Review of the Disability Evaluation Study Design: Third Interim Report

Gooloo S. Wunderlich and William D. Kalsbeek, Editors; Committee to Review the Social Security Administration's Disability Decision Process Research, Institute of Medicine, and Committee on National Statistics, National Research Council

ISBN: 0-309-59391-3, 38 pages, 8.5 x 11, (1999)

Review of Disability Evaluation Study Design

Third Interim Report

Gooloo S. Wunderlich and William D. Kalsbeek, *Editors*

Committee to Review the Social Security Administration's Disability Decision Process Research
Dorothy Rice, *Chair*
Division of Health Care Services
INSTITUTE OF MEDICINE
and
Committee on National Statistics
Commission on Behavioral and Social Sciences and Education
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY PRESS
Washington, D.C.

NATIONAL ACADEMY PRESS 2101 Constitution Avenue, N.W. Washington, DC 20418

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

The Institute of Medicine was chartered in 1970 by the National Academy of Sciences to enlist distinguished members of the appropriate professions in the examination of policy matters pertaining to the health of the public. In this, the Institute acts under both the Academy's 1863 congressional charter responsibility to be an adviser to the federal government and its own initiative in identifying issues of medical care, research, and education. Dr. Kenneth I. Shine is President of the Institute of Medicine.

Support for this project was provided by Contract No. 600-96-27893 between the National Academy of Sciences and the Social Security Administration. The views presented in this report are those of the Institute of Medicine Committee to Review the Social Security Administration's Disability Decision Process Research and are not necessarily those of the funding agencies.

International Standard Book Number 0-309-06830-4

Additional copies of this report are available for sale from the National Academy Press, 2101 Constitution Avenue, N.W., Box 285, Washington, D.C. 20055. Call (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area), or visit the NAP's home page at **www.nap.edu.** The full text of this report is available at **www.nap.edu/readingroom.**

For more information about the Institute of Medicine, visit the IOM home page at: **www.iom.edu.** For more information about the Committee on National Statistics, visit the CNSTAT home page at **www2.nas.edu/cnstat.**

Printed in the United States of America.

The serpent has been a symbol of long life, healing, and knowledge among almost all cultures and religions since the beginning of recorded history. The serpent adopted as a logotype by the Institute of Medicine is a relief carving from ancient Greece, now held by the Staatliche Museen in Berlin.

COMMITTEE TO REVIEW THE SOCIAL SECURITY ADMINISTRATION'S DISABILITY DECISION PROCESS RESEARCH

DOROTHY P. RICE (*Chair*), Professor Emeritus, Department of Social and Behavioral Sciences, Institute for Health and Aging, University of California at San Francisco

MONROE BERKOWITZ, Professor of Economics, Emeritus, and Director, Disability and Health Economics Research, Rutgers University, New Brunswick, New Jersey

RONALD S. BROOKMEYER, Professor of Biostatistics, The Johns Hopkins University School of Hygiene and Public Health

GERBEN DEJONG, Director, National Rehabilitation Hospital Research Center, and Professor of Family Medicine and Adjunct Professor, Georgetown University Institute of Public Policy, Washington, D.C.

MARSHAL F. FOLSTEIN, Chairman and Professor of Psychiatry, Tufts University School of Medicine; and Psychiatrist-in-Chief, New England Medical Center, Boston, Massachusetts

ROBERT M. GROVES, Director of the Joint Program on Survey Methodology at the University of Maryland at College Park; and Professor and Program Director, Survey Research Center of the University of Michigan at Ann Arbor

ALAN M. JETTE, Professor and Dean, Sargent College of Health and Rehabilitation Sciences, Boston University

WILLIAM D. KALSBEEK, Professor of Biostatistics and Director, Survey Research Unit, University of North Carolina at Chapel Hill

JERRY L. MASHAW, Sterling Professor of Law and Management and Professor, Institute for Social and Policy Studies, Yale University

CATHARINE C. MASLOW, Director, Initiative on Alzheimer's and Managed Care, Alzheimer's Association, Washington, D.C.

DONALD L. PATRICK, Professor of Health Services, University of Washington School of Public Health and Community Medicine

HAROLD A. PINCUS, Senior Scientific Consultant, Robert Wood Johnson Foundation, the John D. & Catherine T. MacArthur Foundation, and the RAND Corporation

JOHN A. SWETS, Chief Scientist, Information Sciences, BBN Technologies, Cambridge, Massachusetts

EDWARD H. YELIN, Professor of Medicine and Health Policy, Department of Medicine and Institute for Health Policy Studies, University of California at San Francisco

Study Staff

GOOLOO S. WUNDERLICH, Study Director
KATHLEEN NOLAN, Research Associate
NICOLE AMADO, Senior Project Assistant

BOARD ON HEALTH CARE SERVICES
INSTITUTE OF MEDICINE

COMMITTEE ON NATIONAL STATISTICS
NATIONAL RESEARCH COUNCIL

REVIEWERS

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the NRC's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making the published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their participation in the review of this report:

NORMAN BRADBURN, Senior Vice President for Research of the National Opinion Research Center (NORC), University of Chicago

JOAN LEON, Private Consultant, Berkeley, California

RODERICK J. A. LITTLE, Chair, Department of Biostatistics, University of Michigan

EDWARD B. PERRIN, Professor Emeritus, University of Washington

SEYMOUR SUDMAN, Professor, University of Illinois-Urbana

While the individuals listed above have provided constructive comments and suggestions, it must be emphasized that responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Contents

1

Introduction

The Social Security Administration (SSA) has been engaged in a major effort to redesign the process for determining disability for cash benefits and medical assistance under its Social Security Disability Insurance (SSDI), Title II of the Social Security Act and the Supplemental Security Income (SSI), Title XVI of the Social Security Act. Because of the complexity and far-reaching impact of its efforts, SSA concluded that the redesign effort requires extensive research, testing, and validation, as well as further development of some of its components before national implementation. The effect of the new determination process on the number and characteristics of future beneficiaries also needs further study. The agency asked the National Academies to provide ongoing independent and unbiased review of, and recommendations on, its current and proposed research as it relates to the development of a revised disability decision process including the approach, survey design, and content of the complex multiyear Disability Evaluation Study (DES).

The present report is the third in a series of short interim reports of the National Academies' Committee to Review the SSA's Disability Decision Process (hereafter referred to as "the committee"). The first interim report reviewed and commented on the general features of the proposed survey design, data collection plan, coverage and sampling for the DES. It provided a preliminary examination of these components as described in the scope of work in the draft request for proposal (RFP) developed by SSA for a contract to conduct a DES. In that report the committee made no attempt to comment on the content of the questionnaires, specific measures of functional capability, or the content of the medical examinations and medical and diagnostic tests proposed for the DES.

The second interim report contained a preliminary assessment of the adequacy of SSA's research plan for developing a new disability decision process and the timeline for its completion. In that context, the report outlined a framework for a research design and reviewed the general features and directions specified by SSA in the scope of work in the relevant RFPs for the conduct of the research. The report identified critical elements of a research design that were missing from SSA's plans and offered suggestions for changes in priorities and improvements in the research projects underway and others yet to be developed.

This third interim report relates directly to one of the contract tasks-review of the design, approach, and content of the DES, as proposed by SSA's contractor for the survey, Westat, Inc*. This report is limited to a brief review of the sample design (including that of the pilot study), instruments and procedures, and response rates goals developed by Westat and provided by SSA in June 1999 to the committee for its review and recommendations (Westat, Inc., 1999 a-c). It also comments on the proposed timeline for initiation of each phase of the survey.

* The committee conducted much of the work on this report through a subcommittee composed of William Kalsbeek *Chair*, Ronald Brookmeyer, Gerben DeJong, Robert Groves, and Catharine Maslow. The full committee reviewed the draft report, and subsequent revisions were made in response to comments from committee members. Thus, the report reflects the collective thinking of the committee on the issues addressed.

2

Overview of the Disability Evaluation Study

This chapter describes briefly the general features of the Disability Evaluation Study (DES) as planned by the Social Security Administration (SSA) and identifies the key survey design and sampling plan, data collection plans, and operational decisions made to date regarding the national survey and its pilot study[1].

The DES is a complex, national sample survey to estimate the number and characteristics of a broad range of people with disabilities that affect their ability to work and carry out activities of daily living. In addition to personal interviews, medical and physical examinations will be conducted on a nationally representative sample of the noninstitutional population 18-69 years of age.

The primary objectives of the DES are to:

1. Estimate the total number and characteristics of people who are severely enough impaired such that, but for work or other reasons [2], would meet SSA's statutory definition of disability. (This group would represent the universe of potentially eligible nonbeneficiaries who could apply and meet the current criteria, but who are not now receiving benefits.)
2. Identify the factors that enable persons with disabilities who could qualify for benefits to remain in the workforce.
3. Identify the variables needed to monitor and assess, in a cost-effective manner, future changes in the prevalence of disability.
4. Identify the number and types of people who could be affected by a change in the disability decision process.

[1] The information in this chapter is excerpted from the reports prepared in June 1999 by Westat, Inc. (1999a-c) for SSA.

[2] The term work for SSA's purposes refers to substantial gainful employment, which is generally about $700 per month for 1999. Other reasons include people who have chosen not to apply for disability, who have too many assets, who rely on family for support or those who are unaware of the program.

STUDY DESIGN

Sample Design

The sample design for the DES is driven by the following four core objectives (Westat, 1999b, p. 5): The design should:

1. yield a sample of the various subgroups of working-age people with severe enough disabilities to be likely to be eligible for disability benefits for SSA purposes if they applied;
2. yield a sample of the "borderline" group of people with disabilities sufficient to permit estimates of the number and characteristics of those who might become eligible, or cease to be eligible, if the current SSA disability decision criteria are altered;
3. yield a sample of people with only mild or no disabilities sufficient to permit comparisons with the population with disabilities on measures of physical and functional performance and medical conditions in the population; and
4. yield a sample of people receiving disability benefits under SSDI and or SSI.

The sample for the DES is a dual-frame, multistage, stratified probability sample design. The first stage will be a stratified sample of primary sampling units (PSU) selected with probability proportional to size. Within the PSUs, households with persons 18-69 years of age will be subsampled at rates designed to yield a nationally representative sample. Households with telephones will be selected for telephone interviews by list-assisted random digit dialing (RDD) sampling. Households without telephones will be selected for in-person interviews using standard area probability sampling methods (i.e., segments within PSUs, listing of dwelling units, sampling of households). In addition, a small subsample of households with telephones in the area sample will be selected for in-person administration of the initial screening interview. Group quarters will be chosen mainly from lists for the sample PSUs and will be sampled for in-person interviews.

Sampling Frame

PSU Definition

The first-stage sample will consist of 80 PSUs[3] selected from more than 3,000 counties and independent cities in the United States using probability proportional to size (PPS) sampling. To limit the travel distance between the respondents' homes and the mobile examination centers (MEC), Westat has decided to use individual counties as PSUs for this study instead of the typical metropolitan statistical area or groups of contiguous nonmetropolitan counties.

[3] After analyzing the effects of a reduction in the number of PSUs on precision, survey cost, data quality and other issues, SSA and Westat decided to use 80 PSUs instead of 100 as originally proposed (telephone communication on October 5, 1999).

Sampling Frame Construction

The PSUs selected for the survey will be used for both the area sample and the RDD sample. Westat plans to construct the area sampling frame from the Census Bureau P.L. 94 data tape[4] that contains housing unit counts and geographic information for every block in the United States. Blocks and block groups are uniquely associated with counties in this data tape. In the telephone sampling frame, telephone numbers within the same area code and exchange can sometimes straddle two or more counties. To solve this problem Westat plans to adopt an "unbiased" rule in which an exchange is associated with the county only if a plurality, or a specified percentage, of households in the exchange is contained within the county. Under this approach each exchange will be associated with only one county. However, the definition of the county for sampling households in the telephone frame will not match the exact geographical definition of the county. Westat expects little loss in efficiency in employing different PSU definitions for the area and RDD samples.

Measure of Size for Sampling PSUs

In sampling units of unequal population sizes (counties, in this study), it is standard practice to sample them with PPS or some other measure of size. In the DES, selection of the measure of size is driven by the need to equalize the workloads at the MECs. Using an estimate of the number of persons aged 18-69 for a county as its measure of size would lead to an equal-probability sample with approximately equal sample sizes per county for the screening interview. Westat plans to use, for that purpose, the 1997 county-level population estimates[5] available from the U.S. Census Bureau. To equalize the workloads at the MEC, these estimates will be adjusted by county-level disability statistics.

SAMPLE SIZES

The SSA has set a target to identify and complete all data collection for a total sample of 5,665 persons made up of the following four study groups or strata:

1. a "core" group of nonbeneficiaries with severe disabilities (the likely eligible group);
2. people with significant but lesser impairments (the "borderline" cases);
3. people with only mild or no disabilities; and
4. current SSDI and or SSI disability beneficiaries, who will be included primarily for the purpose of benchmarking the distinctive characteristics of the core group.

Westat assumes an initial sample size of 98,095 persons, about 89 percent of whom will be sampled using the RDD component. The initial sample sizes for the RDD component of the survey are:

[4] The tape contains basic 1990 census estimates of population down to the block level used for redistricting purposes as required under PL 94-171.

[5] Estimates are based on 1990 Census data adjusted using administrative records.

- 80 PSUs;
- 110,238 telephone numbers have residential/nonresidential status determined;
- 54,016 telephone numbers determined to be residential;
- 92,636 persons 18-69 years of age sampled in identified households; and 86,955 eligible persons for whom the initial screening interview is completed.

The initial sample sizes for the area frame component can be summarized as follows:

- 80 PSUs;
- 2,000 total area segments, 25 per PSU;
- 64,124 total occupied dwelling units canvassed, 25 per segment;
- 4,168 nontelephone households canvassed, all sampled;
- 59,956 telephone households canvassed, 3,504 sampled;
- 12,378 total persons 18-69 years old sampled, 6,189 each in nontelephone and telephone households; and
- 11,444 persons for whom in-person screening interview is completed.

As stated previously, noninstitutionalized persons residing in group quarters are included in the DES. After the selection of PSUs, Westat plans to develop PSU-wide lists of group quarters. Group quarters will be sampled from these lists within each PSU, with sampling rates designed so that people living in group quarters will have the same selection probabilities as people living in households. Group quarters for which lists are not available and those that are missed during list construction will be identified during the area sample listing process.

Sampling Methods

Stratification Variables

To ensure geographical spread and demographic diversity of the sampled PSUs, the PSU frame will be stratified by geographic region, metropolitan status, income level, minority status, prevalence of work disability using 1990 census data, and population size. In addition, because research on work disabilities has shown marked differences by geographic region, Westat is considering stratifying by prevalence of work disability to ensure appropriate representation of areas with very high and very low levels of disability.

Once the sampling strata have been specified, the largest counties will be included in the sample with certainty. Westat expects five certainty PSUs. The remaining counties then will be assigned to strata of approximately equal size. Counties too small to be sampled as individual units will be grouped to meet minimum measure-of-size requirements. Finally, two PSUs will be selected from each stratum with probability proportional to size.

Within-PSU Sampling

The second-stage sampling unit for the RDD sample is the telephone number. The sample of telephone numbers will be selected from the exchanges linked to sampled counties. The sam

ple will be unclustered within the PSUs; the number of households with telephones to be screened per PSU will be roughly equal across the noncertainty PSUs.

As indicated earlier in this chapter, the DES sample design includes an area sample of households without telephones to complement the RDD sample. To obtain a self-weighting sample of households of sufficient size for the initial screening process, about 64,124 dwelling units in the area sample will be listed and canvassed to include 4,168 nontelephone households in the sample. In the area sample, each PSU will be divided into sampling units or segments, which are defined to be census blocks. Each segment will include 25 dwelling units. Within each PSU, a sample of segments will be selected with probability proportional to the number of dwelling units. A sample of households will be sampled from within each selected segment. Assuming that nontelephone households are sampled at the same rate as telephone households, a total sample of 64,124 households will include about 4,168 nontelephone and 59,956 telephone households. Group quarters will be sampled as part of the area frame. Group quarters will be selected using probability proportional to size. The measure of size will be the best estimate of the number of residents 18-69 years of age living in the group quarter.

The Screener[6]

The main purpose of the DES screening process is to collect information needed for sampling purposes and to classify people aged 18-69 into the four target study groups noted previously. A two-stage screening process will be used. The purpose of the initial screener is to collect information needed for sampling purposes as well as information needed to classify the estimated 98,095 screener respondents into the four study groups. Following the initial screener, a subsample of respondents will be selected for the second stage screening, which will be an in-person interview. The initial screener will be mostly a telephone interview, followed by a face-to-face interview. The initial screener will include 57,712 households, and will result in interviews with about 98,095 adults 18-69 years old. It will involve three types of interviews. The majority of people will be interviewed by telephone, however a face-to-face interview using the same screening questions will be conducted with two groups. One group will include a subsample of approximately 3,609 telephone households, the other will be a subsample of approximately 4,168 nontelephone households. Westat assumes a 10 percent attrition rate for further study participation.

Of the total number of people participating in the initial screener, a sample of about 11,444 will be selected to participate in the follow-up personal interviews. The purpose of this follow-up screener is to confirm the preliminary classification of each person into one of the four categories and to more precisely classify people into severe and borderline-impaired study groups. It will also confirm suspected mental illness and cognitive deficit. Westat expects approximately 10,300 persons to complete the follow-up interview. All of these interviews will be done in person. Upon completion of the follow-up screening, sampled persons will be assigned once again to one of the four study groups, with an expected breakdown of 515 nonimpaired nonbeneficiaries; 1,545 borderline-impaired nonbeneficiaries: 3,090 severely impaired nonbeneficiaries; and 515 beneficiaries. They will then be subsampled to obtain the target sample sizes for each of the study groups.

[6] The term "screener" refers to the screening instrument for the DES.

The total target sample size is 5,665 persons, and the target sample sizes for the four study groups completing the medical examination are as follows:

1. nonbeneficiaries likely to be severely disabled (3,090);
2. nonbeneficiaries who are less disabled (1,545);
3. nonimpaired or slightly impaired nonbeneficiaries (515); and
4. persons receiving disability benefits (515).

This sample of 5,665 will then receive a medical examination.

Response Rates

Westat's assumptions about the sample size that needs to be screened to obtain the required 5,665 persons distributed disproportionately in the four strata for the various components are based on achieving the following response rates:

- 90 percent for the initial screening interview;
- 90 percent for the subsequent in-person interview and medical examination; and
- an overall response rate of 80 percent for the combined interview and medical examination components.

Assuming that these high response rates can be achieved, Westat estimates that a sample of about 98,095 persons in about 57,712 households will be sufficient to yield 5,665 persons for the DES study group.

Proxy Respondents

The issue of the use of proxies arises in this survey because a large number of people in the sample will have disabilities or some kind of functional limitation. Westat plans to avoid proxies whenever possible. However, it may be necessary to collect information from proxies to ensure the highest possible response rate and to obtain as much information as possible from people who have difficulty responding on their own.

Westat's current plans call for a household reporter to answer questions in the initial screener about all working-age adults in the household. Westat is concerned, however, that such reporters may not be able to answer accurately and honestly questions about the mental and cognitive health of other members of the household. Westat is also concerned about the risk of very low response rates if it attempts to interview each person in the household about his or her mental and cognitive health. During the follow-up screener and the comprehensive survey interview, Westat plans to use medical exam proxy assistants in interpreting for and assisting the sample person with medical needs or language problems.

Proxy interviews have varying levels of accuracy depending on the topic of the interview and the relationship of the subject to the proxy. Westat believes that the use of proxies in the initial screening process will make it over-sensitive, for purposes of the initial screener, however, that would be acceptable. Beyond the initial screener, Westat plans to avoid using proxy reporters but does expect to have proxy-assisted interviews. The decision to use or not use a proxy respondent

will be made when the sample person is initially contacted. If the respondent is available and able to complete the interview, the interviewers will be discouraged from accepting a proxy.

PILOT STUDY DESIGN

In response to a recommendation in the committee's first interim report, the DES now includes a large comprehensive pilot study prior to the conduct of the national study. The stated purposes of the pilot study are to experiment with several data collection methods and procedures, and to ensure that the questionnaires are clear and concise, and that all procedures run smoothly and efficiently, with the burden and discomfort placed on the respondent kept to a minimum. The pilot study will also test the effectiveness of the screening instruments and the accuracy of the screening algorithm; determine the best procedures for maximizing the response rates, both total and item; and develop estimates of prevalence rates to determine the final sample sizes for the main study. Finally, the pilot study will test the operational procedures for medical examinations, including the reliability of physician and nurse practitioner examinations; medical examinations performed in the home and in mobile MECs; and the logistics, reliability, and validity of the simulated disability decision process.

Current plans call for the pilot study to be conducted in eight PSUs, with a total sample of 1,000 persons in the final data collection step. A sample of approximately 13,202 households will be contacted in the initial screener. Of the 13,202 households, about 11,882 households, or about 20,316 persons, are expected to complete the initial screener. A single respondent per household will be administered the screener for all persons in the household 18-69 years of age. The screener will result in the assignment of respondents into the categories of interest for the DES. From this group, a sample of 2,000 persons will be selected for follow-up. Assuming a 10 percent attrition rate, Westat expects 1,800 persons to complete the follow-up screener and the comprehensive survey instrument. From these 1,800 persons, 1,000 will be selected for the medical examinations. Included in this 1,000 will be 200 current beneficiaries on the disability rolls, about half of them will be selected from SSA files

Focus Groups, Cognitive Laboratory Tests, and Pretests

Before the pilot study starts, Westat plans to ensure that the instruments are clear, understandable, and concise and to test all explanatory information and procedures. These developmental activities will include focus groups, cognitive laboratory tests, and pretests. These activities will continue into the pilot study to allow further improvements.

Focus Groups will be conducted to learn about four areas:

1. Four focus groups will be held to simulate the disability determination process-two will be held early in the design effort, and two will be held after half of the pilot study simulations are completed.
2. Two focus groups will be held with individuals with disabilities who are not current beneficiaries to review and discuss advance materials and materials provided after the sample person is selected.

3. Two focus groups will be held with current beneficiaries to discuss issues surrounding participation in the study.
4. Three focus groups will be held with respondents on the pretest to discuss reactions to the interviews and examinations.

Each focus group session is expected to last two or three hours.

The objectives of *the Cognitive Laboratory Tests* are to reduce response errors and to ensure that each instrument serves its purpose. Westat is planning to conduct cognitive tests early in the design process with nine individuals for each test to help identify problems and issues related to item flow, item content assessment, and length of the selected sections of the instruments. Each cognitive test session is expected to last from 45 to 90 minutes. About 50 interviews are planned and they will cover various sections of the screeners and the comprehensive survey instrument. These tests will use several techniques to capture respondents' views about the questionnaire items, including postinterview debriefing, think-aloud protocols, and paraphrasing.

Westat plans to conduct *Pretests* before and after approval is received from the Office of Management and Budget (OMB). Westat plans to conduct three sets of nine interviews, for a total of 27 interviews, before OMB approval. One of the stated purposes of these pretests is to provide an initial examination of how well the questions work when they are used as a complete protocol. The pretests also will allow an assessment of overall flow, length of interview, and interviewer-respondent interaction.

Another purpose of the pretests will be to determine which instruments provide the desired results in the follow-up screener, particularly for confirming mental illness and cognitive deficit. As Westat correctly points out, the capacity of the screener to identify and classify people with mental illness is a persistent concern with the DES. Westat plans to test the effectiveness of two options each for confirming mental illness and cognitive deficit. The Composite International Diagnostic Interview, Short Form (CIDI-SF), which is brief and can be administered by a lay interviewer, will be compared with the Structured Clinical Interview for DSM-IV (SCID) which is a highly sensitive and specific instrument that must be administered by a psychiatrist or a psychiatric social worker, to determine if the CIDI-SF can identify the type and level of severity of psychiatric illness as well as the SCID. The Mini-Mental State Examination (MMSE) will be compared with the Mental Status Examination (MSE) to determine which instrument more accurately identifies and classifies people with cognitive impairments. Participants for these tests will be obtained from SSA rolls of people with known psychiatric illness and known cognitive impairments.

After approval for the survey is obtained from the OMB, Westat plans to conduct a pretest of 50 persons representing a cross-section of the population of interest for the DES.

Tests During the Pilot Study

Westat plans to conduct several experiments during the pilot study. Response rates will be a critical factor in determining the final sample sizes for the main study. To increase response rates, Westat plans to conduct four experiments with different data collection methods, four with refusal conversion incentive strategies, and one with medical providers. Another area of experimentation regarding improved response rate relates to current beneficiaries. Three alternative consent forms will be tested in the pilot study for this purpose.

Westat plans to test several process and methodological issues related to medical examinations. Some of these issues concern the efficiency of the MEC examinations, coordination and communication of the administrative and interviewing office and examination center, participant satisfaction, layout of the MEC, functioning of the equipment, data tracking, reporting, and documentation, efficiency of the process of referrals to outside medical providers, and efficiency of the home examination. The pilot study also offers an opportunity to test the comparability of medical examinations administered by physicians and by nurse practitioners, the comparability of examinations given by nurse practitioners at the respondents' homes and in the MECs, and the determination of when home examinations should be offered. A sample of 30 persons will be asked to be examined three times: twice in the MEC-once by a nurse practitioner, once by a physician-and then again in their homes by a nurse practitioner.

The pilot study is also designed to test instrument designs for the DES and more thoroughly test the screens and questionnaires. The tests concern the screener methods used to allocate the general population into the four study groups and will attempt to answer these questions: Does the screening process misclassify individuals? Is the screener sensitive to the full range of people with mental illness? Is the screener sensitive to people with cognitive deficits? The resulting classification analysis will address the sensitivity and specificity of the screening process. Pretests and laboratory testing of the instruments prior to the pilot study, help concentrate on issues of item wording, response options, item sensitivity, sequencing, and flow. Westat plans to fully test the questionnaires in the pilot. Westat believes that the large sample in the pilot study affords the opportunity to conduct item analyses, which could provide clues to items in the questionnaires that respondents routinely do not answer or that offer little or no variability in response. Such information could not be obtained in small pretests. Alternative formats of the content and length of the various sections of the screener and questionnaires also will be tested in the pilot study. The work status section is a case in point.

Westat's proposed plans for the DES include a comprehensive series of tests and experiments covering all aspects of the survey operations, design, response rates, and the effectiveness of the questionnaires before the start pilot and during the pilot study.

3

Committee Reactions

The previous chapter gave a brief overview of the design and procedures for the Disability Evaluation Survey (DES) as proposed by Westat, the survey contractor to the Social Security Administration (SSA). This chapter highlights some of the committee's reactions to the design and its recommendations for improvements.

The committee has reviewed the reports prepared by Westat on the survey procedures, the design of the DES (including that of the pilot study), and the strategies to maximize response rates. The committee commends SSA and Westat for the tremendous amount of work they have done and the progress they have made in designing this complex study. Both SSA and Westat have heeded many of the recommendations made in the committee's first interim report. Their plans are an improvement over the preliminary design for the study reviewed by the committee in 1997 (IOM, 1997). The reports they have prepared are thoughtful, and contain sound ideas; in addition, they have addressed the correct issues for a survey design.

The committee, however, believes that the DES cannot adequately address all of the stated objectives using only a personal or proxy report of the factors related to the person's work environment. The complex interaction between environmental factors external to the individual (such as social, geographic, and economic influences) and the perceptions of the individual also should be studied.

In its first interim report, the committee reviewed and commented on the preliminary design developed by SSA for the DES. In that report, *the committee strongly endorsed the conduct by SSA of a well-designed, carefully pretested, and statistically sound DES. The committee has not changed its position; in fact, it reemphasizes that position*.

The committee, however, continues to be concerned that some of the key problem areas, and the recommendations for resolving them that were identified in 1997, are not addressed adequately in the reports prepared by Westat. In its 1997 report on the DES design, the committee had recommended that the period prior to and including the pilot study should be expanded and extended into a research, development, and testing phase for the survey, with applications to samples of the type that are more traditionally used in methods testing. The national survey should not be launched until the functional assessment instruments, survey operations, and other issues are fully developed, tested and resolved.

The single most important issue surrounding the DES as currently developed and the biggest deficiency in its design is the truncation of the research and development phase that the committee had strongly recommended. The current schedule provides no flexibility in terms of the available time to correct problems and to test alternative solutions if some questions or procedures do not work in the pilot study (Table 3.1).

The schedule developed for the conduct of the various phases of the survey and the time allotted to absorb and retool for the main survey will not permit deliberate and rigorous decisions about revisions of the design, procedures, or questionnaire content. Decisions will have to be made throughout the process, and there is insufficient time to resolve issues and test alternatives. SSA and Westat are aware of these uncertainties. *The committee believes that unless these issues are resolved before the national study, SSA may be taking enormous risks that could compromise the scientific integrity of the national survey and its findings*.

PILOT STUDY

Westat has developed extensive plans for testing and experiments to be conducted before and during the pilot study, but the short period of time allowed for research, development, and testing in the rush to launch the national survey will cause serious logistical inflexibility during the various phases of the survey. *Before starting a national survey, sufficient time should be allowed (a) to conduct and analyze the results of the various pretests, focus groups, and cognitive tests; (b) conduct a comprehensive pilot study with the planned and other built-in experiments; and (c) analyze and test alternative solutions in areas that need resolution as a result of the pilot study*. Under the current time frame, the pilot study may not yield much evaluative information and not enough time is allowed to make modifications in the national study after alternatives are tested. As far as the committee can gather from the documents it reviewed, the pilot study may only show that "the design works" or "the design does not work." Westat will not know if the alternative selected will work in the national study without further tests and evaluation of the alternatives. Unfortunately, the present schedule for the pilot study allows just about 2-3 months from the end of the pilot study (November 2000) to the start of data collection for the main study (January 2001). Although some decisions on instrumentation can be made prior to the end of the pilot study, sample sizes and results will not be sufficient to allow a thorough analysis of issues until close to the end of data collection phase in the pilot study. Even if analysis of some tests and experiments could begin earlier in the analysis phase of the DES pilot study, additional time will be needed to examine the implications and plausibility of several different "adjustments" in the problem areas.

Westat needs to explain to SSA exactly what it plans to do when problems arise in the pilot study. There should be some advance plans about how they will identify the obvious problems, assess the risks they pose, and then what they propose to do about these problems. Moreover, given the complexity of the DES, once all the issues are resolved, Westat should consider conducting a dress rehearsal before starting the national study. *The documents reviewed do not explicitly state that Westat plans to conduct a dress rehearsal and no time is allocated for it in the current timetable*.

TABLE 3-1. Timeline for Survey Activities Proposed by Westat

	Time Schedule	
Survey Activity	No. of Months After OMB Approval	Calendar Month After OMB Approval
OMB Approval Received		October 1 (est.)
Pilot Study		
Print survey instruments	1	October
Conduct listing activities and administer initial screener	2-11	Nov 99-Aug 00
Conduct interviewing and medical exam activities, including targeted referrals	2-12	Nov 99-Sept 00
Obtain medical records	3-13	Dec 99-Oct 00
Ongoing data preparation and processing	3-14	Dec 99-Nov 00
Data analysis and preparation of simulated case folders	5-14	Feb 00-Nov 00
Methodological Report 1: Evaluation of Pilot Study Methods	14-15	Nov 00-Dec 00
Submit protocol revisions to OMB	15	December 2000
Methodological Report 2: Instruments and Procedures for the DES Main Study	15-16	Dec 00-Jan 01
Main Study		
Print revised survey instruments	16	January 2001
Conduct listing activities and administer initial screener	16-26	Jan 01-Nov 01
Conduct interviewing and medical exam activities, including targeted referrals	16-27	Jan 01-Dec 01
Obtain medical records	17-28	Feb 01-Jan 02
Ongoing data preparation and processing	17-29	Feb 01-Feb 02
Data analysis and preparation of simulated case folders	18-29	Mar 01-Feb 02
Descriptive Report 1: Disability Profile of Working-Age Americans: Preliminary Data from the DES	30	March 2002
Descriptive Report 2: Employment Profile of Working-Age Americans: Preliminary Data from the DES	32	May 2002
Descriptive Report 3+: other reports as specified	28-31	Jan 02-Apr 02
Analytic Report 1: Estimates of Trends in SSDI and SSI	30	March 2002
Analytic Report 2: Relationships Between Disability and Work	30	March 2002
Analytic Report 3: Evaluation of Disability Data Collection Questions and Procedures	31	April 2002
Analytic Report 4: The Effects of Disability on Retirement	32	May 2002
Analytic Report 5: Disability Decision Process	32	May 2002
Final project report	33	June 2002

NOTE: OMB = Office of Management and Budget.
SOURCE: data provided by Michele Adler, SSA, June, 1999.

PERFORMANCE OF THE INITIAL SCREENER AND ACHIEVEMENT OF TARGET SAMPLE SIZES

An example of the inadequate time for developmental work is Westat's plan to conduct initial screening to classify people into four disability-status categories. As indicated in the previous chapter, a key element of Westat's plan for the DES is its intent to meet respondent sample size targets for four disability-status categories (with sample size targets for a completed examination indicated in parentheses):

1. non-beneficiaries likely to have severe disabilities (3,090);
2. non-beneficiaries who have less severe disabilities (1,545);
3. non-beneficiaries with no or slight disabilities (515); and
4. people currently receiving disability benefits (515).

The capacity of the screener to work, and for Westat to achieve the response rates needed, will have serious implications in terms of achieving the targets set for the sample sizes for each of the categories. In addition to levels of survey nonresponse at the various steps of the DES design and the extent of differential attrition among key population segments, the ability to hit these targets is heavily dependent on the capacity of the initial screener to correctly classify people screened by telephone or in person. In the current design, Westat does not know if the screener will be sensitive enough to classify the people in the sample in one of the four categories or if the categories themselves will be feasible. Also clouding this issue is the operational interpretation of being "nonimpaired," as well as the qualifiers "severely," "less severely," and "slightly."

The committee is concerned that Westat may not have an adequate opportunity within the present timeframe of the DES contract to adequately deal with category misclassification that is likely to occur with the initial screener. Westat's plan for DES instruments and procedures indicates that a "major objective of the pilot will be to analyze the proposed cut-points (to delineate the four groups) and make adjustments for the main study" (Westat, Inc., 1999a, pp. 2-7). Unfortunately, as indicated above, the present timeline for the pilot study allows just a short period from the end of data collection in the pilot study to the start of data collection for the main study. Even if analysis of the initial screener could begin earlier in the analysis phase of the DES pilot study, additional time will be needed to examine the implications and plausibility of several different "adjustments" that might need to be made as a result of the analysis; to examine the specific effects of attrition on the screening phase of the DES; and to seek the best available national data on the distribution of the four groups that is needed to establish post screening rates of disproportionate subsampling.

In the committee's judgment, the addition of several months to the period between the end of data collection in the pilot study and the start of data collection for the main study will allow the screening and attrition findings of the pilot study to be more carefully considered and used to ensure that sample size targets for the four groups are met. Without this extension, the adequacy of the job done by the initial screener may well become an important issue, in that an adverse outcome could poses a serious risk to the success of the DES. *The committee, therefore, urges Westat to allow enough time for it to carefully examine (a) the implications and plausibility of several different "adjustments "; (b) specific effects of attrition on the screening phase of the DES, and (c) the best available national figures on the distribution of the four groups*.

Targeted Sample Sizes

The committee continues to have several questions and concerns about the rationale for the ability to achieve, as well as for the adequacy of the total sample size, and especially of the allocation of people among the four subgroups. It is not clear how SSA and Westat arrived at this particular disproportionate sample design. The committee has never seen the statistical rationale for setting the sample size targets. What precision targets were used to arrive at these sample sizes? Westat assumes that at response rates of 90 percent for each component of the DES, it should get the planned sample sizes. *The committee believes that the rate is overly optimistic, especially for a population with disabilities*. Even if Westat can achieve these planned sample sizes, the cells very likely will be much too small, especially if SSA stratifies for analytical purposes on more than one disabling condition, and or demographic and socioeconomic characteristics such as age, gender, minority status.

The committee raised these issues in its first interim report (IOM, 1997); it reemphasizes the problems that could arise as a result of sample selection, size, and allocation if adequate advance planning and testing are not undertaken.

SSA should produce a clear explanation of how and why it came up with the sample size targets, particularly for the four study groups. Such an explanation should include: (a) an explicit statement of SSA's plans for the analysis of DES data (with priorities indicated regarding information goals and population subgroups for generating domain estimates); (b) statistical precision and or power requirements for these analyses (particularly those of highest priority); and (c) a clear statement of how they arrived at their target sample sizes based on (a) and (b). Further SSA should require Westat to devote sufficient time after the pilot and based on the pilot data to conduct power calculations to assess in more detail whether the main survey is adequate in size. Based on the results, Westat should re-evaluate and revisit the numbers it originally set as targets, and if necessary, to assess Westat's ability to meet any new targets.

MEDICAL EXAMINATION SCHEDULE

The scheduling of medical examinations in the Mobile Examination Centers (MEC) is another example of logistical inflexibility resulting from the tight timeframe for the survey. The timing of the exams is so tight that there would be little flexibility if several respondents required more time to complete the examination. In addition, Westat is using the National Health and Nutrition Examination Survey (NHANES) model for the MEC part of the study, but it is not clear from the documents the committee reviewed if any tests were undertaken to assess whether the NHANES model will work for people with disabilities. As noted in the previous chapter, Westat plans to expend significant effort in testing and experimenting with different options related to the efficiency of processes and methods before and during the pilot study. *The committee commends Westat's plans but does not think Westat can appropriately conduct these needed tests, analyze the results, test alternatives, and make the necessary modifications within the current timeframe*.

PROXY RESPONDENTS

Westat has addressed many of the concerns about the use of proxies in surveys and has identified useful mechanisms and procedures for interviewing and examining the survey participants; in addition, it plans to use proxies only when unavoidable for the comprehensive interview and the follow-up interview. The committee continues to be concerned about the effect of the use of proxies on the validity of responses. Westat recognizes that the use of proxies is a critical issue for any household survey. It has heightened importance for the DES because the target populations are likely to have difficulty responding due to their impairments, resulting in the potential for higher levels of proxy responses and proxy involvement throughout the administration of the survey (Westat, Inc. 1999c). In addition, Westat must be able to deal with the divergence of responses from proxies and from the sample persons on questions that may lead to misclassification of participants in one of the four categories for the subsequent phases of the survey. For example, proxy responses in the initial screener interview may lead Westat to classify the person as severely impaired, whereas the respondent may be found to be mildly or moderately impaired during the follow-up screener and medical examination.

To reduce respondent burden during the initial screener, a "household reporter" will respond for all working-age adults in the household. Westat is concerned about the validity of data, especially those collected from a household reporter, and particularly about the ability of this reporter to respond to questions about the mental and cognitive health of the other individuals living in the household. It intends to experiment with alternative approaches in the pilot study. The experiments identified include (a) attempting to conduct mental and cognitive interviews with all working-age adults in the household; (b) collecting information from the reporter and then interview each individual; and (c) using the interviews for mental and cognitive ability (e.g., MMSE and CIDI) with the household reporters, asking them to answer these questions for the other adults in the household, and then selecting those who test positive on the interviews, based on the reporters' responses for the follow-up screener and the comprehensive interview. *Adequate time is required to conduct these experiments, analyze the data and make changes based on the findings*. Moreover the training needed for the CIDI-SF and the SCID is time consuming, as is the documentation and evaluation of quality control. Sufficient time will be needed after the pilot study to compare the two systems and to develop and monitor the assessment.

Proxies have been shown to be valid reporters for much of the information that is sought in the initial screening instrument (e.g., Epstein et al, 1989; Bassett, et al., 1990; Long, et al., 1998; Sneeuw et al, 1997, 1998). The screener has been designed so that the various mental and cognitive exams will only be administered to the respondent because they cannot be accurately or reliably completed by proxies. However, in addition to difficulties of detecting mental disability through proxies, research also has shown differences in proxy versus respondent reports in other areas, including a tendency for proxies to overreport or underreport the severity of disability (Epstein, et al, 1989; Hays, et. al., 1995; Grootendorst, et al., 1997; Magaziner et. al, 1997).

QUESTIONNAIRE AND PROCEDURES ISSUES

Questionnaire Development

Much of the work on questionnaire development and procedures for data collection could be tested with small samples using different approaches, including the laboratory research. In

creasingly, survey operations, especially for large-scale surveys, are being preceded by small-scale, carefully structured, field testing coupled with laboratory research, so that definitions, questionnaire wording and the order effects of the questions, response burden on the respondents, and concepts critical to the results of the survey can be better understood.

Questionnaire pretesting can involve a number of different methods. One qualitative method often used in questionnaire pretesting is the cognitive interview in which a respondent is interviewed one-on-one by a trained questionnaire design specialist. The respondent is asked to think aloud as he develops his answers. This technique can be useful in detecting problems with questionnaire items such as retrieval difficulties (Sudman, et al., 1996). Another approach is to ask more directed probes after the respondent has answered the question. These probes can help assess the respondent's comprehension of some term or concept (Forsyth and Lessler, 1991). These techniques also can be helpful in developing successful probes to be used with open-ended questions. While understanding how respondents arrive at their answers will be important, the answers themselves will also provide invaluable information in the development of response options for some questions. The responses given by cognitive interview participants should provide some idea of what responses to expect for these questions.

The committee is pleased to note that Westat's plans include field testing and cognitive testing, followed later by a larger pretest and then a large pilot test. *Westat's detailed plans for testing and experimentation also should include the applicability of the NHANES mode of medical examination for people with severe disabilities and functional assessment instruments*. Moreover, in the course of the workshop titled "Measuring Functional Capacity and Work Requirements" held in 1998, participants generally agreed that there is no one instrument available to assess the functional capacity to work that could be incorporated into the DES or into the disability decision process. Many participants were of the opinion that more research and experimentation are needed in this area (IOM, 1999).

The committee notes that Westat has incorporated many measures and procedures to include people with mental or cognitive impairments. However, the committee is especially concerned about the applicability of the questions and procedures to this group of people. For instance, some people with chronic mental illness and traumatic brain injury are able to get jobs but cannot retain the jobs for more than a few days for various reasons related to their condition. These individuals are likely to be eligible or borderline eligible for disability benefits under SSDI and or SSI. *The work status questions in the various sections of the DES generally do not address this situation; they need to be reviewed to ensure that these questions are appropriately worded and tested using cognitive processes and then again in the pilot study*.

Another example is the use of the term "work." People understand the meaning of this wordterm work differently, and people with disabilities are often encouraged to think of sheltered work and activity programs as "work." For purposes of the DES, it is important to differentiate these different situations. Although this distinction is made in many places in the current DES questionnaires, all sections of the DES questionnaires should be reviewed to ensure that every time respondents are asked about work, it would be clear whether their answer refers to regular work, sheltered work, or an activity program.

Respondent Burden

Each of the DES survey instruments is lengthy and complex, thus creating a risk that respondents will be unwilling or unable to provide useful data to SSA (see Table 3.2). For example, SSA has noted that that the Comprehensive Survey Interview will impose a burden on some respondents who have a complicated medical history, considerable income or assets, and a complex work history. The committee agrees and expects that other DES components will also impose a significant burden on those and other respondents. Another concern is the initial screener, because its results will be used to sort individuals into the four categories. For this screener, one household member will be asked to respond to numerous questions, including questions about mental and emotional problems, for all household members 18-69 years of age. If the respondent does not answer these questions correctly for all household members, individuals who have conditions that should result in their selection for the follow-up screener may be missed.

Because of its length and complexity, SSA and the committee agree that the instrument will have to be pared down between the end of the pilot study and the start of the national study. SSA first must decide which questionnaire items are to be eliminated, and then the shortened version must be evaluated and field-tested to ensure its viability as an instrument that is able to meet the study's goals. These steps will take several weeks or months to be done well. The committee believes that the current DES contract does not allow sufficient time for these steps.

SUMMARY

The pilot design generally looks reasonable, but the committee seriously doubts that enough time is allotted to determine what changes are needed to implement those changes before the conduct of the national survey. The type of research and experimentation described in the previous sections that is planned to be conducted before and during the pilot study would facilitate final resolution of the various issues identified, and appropriate changes could then be made in the design and procedures before the start of the main survey. In order to resolve the problem areas in a satisfactory manner more time will be needed especially between the completion of the pilot study and the start of the national study. The current plan allows 2-3 months to assess the findings of the pilot study prior to the start of the main study. Given the potential for adverse outcomes in the current design of the pilot study, this amount of time for reflection and response is likely to be wholly inadequate.

Recommendation: The committee strongly recommends that SSA revise the project schedule to allow significantly more time to plan and analyze the pilot study and test alternative solutions for problem areas before starting the national study.

While the committee is unable to prescribe specific time needs, a number of plausible scenarios may provide some guidance to SSA on the amount of additional time that may be required for the DES timeline. One possibility is an inability to meet response rate targets set for the screening, interview, and examination phases of data gathering under *each* of the data collection options and incentive levels proposed by Westat. This outcome could easily require 3-4 additional months to develop and field test alternatives to achieve these targets. In another scenario,

development of the initial and follow-up screeners may require one or more additional iterations to better achieve sample size targets for disability status groups. These iterations could add several months to the completion of the pilot study phase of the survey. Finally, planned quantitative assessments of examination instruments may reveal unacceptably low levels of reliability and validity. This situation would demand 4-5 additional months for another round of instrument redesign and field testing before SSA could comfortably implement this phase of data gathering in the national study. Simultaneous occurrence of all of these scenarios could force a delay of a year or more, during which time the entire design of the DES is revisited and new survey strategies are devised and tested. The committee recognizes that increasing the time and level of research between the pilot study and the national survey may have cost implications.

TABLE 3-2. Components of Survey Instrumentation for the Disability Evaluation Study

Initial Screener: To be conducted primarily through telephone interviews with one household member who will be asked to answer questions for all household members age 18-69.

- Household enumeration
- Functional limitations
- Health status
- ADLs and IADLs
- Current recipients of Social Security disability benefits
- Work status

Follow-Up Screener: To be conducted in person with each subject selected into the sample.

- Household enumeration
- Functional limitation scales
- ADLs and IADLs
- Mental status examination
- Composite International Diagnostic Interview 12-month short screening scales
- Performance measures (actual tasks to be completed by the subject)
- Perceived disability

Comprehensive Survey Interview: To be conducted in person with each subject in the sample.

- Health behaviors
- Medical history
- Access to health, educational, and vocational services
- Social and community living
- Work status, history, and environment
- Program knowledge
- Economic resources

Medical Examination: To be conducted in person with each subject in the sample.

- Core medical examination
- Cardiovascular system module
- Endocrine system module
- Gastrointestinal system module
- Genitourinary system module
- Hemic and lymphatic system module
- Immune system module
- Musculoskeletal system module
- Neoplastic disease module
- Neurological disease module
- Respiratory system module
- Dermatology system module
- Performance measures, including special senses, speech, and actual tasks to be completed by the subject. **Procedures for Collecting the Medical Evidence of Record** D**isability Decision Process**

NOTE: ADL = activities of daily living; IADL = instrumental activities of daily living.
SOURCE: Westat, 1999a.

4

Concluding Comments

The Disability Evaluation Study (DES), if well designed could be the cornerstone for long-term disability research. It will be of fundamental importance to future analyses by the Social Security Administration (SSA) and other researchers. It will provide information that would guide SSA in making decisions about its disability programs and will play a key role in projecting and understanding disability rolls in the future. Moreover, it will lay the groundwork for future surveys. The committee is gratified to note that SSA decisionmakers have given high priority to research and to policy development based on that research. The DES is key to how SSA will deal with the population with disabilities in the future. Until now, SSA has mainly focused on streamlining the claims process by making it more efficient and less time consuming. To ensure effective planning, SSA must explore the fundamental characteristics of who is disabled, how many more people might become disabled, and what can be done to assist people to remain in the workforce. The SSA has not collected such basic information for nearly 20 years and it is long overdue.

The committee, however, has serious reservations about the timeframe for the conduct of the survey. The current plan provides little flexibility in terms of the amount of time available to make deliberate and rigorous decisions on issues of design, procedures, and questionnaire if problems are uncovered during the pilot study. The 3 months allowed between the end of the pilot study and the start of the national survey is clearly inadequate for the kind of analysis and further testing that will be needed to resolve any issues that may arise. The committee understands that the documents reviewed in preparing this report are "works in progress" and that Westat and SSA are already addressing many of the issues raised in this report. Nevertheless, the committee believes that unless the period for testing, analysis, and development is extended, SSA could encounter serious problems during the national survey.

References

Bassett, S.S., Magaziner, J., and Hebel, J.R. Reliability of Proxy Response on Mental Health Indices for Aged, Community-Dwelling Women. *Psychology of Aging* 5(1): 127-132, 1990.

Epstein, A.M., Hall, J.A., Tognetti, J., Son, L.H., and Conant, Jr., L. Using Proxies to Evaluate Quality of Life. Can They Provide Valid Information about Patients' Health Status and Satisfaction with Medical Care? *Medical Care* 27(3 Suppl.):S91-S98, 1989.

Forsyth, B.H., and Lessler, J.T. Cognitive Laboratory Methods: A Taxonomy. In Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., and Sudman, S., eds., *Measurement Errors in Surveys*. New York: Wiley, 1991.

Grootendorst, P.V., Feeny, D.H., and Furlong, W. Does it Matter Whom and How You Ask? Inter-and Intra-Rater Agreement in the Ontario Health Survey. *Journal of Clinical Epidemiology* 50(2): 127-135, 1997.

Hays, R.D., Vickrey, B.G., Hermann, B.P., Perrine, K., et al. Agreement Between Self-Reports and Proxy Reports of Quality of Life in Epilepsy Patients . *Quality of Life Research* 4(2): 159-168, 1995.

Institute of Medicine (IOM). *Disability Evaluation Study Design: First Interim Report*. Institute of Medicine. Washington, D.C., National Academy Press, 1997.

Institute of Medicine (IOM). *Measuring Functional Capacity and Work Requirements: Summary of a Workshop*. Institute of Medicine. Washington, D.C., National Academy Press, 1999.

Long, K., Sudha, S., and Mutran, E.J. Elder-Proxy Agreement Concerning the Functional Status and Medical History of the Older Person: The Impact of Caregiver Burden and Depressive Symptomology. *Journal of the American Geriatric Society* 46(9): 1103-1111, 1998.

Magaziner, J., Zimmerman, S.I., Gruber-Baldini, A.L., Hebel, J.R., and Fox, K.M. Proxy Reporting in Five Areas of Functional Status. Comparison with Self-Reports and Observations of Performance. *American Journal of Epidemiology* 146(5):418-428, 1997.

Sneeuw, K.C., Aaronson, N.K., de Haan, R.J., and Limburg, M. Assessing Quality of Life After Stroke. The Value and Limitations of Proxy Ratings. *Stroke* 28(8): 1541-1549, 1997.

Sneeuw, K.C., Aaronson, N.K., Sprangers, M.A., Detmar, S.B., Wever, L.D., and Schornagel, J.H. Comparison of Patient and Proxy EORTC QLQ-C30 Ratings in Assessing the Quality of Life of Cancer Patients. *Journal of Clinical Epidemiology* 51 (7):617-631, 1998.

Sudman, S., Bradburn, N.M., and Schwartz, N. *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass Publishers, 1996.

Westat, Inc. Disability Evaluation Study-Instruments and Procedures: Task 4, Report 1. Submitted to Social Security Administration, Washington, D.C., June 1999a. (Unpublished.)

Westat, Inc. Disability Evaluation Study-Final Sample Design Report: Task 4, Report 2. (Includes Pilot Study Design.) Submitted to Social Security Administration, Washington, D.C., June 1999b. (Unpublished.)

Westat, Inc. Disability Evaluation Study-Plans to Meet Response Rate Goals: Task 4, Report 3. Submitted to Social Security Administration, Washington, D.C., June 1999c. (Unpublished.)

Appendix

REVIEW OF THE SOCIAL SECURITY ADMINISTRATION'S DISABILITY DECISION PROCESS RESEARCH

STUDY MANDATE

The study will review and provide advice on the scope of work, design, content of the survey, and the approach and scientific methods of completed and planned research as the Social Security Administration (SSA) develops the new disability decision process. The study will focus on the population 18-69 years of age. Although the committee is given latitude in setting its own agenda and designing its plan of work, the topics it explores will include:

- Review of the research plan and timeline for developing a new decision process for disability;
- Review of the preliminary design of the Disability Evaluation Study (DES) research efforts, the scope of work for the DES, and the design and content of the survey, as proposed by the survey contractor, as well as SSA's plans to integrate the decision method and DES research effort, identifying statistical design, methodological and content concerns, and other outstanding issues;
- Examine the results of completed research including research into existing functional assessment instruments and subsequently identified research for SSA's redesign efforts, and provide advice for adopting or developing functional assessment instruments or protocols for the redesigned disability process and the DES in particular; and
- Assess the results and findings of the research undertaken by SSA, comment on future research proposals, and offer advice on the analysis of the consequences of alternative disability determination processes. Some of the topic areas that might be considered include functional assessment of work-related limitations of physical and mental impairments; disability decision processes (including screening mechanisms); testing and validating decision processes for determining disability; and age, education, and work experience.

Acronyms and Abbreviations

ADL	Activities of Daily Living
CIDI-SF	Composite International Diagnostic Interview, Short Form
DES	Disability Evaluation Study
IADL	Instrumental Activities of Daily Living
MEC	Mobile Examination Centers
MMSE	Mini-Mental State Examination
MSE	Mental Status Examination
NHANES	National Health and Nutrition Examination Survey
OMB	Office of Management and Budget
PPS	Probability Proportional to Size
PSU	Primary Sampling Unit
RFP	Request for Proposal
RDD	Random Digit Dialing
SCID	Structured Clinical Interview for DSM-IV
SSA	Social Security Administration
SSDI	Social Security Disability Insurance
SSI	Supplemental Security Income

Biographical Sketches of Committee Members

Dorothy P. Rice, Ph.D. (*Chair*) is Professor Emeritus of Social and Behavioral Sciences at the School of Nursing, University of California at San Francisco (UCSF) and holds joint appointments at the Institute for Health and Aging and the Institute for Health Policy Studies at UCSF. From 1983 to 1994, she was Professor-in-Residence at UCSF. Previously she served as Director of the National Center for Health Statistics and was Deputy Assistant Commissioner for Research and Statistics at the Social Security Administration. Professor Rice's major research interests and expertise include health statistics; survey research, design, and methods; disability; chronic illness; and the economics of medical care. She has achieved national and international renown for her leadership role, extensive research, and scholarly publications. Professor Rice has received numerous awards including an honorary Doctor of Science from the College of Medicine and Dentistry of New Jersey. She is a Fellow of the American Public Health Association and the American Statistical Association, and a member of the Institute of Medicine.

Monroe Berkowitz, Ph.D., is Professor Emeritus of Economics and Director of Disability and Health Economics in the Bureau of Economic Research at Rutgers University. He has served as a consultant to various government agencies including the Social Security Administration, the World Health Organization, and the American Association for the Advancement of Science. Dr. Berkowitz is a leading authority on the economics of disability and rehabilitation in public programs (SSA disability insurance and worker's compensation), private disability insurance, and public and private rehabilitation systems; and has conducted extensive comparative analysis of foreign systems. He is a member of the National Academy of Arbitrators, the National Academy of Social Insurance, the American Economic Association, and the Industrial Relations Research Association.

Ronald S. Brookmeyer, Ph.D., is Professor of Biostatistics and Epidemiology at the Johns Hopkins University School of Hygiene and Public Health. He has been a Visiting Biostatistician at the National Cancer Institute and the International Agency for Research on Cancer in Lyon, France. Dr. Brookmeyer's research interests and expertise are in statistical modeling and methodology, biometrics, and epidemiology. He is the recipient of the Spiegelman Gold Medal awarded by the American Public Health Association for contributions to health statistics. He is a

Fellow of the American Statistical Association and the American Association for the Advancement of Science, and a member of the Biometrics Society and the Society for Epidemiological Research.

Gerben DeJong, Ph.D., is Director of the National Rehabilitation Hospital Research Center and Professor of Family Medicine and Adjunct Professor at the Georgetown University Institute of Public Policy. Prior to coming to Washington, D.C., he served as Associate Professor in Rehabilitation Medicine at the Tufts University School of Medicine. Dr. DeJong has a special interest in managed care's impact on medical rehabilitation-people with disabilities and other vulnerable populations; health outcomes measurement, and medical ethics. He is probably best known for his seminal work on disability and health policy and the independent living movement. Dr. DeJong was a Fulbright Scholar in the Netherlands on the research staff of the Social Security Council. He is a member of the American Congress of Rehabilitation Medicine, the Association for Health Services Research, and the National Academy of Social Insurance.

Marshal F. Folstein, M.D., is Chair and Professor of Psychiatry at Tufts University School of Medicine and Psychiatrist-in-Chief at the New England Medical Center (NEMC). Prior to joining NEMC, he was Eugene Meyer III Professor of Psychiatry and Medicine at the Johns Hopkins Medical Institutions. His expertise and research interests are in neuropsychiatry, disability research, and Alzheimer's Disease. Dr. Folstein created the Mini-Mental State Examination, widely used for assessing cognitive mental status in medical patients and in population surveys. He is a Fellow of the American College of Physicians, the American Psychiatric Association, and the Gerontological Society; and a member of the American Neurological Association and the Society for Epidemiological Research.

Robert M. Groves, Ph.D., is a Professor of Sociology and Research Scientist at the Institute for Social Research at the University of Michigan, and is Director of the Joint Program in Survey Methodology, based at the University of Maryland, a National Science Foundation-sponsored consortium of the University of Maryland, University of Michigan, and Westat, Inc. From 1990 to 1992, Dr. Groves was an Associate Director of the U.S. Census Bureau, on loan from Michigan. He has over 25 years of experience with large scale surveys, and has investigated the impact of alternative telephone sample designs on precision, the effect of data collection mode on the quality of survey reports, causes and remedies for nonresponse errors in surveys, estimation and explanation of interviewer variance in survey responses, and other topics in survey methods. His current research interests focus on theory-building in survey participation and models of nonresponse reduction and adjustment. He is a fellow of the American Statistical Association, an elected member of the International Statistical Institute, former President of the American Association for Public Opinion Research, and currently Chair of the Survey Research Methods Section of the American Statistical Association.

Alan M. Jette, Ph.D., is Professor and Dean of Boston University's Sargent College of Health and Rehabilitation Sciences, and Professor of Social and Behavioral Sciences at the Boston University School of Public Health. His previous appointments have included: Chief Research Scientist, New England Research Institute; Associate Professor, Massachusetts General's Institute of Health Professions; and Assistant Professor, Division on Aging, Harvard Medical School. Dr. Jette's research interests include measurement, epidemiology, and prevention of disability and

the critical evaluation of treatment outcomes in the medical and rehabilitation fields. He has developed several disability outcome instruments, widely used in health services research in the United States and abroad. Dr. Jette recently directed several health services research projects focusing on disability prevention, home care and geriatric rehabilitation.

William D. Kalsbeek, Ph.D., is Professor of Biostatistics and Director of the Survey Research Unit at the University of North Carolina-Chapel Hill. His prior experience includes statistical research with the Office of Research and Methodology at the National Center for Health Statistics and at the Sampling Research and Design Center at the Research Triangle Institute in North Carolina. Dr. Kalsbeek's research interests and areas of expertise are in biostatistics, survey design and research, spinal cord injuries, and assessment; and is well known for his work in survey methods. He is a Fellow of the American Statistical Association, and a member of the Biometrics Society and the American Public Health Association.

Jerry L. Mashaw, LL.B., Ph.D., is Sterling Professor of Law and Management and Professor at the Institute of Social and Policy Studies at Yale University. He is a leading scholar in administrative law and has written widely on social insurance, social welfare issues, and disability policy. Dr. Mashaw recently chaired the National Academy of Social Insurance's Disability Policy Panel. He is a Fellow of the National Academy of Arts and Sciences and founding co-editor of the Journal of Law Economics and Organization.

Catharine C. (Katie) Maslow, M.S.W., is Director of the Initiative on Alzheimer's and Managed Care at the Alzheimer's Association. Prior to this, she was at the U.S. Office of Technology Assessment (OTA), and has experience in public welfare, mental health, and nursing home settings. Her research and consumer interests include aging, disability, criteria for long-term care, client assessment, and Alzheimer's Disease. Ms. Maslow is a member of the National Association of Social Workers, the American Public Health Association, the Gerontological Society of America, and the American Society on Aging.

Donald L. Patrick, Ph.D., M.S.P.H., is Professor of Health Services and Director of the Social and Behavioral Sciences Program at the University of Washington School of Public Health. He holds adjunct appointments in epidemiology, sociology, and rehabilitation medicine and is a senior investigator at the University's Center for Disability Policy and Research and the Northwest Prevention Effectiveness Center. He is also Director of the U.S. Field Centre for the World Health Organization quality-of-life measures. Dr. Patrick's research interests and expertise are in health services, public health policy for people with disabilities and older adults, and quality-of-life assessment. He is a Fellow of the Association of Health Services Research, and a member of the American Public Health Association, the British Society of Social Medicine, and the Society for Disability Studies. He was the inaugural president of the International Society for Quality of Life Research and is a member of the Institute of Medicine.

Harold A. Pincus, M.D., serves as a senior scientific consultant for the Robert Wood Johnson Foundation, the John D. and Catherine T. MacArthur Foundation and the RAND Corporation. Dr. Pincus was the Deputy Medical Director of the American Psychiatric Association (APA) and founding director of the APA's Office of Research. He is Adjunct Professor of Psychiatry and Behavioral Sciences at Duke University Medical Center, a Clinical Professor of Psychiatry and

Behavioral Sciences at George Washington University, and a Clinical Professor of Psychiatry at the Uniformed Services University of Health Sciences, F. Edward Hebert[Hébert] School of Medicine. He has led major health policy and services research and training projects, and co-directs the Practice Research Network, a practice-based psychiatric research network. His research interests are in the relationships between mental health and general medical care; the diagnosis, classification, and treatment of mental disorders; and functional assessment and rehabilitation. Dr. Pincus is the 1997 recipient of the William C. Menninger Memorial Award of the American College of Physicians for distinguished contributions to the science of mental health.

John A. Swets, Ph.D., is Chief Scientist for Information Sciences at BBN Technologies in Cambridge, Massachusetts, a lecturer at Harvard Medical School on health care policy, and Senior Research Associate in Radiology at the Brigham and Women's Hospital. His research interests are behavioral modeling and analysis, specifically in applied signal detection theory to human perception and decision-making. Dr. Swets' theory created a new paradigm for the study of human sensory systems and addressed new areas in psychology and medicine. He is a member of The National Academies (NAS); the National Research Council's Board on Behavioral, Cognitive, and Sensory Sciences; and the immediate past chair of the NAS Commission on Behavioral and Social Sciences and Education. Dr. Swets is a Fellow of the American Association for the Advancement of Science, the American Psychological Association, and the American Psychological Society.

Edward H. Yelin, Ph.D., is Professor of Medicine and Health Policy at the University of California, San Francisco, where he has primary academic appointments in the Department of Medicine and Institute for Health Policy Studies. He is also the Director of the Arthritis Research Group at UCSF. Dr. Yelin's research interests concern the impact of managed care on persons with chronic conditions and disability and employment problems among persons with disabilities. He has over 80 publications in these areas, including *Disability and the Displaced Worker* (Rutgers University Press). Dr. Yelin is a member of the American Public Health Association and American College of Rheumatology. He has received many academic awards, including the Distinguished Scholar Award from the Association of Rheumatology Health Professionals.