second Longitudinal Study (DSLS)	Men, women 46-70	235	Longitudinal	-Retirement status	-Job attitudes, self-concept, subjective well being		
Jaslow (1976)	Not specified	Women, 65+, national probability sample	2398	Cross-sectional	-Retirement status	-Morale	-Morale higher among employed than among retired women	
Katona et al. (1969)	1966 Survey of Consumer Finances	58-61 in 1969, auto and agricultural implement workers, retired and non-retired	1123 (heads of households)	Cross-sectional	-Retirement status -Planned retirement age	-Job satisfaction and related measures, -Leisure plans	-Job satisfaction not related to retirement status or planned age -Having trouble keeping up with job related to planned age	
Schmitt et al. (1979)	Original	Male and female Michigan civil servants 55 $\leq$ 65, (year of data collection not specified), retired and working	672	Cross-sectional mail survey	-Retirement status	-Various measures of perceptions of job and job satisfaction -Locus of control, self esteem	-Retirees report jobs less intrinsically satisfying -locus of control and self esteem not related to early retirement	.10- attitudinal only, .22- demographic, .28- total set of variables
Schmitt & McCune (1981)	Original	Michigan civil servants employees, 55-70, retired and not (year of data collection not specified)	892	Longitudinal (respondents interviewed one year after initial interview)	-Retirement status	-Job satisfaction and related measures	-Those who retired early viewed their jobs as less involving and challenging, although retirees and non-retirees were equally satisfied with their jobs	.36 (canonical correlation for all variables)

[a] only results with $p<.05$ are listed (for studies that provide levels of statistical significance of results).

	Study Characteristics				Variables			
Reference	Data Base	Population	N	Design	Dependent	Independent	Results	R^2
Parnes & Nestel (1975)	NLS	Men, 50-60 in 1971	5020 (9% Retired)	Cross-sectional	-Retirement status -Expected retirement age	-Job satisfaction -Work commitment	-Work commitment, job satisfaction related to retirement status -Job satisfaction positively related to expected retirement age	.11 for all variables and retirement status, .23 for all variables + expected retirement
Pollman & Johnson (1974)	Original	UAW male machine operators, 60-65 (time of data collection not specified) midwest	700 (1/3 retired)	Cross-sectional mail Survey	-Retirement status	-Changes in old job -Job transfer	-Recent and expected job changes greater for retirees	
				Studies Examining Planned Retirement Age				
Atchley (1983)	Original	Pre-retirees, men and women, 50+ in 1975, small town community	346	Cross-sectional	-Planned retirement age, attitude toward retirement	-Attitude toward job -Goal directedness -Life satisfaction -Self confidence	-Attitudes toward work + retirement uncorrelated -For women the less positive the attitude toward retirement, the higher the planned retirement age -For men, positive attitude toward job related to higher planned retirement age	

Reference	Data Base	Population	N	Design	Variables: Dependent	Variables: Independent	Results	R^2
Studies Examining Planned Retirement Age								
Barfield & Morgan (1970)	Original	-Family heads, nationwide, 35-59 in 1966-67 -Non-retired auto workers	1652 (national) 646 (auto)	Cross-sectional	-Planned retirement age (before, at, or after 65)	-Attitudes towards work in general -Various job satisfaction, work involvement measures	-More people with negative work attitudes, lower job satisfaction planned to retire before 65	
McPherson and Guppy (1979)	Original	Employed men, 55-64 (year of data collection not specified), one urban community	269	Cross-sectional	-Planned retirement age (before, at, or after 65)	-Job satisfaction -Perception of pension adequacy -Leisure orientation	-People satisfied with job, and with greater leisure orientation more likely to plan early retirement	.11 job satisfaction (eta), .24 leisure orientation (eta)
Rose & Mogey (1972)	VA Normative Aging Study	Employed men, age not specified	2000	Cross-sectional	-Preferred retirement age	-Various job satisfaction measures	-Job satisfaction measures not significantly related to preferred retirement age	
Studies Examining Attitudes Toward Retirement								
Fillenbaum (1971)	Original	Non-academic employees of a university, 25+ (year of data collection not specified	200	Cross-sectional mail survey, (56% response rate)	-Attitude toward retirement (3 item scale)	-5 job satisfaction measures	-More people reporting less chance of increasing skills on the job viewed retirement as a good thing	

Reference	Study Characteristics: Data Base	Population	N	Design	Variables: Dependent	Independent	Results	R^2
Studies Examining Attitudes Toward Retirement								
Harris (1981)	Original	U.S. population, 18+ in 1981	3427	Cross-sectional personal interviews	-Attitude toward retirement	-None	-45% of people "Look forward to retirement"	
Goudy et al. (1975)	Original	Employed men, 50+ (year of data collection not specified), midwestern city	1922	Cross-sectional personal interviews	-Attitude toward retirement (11 items)	-Job satisfaction	-People satisfied with jobs less likely to work forward to retirement	
Johnson & Price-Bonham (1980)	Original	Employed, married women, 50+ (year of data collection not specified), one community	59, non-probability sample	Cross-sectional personal interviews	-Attitudes toward retirement	-Various measures of job satisfaction, primary role perception, and life satisfaction	-Higher life satisfaction associated with less resistance to retirement -Higher job satisfaction associated with fewer negative retirement stereotypes	.57 for all variables
Price-Bonham & Johnson (1983)	Original	Employed, married women, 55-63 (year of data collection not specified), professionals and non-professionals, one community	100, non-probability sample	Cross-sectional personal + mail interviews	-Attitudes toward retirement	-Various measures of job satisfaction, primary role perception, and life satisfaction, sex role attitudes	-Prof. women with higher work commitment had more negative attitudes toward retirement -For non-professionals, positive retirement attitudes associated with more egalitarian sex role attitudes, and lower life satisfaction.	.50 for all variables

Reference	Study Characteristics: Data Base	Population	N	Design	Variables: Dependent	Independent	Results	R^2
Studies Examining Reasons For Retirement								
Evans & Ekerdt (no date)	Original	Recent male retirees, 47-76 in 1981, from VA Normative Aging Study	70	Cross-sectional	-Reasons for retirement	-None	-Job dissatisfaction cited as primary reason by 21.4% + as contributing reason by 58.6%	
Messer (1969)	Original	U.S. Civil Service early retirees, <65 (year of data collection not specified)	3299	Cross-sectional retrospective mail survey	-Reasons for retirement	-None	-17.9% left due to job dissatisfaction -22.6% left while still time to enjoy retirement	
Palmore (1965)	1963 SSA Survey of the Aged	Men + women, 62 + in 1963	7701 men, 9660 women	Cross-sectional	-Reasons for retirement	-None	-30% of men retired for voluntary reasons (e.g., preferred leisure) -Over 50% of women retired for voluntary reasons (e.g. preferred leisure)	
TIAA-CREF (1983)	Original	TIAA-CREF annuitants, 60-90 in 1982	Not specified	Cross-sectional mail survey (84% response rate)	-Reasons for retirement	-None	-42% sought retirement as an attractive alternative to current employment -10% viewed current employment as no longer suitable	
U.S. GAO (1982)	Original	Newly entitled SS beneficiaries in 1980	1709	Cross-sectional mail survey (80% response rate)	-Reasons for applying for SS benefits before vs. at or after 65	-None	-Desire to do things enjoyed or be with family cited as major reason by 35% of both early + other retirees	

APPENDIX II

SUMMARY OF INTERVIEWS WITH PRE AND RECENT RETIREES

These notes are not verbatim, but capture the flavor of the questions and responses in each interview session.

INTERVIEW ONE

Respondent is a single woman, age 62, who is currently working in a professional capacity.

Q: Have you thought about retirement?

R: Yes. I plan to retire when I reach 65, but I'll work part-time after that. I would like to leave my job when I'm in good enough health to enjoy the rest of my life. I want to start something new, but I'll still think of myself as retired.

Q: What do you think about when you think about retirement?

R: My job. I like my job, and I like the people here, but I feel older than most people at work. I would like to be with people my own age. Things are done differently on the job now, with a lot of people using computers. I don't want to learn that.

Q: For how long have you been thinking about retirement?

R: I always thought I'd retire at about 60 to 65, but when I was younger I thought I'd retire and do nothing. Now I know I must do something, like teaching or consulting.

Q: What do you think are the advantages of retirement? The disadvantages?

R: (RESPONDENT PULLED OUT A CARD. SHE HAD WRITTEN "ADVANTAGES" OF WORK ON ONE SIDE, AND "ADVANTAGES OF RETIREMENT" ON THE OTHER.)

Advantages of Work: structured life; feel needed; involved in interesting projects; association with colleagues; slightly more money; professional recognition; prestige of having a job in a work-oriented society.

Advantages of Retirement: freedom to do what you want; more time with friends and family; probably better health and more exercise; get away from snow, ice, and cold weather; learn new things; may have option for more interesting work, irrespective of pay; no longer need to commute.

Q: Will you consider yourself retired, even though you plan to work after you leave your present job?

R: Yes. Once you accept a pension, you don't have a career orientation anymore. You know you can quit any job you're working at.

Q: Do you worry about your finances in retirement?

R: No. Before they retire, people generally underestimate what their income and overestimate what their expenses in retirement will be. They don't realize that their expenses will go down.

ANALYSIS: This woman will have a good pension when she retires. She knows what her benefits will be and doesn't seem to be worried about the financial aspect of retirement. But she does seem concerned that retirement will take away the structure she has in her life and the prestige she gets from working. She seems to have dealt with this concern by planning to work, although part-time, after retirement.

INTERVIEW TWO

Respondent is a married woman, age 62, who retired 4 years ago after 30 years as a teacher. Her husband has been retired for two years.

Q: How did you happen to retire when you did?

R: I taught for over 30 years all over the country, following my husband who was career military. I started thinking about retiring two years before I actually did it; I had 37 kids in my classroom that year.

I talked with my husband about retirement and my family also thought that I should retire. I agreed, since I thought my health would be better---I had hypertension.

Q: When you considered retirement, what sold you on it?

R: I wanted to be able to do things for myself and for my community. I had guilt feelings about not being able to serve the community, especially about not being able to work for my church. Retirement would give me the time I needed to do these things.

I only had a slight hint of what my pension would be. In fact, I overestimated it.

Q: Why did you retire at the particular age you did (62)?

R: Society makes 62 or 65 the year to retire. It's almost a custom. However, more and more among educators, I'm hearing people say: "I've served my 30 years and I want out." The system also pressures people to leave when they reach their 60s.

ANALYSIS: This woman seems to have retired because of family pressures, custom, and perhaps concern about her health. Job dissatisfaction also seemed to be a factor, as well as the desire to commit her energies elsewhere (church and volunteer work). Finances may have been a consideration, though not a major one since she didn't know what her pension would be.

INTERVIEW THREE

Respondent is a 47 year old married man who retired one year ago after about 30 years of service as a blue collar worker in a government agency.

Q: How did you come to retire when you did?

R: I knew I'd be getting a pretty good pension. I also couldn't see myself advancing in my job.

Q: When did you start thinking about retirement?

R: I thought about it for two years. I knew I'd be eligible when I had 30 years of service. When I had 23 years of service, I started my own business, so I'd have something to do when I retired.

Q: Do you consider yourself retired?

R: I'm retired because I'm drawing a pension.

Q: What are the advantages of retirement? The disadvantages?

R: I can do what I want when I want to. Before I felt a responsibility to be on the job everyday. Now my brother can look after my business when I'm not there. The disadvantages of retirement? Right now I can't think of any.

Q: What steps did you take to prepare for retirement?

R: I talked to my retirement counselor about a year before. He gave me the in's and out's of retirement. The biggest consideration for me was would I be able to survive my retirement if my business went bad (his pension is about 60 percent of his former salary.) Before I retired, I talked it over with my wife too.

ANALYSIS: The biggest consideration here seems to have been pension eligibility and also pension amount (replacement value). Perhaps some job dissatisfaction (inability to advance) and desire to be out on his own also were involved, but pension was what made it all possible.

INTERVIEW FOUR

Respondent is a 62 year old married woman, who retired last year after over 40 years as a teacher.

Q: How did you happen to retire when you did?

R: It was the most difficult decision. I loved my work. I'd been teaching in the same district for 40 years. I have no kids, and the children were my life.

But I wanted to go out on top. And I did. I went out with a standing ovation. They didn't say: "She should have retired last year."

Q: What else went into your decision?

R: I considered my pension, but I was eligible to retire when I was 50. But every year after I was 50 I'd say---I'll teach one more year. I was afraid to give up the security of my job. I kept saying, what will I do? If you've had your own kids, it's different. You've been a Brownie mother, you have other interests. But I didn't. The kids at school were my kids.

Q: Was your retirement decision made jointly with your husband?

R: Not really. My husband wanted me to retire, but wouldn't influence me one way or the other. What it came down to was that I felt very good about my job, but also wanted freedom from the 9 to 4 routine. So, although the decision to leave was very emotional (I cried when I handed in my papers), I finally made it. But I'll still keep in touch. I may substitute and also will get more involved in church work.

ANALYSIS: For this woman, a primary motivator seems to have been commitment to work. Since she was so involved in her career (and probably in the sense of community associated with being a longtime employee of one school system), the decision to break with the work role was a very difficult one. However, her degree of commitment also meant that she worked very hard to do a good job; keeping up that level of energy indefinitely was something she chose not to do. As she put it, she wanted to go "out on top," and still have some time left for other pursuits. While finances seemed to play a part, and her impending eligibility for Social Security may even have been a precipitating factor, feelings about her job (tiredness, wanting to be excellent) seemed to have been a much more important motivator.

Perhaps this woman's experience reflects that of other career women. That is, perhaps women who have been working all their lives will follow the early retirement pattern typical of men.

INTERVIEW FIVE

This is a summary of a group interview session with seven school system employees who had retired recently or were considering

retiring in the next few years. All occupation groups were represented, including supervisors, a social worker, a secretary, and several teachers. The school system retirement counselor also was present.

Q: How important are financial considerations in the decision to retire?

R#1 (female, retired, divorced secretary): Finances are important because if one is old enough to retire, it won't be easy to get a supplemental job. This is cause for concern.

R#2 (male, planning retirement within a year, married, supervisor): Finances have to be the most important consideration. If you can't afford to retire, you can't retire.

R#3 (female, mid-40s, single, has a supervisory position): I will be eligible for retirement in eight years. I am considering manay options, but for me the biggest factor is financial, because I am self-supporting. I need to think about what the economy will look like in a few years, and how inflation may affect my benefits. I'm using a formula to project various retirement possibilities. I have about 25 different options based on varying assumptions about taxes, medical expenses, inflation, and so on.

Q: Did any of you defer retirement, even though you were eligible for retirement benefits?

R#1 (female, retired secretary): Yes, I wasn't ready to retire. I was happy with what I was doing and didn't feel my age. But my supervisor retired a year before I did. She was a factor in my starting to think about retirement. I am divorced and self-supporting. I had a fear about whether I would have to change my lifestyle in retirement. I did a lot of thinking and research about how I could maintain my present lifestyle.

R#4 (female, late sixties, retired for one year, single, social worker): It was the opposite for me. My benefits were a lot lower than I thought they would be. I was one who could not retire, but I retired anyway. I retired because I wanted to get out of the job. Right now I'm surviving on my savings. I know this can't go on forever, but it will go on until I have enough options and until I feel like I'm retired. I'm just not ready to think about another job.

R#5 (female, late 50's, divorced, secretary): I'll retire one year before I'm eligible because I got excited about it and have another business, but don't have time to pursue it. It's a business of my own; its lucrative. I also have offers of two jobs and I'm considering some volunteer work. There are so many things that I want to do. I find that even though I like my job as a secretary, I feel that I've already lived several lifetimes and I've got another one coming up.

Q (to R#5): Was there any particular event that precipitated your thinking about retirement?

R#5: I'd been talking with friends of mine who had retired in the past year and each one is totally excited. I had been searching for someone else who is single and retired, because it's a totally different ballgame for us. I've played with the retirement idea for about six months. I talked with the retirement counselor and he gave me some figures and they looked good.

Q; How important is health in the decision to retire?

R#6 (female, married, mid-fifties, teacher): Health is important fo me. I want to be able to retire while I'm in good health. I love teaching and my students, but I'm overloaded with paperwork and don't have any preparation time. I don't want to work all my life and never be able to enjoy it. It's wrong to increase the Social Security eligibility age to 67 and 70 after we've had to pay in all these years. By 70, you just don't get out of bed. You're exhausted and can't enjoy it, and I want to be able to enjoy it.

Q: How do people decide at what age to retire?

R#3 (female, mid-forties, single, supervisor): It's very personal. There's no one reason. I am concerned about whether my health will stand up over the next few years. I'm feeling drained. I'm feeling a need to protect myself. Retirement is becoming more important as I look to that need.

R#6 (female, married, mid-fifties, teacher): I am used to a way of living and I'm not ready to turn it off. That's one of the reasons I'm still working. I'm concerned about losing my health and life insurance benefits once I stop working as well as being able to afford utility bills.

Q: What is the role of IRA's in the retirement decision?

R#2 (male, supervisor, planning retirement within one year): Our IRA has been beneficial. We were able to get a tax break and additional money for my wife's retirement. It figured into my decision. I will boost my income between the time I retire and the time I become eligible for Social Security.

R#5 (female, late 50s, secretary): Four or five years ago someone showed me how I could have a tax-sheltered annuity and have it taken out before I see it. I've increased and increased my contributions up to nearly 20 percent of my income. The last two years were a test of whether I can live in poverty. I've cut down my income to the point where I could test myself, see how little I could live on, and put the rest into savings. It's what I'll have to live on after I retire, until I get another job.

Q: Are there differences in retirement decisions between women with uninterrupted career patterns and women with interrupted careers?

R#7 (Retirement counselor): The career person tends to look more at early retirement. People with interrupted careers tend to work longer (into their sixties) for two reasons. They didn't enter until later so that if they like their jobs they want to stay with them longer. They also want to build up a little more security, for example, credit for Social Security and retirement.

R#2 (male, supervisor): My own situation is probably more common for men than for women. If I hadn't had people dependent on me, I wouldn't have retired totally, but I would have changed jobs. But because of retirement and Social Security credit I'd built up and benefits and job longevity, you don't tend to change jobs after some point. This is more common with men, but it's occurring more now with women than it used to.

APPENDIX III

ASSESSMENT OF EXISTING DATA BASES FOR SECONDARY ANALYSIS

To determine if our research questions could be answered through secondary analysis, we assessed extant data bases that have potential value to studies in the retirement area. Secondary analysis would have involved reanalyzing a data base that included the population and content of interest in this study. Since it would not have involved any new data collection, secondary analysis would have been ideal had an appropriate data base been available. Unfortunately, that was not the case. Therefore, the secondary analysis option was rejected. The data bases, and the assessment process and its results are described in this paper.

Alternative Data Bases

We considered 12 data bases for secondary analysis (see Exhibit 1). They were selected because they had some content of interest to us. The descriptions of them that follow are from Boyce (1984):

1. Longitudinal Retirement History Survey (RHS). The RHS was developed to study the retirement decision. It is a longitudinal study of 11,153 men and mostly single women aged 58 to 63 in 1969. New data were collected every one or two years from 1969 to 1979.

2. National Longitudinal Survey of Labor Market Experience (NLS). The NLS was developed to study the work patterns and experiences of a representative sample of the U.S. population as of 1966. It is a longitudinal study of six age cohorts. Data collection on four of these cohorts began in 1966-1967 and has continued periodically since then. Two additional cohorts entered the study in 1979.

3. Michigan Panel Survey of Income Dynamics (MPSID). This longitudinal survey began in 1968; data have been collected annually since then. It focuses on short-term changes in family economic status. The sample includes 4900 families, with low-income families oversampled. Later waves included any new families formed by the initial respondents.

4. Survey of Income and Program Participation (SIPP). The SIPP is a panel study of households that has recently entered the data collection stage. It focuses on finances and participation in social programs. Once a household is sampled, all adults at that address are administered the questionnaire. Two staggered panels of 20,000 households will be followed for two and a half years.

5. Current Population Survey (CPS). The CPS is a monthly rotating survey of about 60,000 households sponsored by the Bureau of Labor Statistics. It routinely collects

Exhibit 1

Assessment of Extant Data Bases for Use in the Secondary Analysis Option

		Population Coverage				Content Coverage[a]					Sampling Plan		
Data Base	Recency (year of last collection)	Age[b]	Gender[c]	Retirement Status[d]	Marital Status[e]	Area of residence[f]	Demographic	Financial	Employment and work-related	Health	Attitudinal	Sample size	Type of sampling
Longitudinal													
RHS	1979	-	+	+	-	+	+	+	+	+	+	+	+
NLS	1984	-	+	+	+	+	+	+	+	+	+	+	+
MPSID	1984	+	+	+	-	+	+	+	+	+	+	+	+
SIPP	1984	+	+	+	+	+	+	+	+	+	-	+	+
CPS	1984	+	+	+	+	+	+	+	+	-	-	+	+
Cross-sectional													
DCPP	1979	+	+	-	+	+	+	+	+	-	-	+	+
SCF[g]	1983	+	+	+	+	+	+	+	+	-	-	+	+
NBS	1982	-	+	-	+	+	+	+	+	+	-	+	+
AIA	1981	+	+	+	+	+	+	+	-	+	+	+	•
TIAA-CREF	1982	-	+	-	+	+	+	+	+	+	+	+	-
GAO	1980	-	+	-	+	+	+	+	-	+	-	+	+
SDW[h]	1978	+	+	•	+	+	+	+	+	+	+	+	+

Note. A positive sign ("+") indicates that the data base meets the criterion; a negative sign ("-") indicates that the data base does not meet the criterion.

[a] To meet the criterion for a category of factors, the data base must contain items for the category.
[b] To meet the criterion for age, the data base must represent persons 40-70 years old.
[c] To meet the criterion for gender, the data base must represent males and females.
[d] To meet the criterion for retirement status, the data base must represent retirees and non-retirees.
[e] To meet the criterion for marital status, the data base must represent married and non-married persons.
[f] To meet the criterion for area of residence, the data base must represent persons nationwide.
[g] For most content categories, the unit of analysis is the household (i.e., individual level analysis is precluded).
[h] Respondents were disabled.

information on demographics and employment. Although it is most often treated as a cross-sectional survey, the CPS can provide longitudinal data.

6. Survey of the President's Commission on Pension Policy (PCPP). The purpose of the PCPP was to acquire information on the retirement income that will be available to future retirees. The unit of analysis for the survey is the family. Only families with at least one family member in the labor force in 1979 were included.

7. Survey of Consumer Finances (SCF). Data for the SCF were collected in 1983 in two parts. Individual financial data were collected from 4500 households, with heavier sampling of high-income households; detailed pension coverage information was collected from actual pension providers. A second wave of data collection is planned for 1985.

8. New Beneficiary Survey (NBS). This survey was based on a sample of 18,6000 social security beneficiaries who were entitled to benefits in 1980-1981.

9. Aging in America (AIA). This public opinion survey is an update of a similar survey conducted in 1974. The total sample is 3427, with heavier sampling of minorities and older Americans.

10. Teachers Insurance and Annuity Association-College Retirement Equities Fund (TIAA-CREF). The TIAA-CREF survey sampled approximately 2000 individuals who were at least 60 years old in 1982 and receiving retirement income from TIAA-CREF. It focused on the opinions and attitudes of annuitants. A similar survey of TIAA-CREF annuitants was conducted in 1972.

11. U.S. General Accounting Office (GAO). This 1980 survey focused on finances, housing, and reasons for retirement. It sampled early retirees (i.e., those who retired at or before 65).

12. 1978 Survey of Disability and Work (SDW). This survey is an updated version of ones conducted in 1972 and 1974. It surveyed 12,000 disabled individuals on program knowledge, work incentives, and health.

Assessing the data bases.

We judged each of the data bases against the following sets of criteria:

1. Population coverage. Data bases should represent men and women, 40 to 70 years old, retirees and preretirees, married and nonmarried persons, from all regions of the country. This criterion is dichotomous: either a data base includes all the subgroups or not. Failure to meet it

eliminated a data base from consideration for possible secondary analysis.

2. Content coverage. Data bases should provide in-depth information on the categories of factors identified as potentially important influences on retirement decision-making. These categories are demographic, financial, employment and work-related, health, and attitudinal. Data bases were rated on the quantity and quality of the items they have for each of the five categories of factors. Data bases were eliminated if they completely exclude items for any category. (Although it wasn't necessary, data bases also would have been rejected if it had appeared that their items failed to cover adequately all aspects of a category or were weak psychometrically.)

3. Sampling plan. Data bases should be built on a probability sample. They also should have a sample of at least 1200 repondents to allow for subgroup analyses. Data bases that do not meet these criteria were eliminated from consideration for secondary analysis.

4. Recency. Data bases should have collected data within the past five years (i.e., no earlier than 1979). Because five years is a somewhat arbitrary cutoff, this criterion was used more as a guideline for assessing data bases than as a strict requirement.

5. Design. Longitudinal data bases were rated higher than cross-sectional ones for three reasons. First, longitudinal data have potential for use in cohort analyses. This feature is important to us since one of the objectives of a follow-on to the Women's Retirement Project, the "More Comprehensive Model Project," may be to predict the retirement trends of younger cohorts with a model developed with data on older cohorts. Prediction assumes that differences between the cohorts are determined by influences associated with birth cohort membership (e.g., the period in which cohort members developed) rather than those associated with aging. Violation of this assumption threatens the validity of the predictions. The assumption can be examined with longitudinal data that permit cohort analyses but not with cross-sectional data. A second reason for giving preference to longitudinal data bases is, owing to their several data collection points, they are likely to contain contemporary rather than retrospective data on key variables (e.g., health status of retirees when they were last employed) and more detailed data overall. Another reason is longitudinal designs permit analyses of individual change over time (e.g., on attitudes towards retirement) and the forces that contribute to it.

Exhibit 1 shows how the 12 data bases fared when they were judged against these criteria. Only three data bases survived the population coverage criteria: SIPP, CPS, and AIA. (Although, SCF and SDW appear to meet the population

criterion, they suffer other related flaws. SCF uses the household as the unit of analysis for most variables and thus precludes individual level analyses. SDW has data on disabled persons only.) The other data bases omit one or more subgroups and were rejected. When we assessed the remaining three data bases against the content criterion, we found that each of them neglects altogether at least one category of factors (see Exhibit 1). They too were eliminated.

In sum, secondary analysis was not a viable option for the women's retirement project. None of the data bases met population and content coverage criteria; none could be used alone to answer our research questions. Other data source options needed to be considered.

United States
General Accounting Office
Washington, D.C. 20548

Official Business
Penalty for Private Use $300

First-Class Mail
Postage & Fees Paid
GAO
Permit No. G100