

SURVEY DESIGN AND DATA COLLECTION ISSUES IN THE DISABILITY EVALUATION STUDY

Carla Maffeo, William Frey, Graham Kalton, Westat
Carla Maffeo, Westat, 1441 West Montgomery Ave. Rockville, MD 20850

Key Words: Disability, National Survey, Disability Screening, Social Security, Work Disability

Introduction

The Social Security Administration (SSA) provides cash assistance and medical benefits through its disability insurance program to Americans who are unable to work due to a medically determinable impairment. The program currently spends \$51 billion each year in monthly benefits to 6.4 million workers and their families.

Since the early 1980s, SSA's disability program has experienced unprecedented growth in the number of Americans applying for benefits as well as in the number of persons determined eligible for benefits. During this period, the program has also experienced fewer terminations and longer stays on the rolls as technological advances result in people living longer with severe disability. Taken together, these trends have policy analysts concerned about the ability of the program to meet the future needs of disabled workers. Consequently, the Social Security Administration and the Congress commissioned the National Study of Health and Activity (NSHA) to explore the nature and extent of disability in the United States. The NSHA is designed to examine the following questions:

1. What is the prevalence of severely disabled persons in the United States who, but for work status, would be eligible for disability benefits?
2. What enables some severely disabled persons to remain in the workforce?
3. Which self-reported measures are good predictors of work disability?
4. What is the impact of changing the age for retirement benefits on the Disability Program?

In order to answer these questions, the DES will collect information from a representative sample of persons in the U.S. who will be screened to establish disability status and subsampled for an extensive interview and medical examination, which will be conducted in mobile exam centers (MECs). Medical records will be obtained for examined persons to supplement the medical exam data. To determine which study participants would be eligible for the

program, except for work status, the SSA disability determination process will be simulated. Disability folders will be developed from the survey and medical record data and sent to disability examiners for review. The decision will be appended to the data file to be used in analysis.

The definition and process for establishing eligibility for disability benefits through the disability insurance program is both lengthy and complex. Applicants must present medical and work history evidence that supports their claim. A disability claim examiner follows a 5-step process in adjudicating the evidence. The process includes the following steps:

1. Establish that the claimant is no longer working, *and*
2. Establish that the claimant's impairment is medically determined, severe, and is expected to last 12 months, *and*
3. Establish the presence of a medical listings level impairment (resulting in automatic program eligibility), *or*
4. Establish the residual functional capacity does not allow the claimant to continue in the same line of work, *and*
5. Establish that the claimant cannot do other work on the national economy.

This paper examines five major challenges to achieving the goals of the NSHA. These challenges include screening to identify persons who are potentially eligible for disability benefits, efficient sampling of these persons, attaining acceptable response rates, conducting useful medical examinations, and constructing simulated disability folders. The issues surrounding each challenge and plans to resolve them are discussed. An extensive pilot study is underway to test responses to each challenge.

1. Screening for Disability

The current plan for the NSHA is to select participants through a dual stage screening process. Data from an *initial telephone screen* (entitled household screen or HS) with a single knowledgeable household reporter will be used to classify all household members (aged 18 to 69) into four study

categories (current SSI/SSDI beneficiary, severely impaired nonbeneficiary, moderately impaired nonbeneficiary, not impaired). A predetermined sampling fraction will be applied to each category in order to sample persons who will then participate in an *in-person interview* (entitled SP interview). The interview is intended to confirm the preliminary HS classification.

The identification of persons with serious mental illness presents a particular challenge to the screening methodology. This challenge lies primarily with the initial screen, where the episodic nature of some serious mental illness could easily result in false negative classifications and where there is a paucity of proven telephone screen strategies.¹ Unless this population can be effectively identified at the household level, there will be little or no opportunity to confirm the identification and provide a clinical and functional profile during the followup screen using available screening methodologies.

In response to this challenge, we have planned a two-level assessment system for mental illness imbedded in the two screens. The goal in the HS is to be highly sensitive to the population of persons with mental illness. The number of false negative identifications should be minimized, while considerable tolerance for false positives is acceptable. The concern here is to cast a large net for persons with mental illness, from which any false positives may later be removed by more specific screening conducted face-to-face by well-trained non-clinician interviewers. In the SP interview, the goal will be to implement a process that is also sensitive to this population, yet more specific. This task requires a more extensive set of questions. At this second step, there is still reason to tolerate some number of false positives, keeping in mind the possible tradeoffs between cost and the risk of failure to identify the target population.

1.1 Composition of the Household Screen Component for Mental Illness

Persons with mental illness, like those with physical illness, can have a wide range of functional limitations. However, the effects of mental illness may

¹ Mental illness among the homeless and institutionalized populations is not addressed in this paper. The NSHA is not expected to include either of these populations in the national probability sample. However, it should be recognized that projections of the size of these populations impact the expected incidence in the national probability sample.

be overlooked if measures of functioning are restricted to those that reflect the impact of musculo-skeletal or cardiopulmonary disease (e.g. mobility, lifting). For this reason, several measures of functioning were included in the HS because they are expected to be sensitive to the effects of mental illness on disability. Each set of items is discussed in more detail below.

Instrumental Activities of Daily Living (IADLs and Activities of Daily Living (ADLs). Six activities are included as IADLs in the HS: preparing meals, shopping for personal items, managing money, doing light work around the house, doing heavy work around the house and managing medications. Another six ADLs form a part of the HS, including getting around in the home, getting in/out of bed, dressing, taking a bath/shower, using the toilet, and eating

In an analysis of Baltimore ECA data, Bassett et al. (1998) examined disability among people with no psychiatric disorders and those with one of several psychiatric diagnoses. Included among the psychiatric disorders were cognitive impairments (e.g., dementia, delirium, amnesia, mental retardation); substance abuse, schizophrenia, mood disorders, adjustment or anxiety disorders, and personality disorders. Compared to persons with no psychiatric disorders, difficulty in IADL tasks (which included ability to get around in own neighborhood, keep track of money and bills, clean house, prepare meals, use telephone) was significantly higher for persons with a cognitive impairment, schizophrenia, mood disorders, or adjustment/anxiety disorders. In addition, ADL difficulty (including ability to get around in your home, get in and out of bed, dress and undress, take a bath or shower, use the toilet, use a knife and fork to cut up meat) was significantly higher for three of these groups (all but mood disorders). These authors also found the rate of labeling oneself as disabled was significantly higher among persons with cognitive, substance abuse, schizophrenia and personality disorders.

Problems with Interpersonal Interactions. Two items are included in the HS: 1) PERSON has trouble making or keeping friendships, and 2) PERSON has a lot of trouble getting along with other people in social or recreational settings. Barker et al. (1989) found high percentages of persons with serious mental illness who reported difficulty making and keeping friendships -- 46% overall, 60% to 70% among those with work limitations.

Severe Disability in Role Functioning Related to Mental Health. Two items are included in the HS: 1)

During the past 12 months, has (PERSON) accomplished less than (PERSON) would have liked to as a result of emotional problems, such as feeling depressed or anxious? 2) During the past 12 months, did PERSON not do work or other activities as carefully as usual as a result of any emotional problems, such as feeling depressed or anxious?

Bassett et al. (1998) found higher rates of being kept from usual activities over the previous 3 months among persons with cognitive, substance use, mood, and adjustment/anxiety disorders compared to persons with no psychiatric disorders. In a cross-cultural study, Ormel et al., (1994) found higher disability levels among persons with psychiatric disorders using a seven-item disability scale. Their scale included measures of physical functioning (climbing stairs, bending, lifting or stooping) as well as measures of social role functioning (the person reduced or stopped doing activities they once did, and the person had been unable to do things that the family expected as part of daily routine).

Health Conditions (Diagnoses). In addition to measuring areas of functioning, we decided that it was necessary to include a set of mental illness health conditions as a “pick list.” This was done as a precautionary measure in the pilot study to determine whether a subset of persons with mental illness would not report problems with daily functioning. Therefore, a list of diagnoses similar to that used in other national household surveys (the 1989 Mental Health Supplement to the NHIS, the NHIS-D) is included in the HS. The first question asks the following:

During the past 12 months, has PERSON been told by a doctor or mental health professional that PERSON has any of these mental or emotional conditions: schizophrenia, paranoid or delusional disorder, manic episodes or manic depression also called bipolar disorder, major depression, obsessive-compulsive disorder, severe personality disorder, severe anxiety disorder such as panic disorder or phobia, Alzheimer's disease or other dementia, severe eating disorder, alcohol abuse disorder, drug abuse disorder.

A second question asks:

During the past 12 months, did anyone in the family have any other mental or emotional condition? Include only those conditions that seriously interfere with PERSON's ability to work

or attend school, or to manage PERSON's everyday activities?

Although grouped with diagnoses, this latter question also taps role functioning related to mental health.

1.2 Composition of Followup Screen in the SP Interview

Individuals selected to participate in an SP interview receive as part of that interview a followup screen as well as a lengthy series of questions on work, income, health providers, social support and independent living.

The screening part of the interview consists of the entire HS (done for the purpose of evaluating proxy/self-report bias), and three standardized screening instruments to scale respondents' mental status, cognitive ability, and physical ability. The standardized measures include the Composite International Diagnostic Interview, Short Form (CIDI-SF), Mini-Mental State Examination (MMSE), and 8 physical performance measures. Each measure is discussed in more detail below.

Composite International Diagnostic Interview, Short Form. The CIDI-SF is a recently developed brief version of the CIDI, designed specifically to screen for mental illnesses in the National Health Interview Survey (NHIS), a survey sponsored by the National Center for Health Statistics (Kessler et al., 1997). The CIDI-SF was constructed using the stem questions and a subset of branch questions from the full CIDI. While further validation research is currently underway, Kessler reports a high degree of correspondence between the caseness predictions of the CIDI-SF and the full CIDI for each of nine clinical syndromes. The syndromes included in the validated basic CIDI-SF are major depressive disorder, generalized anxiety disorder, agoraphobia, panic attack, drug dependence, social phobia, simple phobia, alcohol dependence, and the recently added obsessive compulsive disorder. The average administration time for the NHIS version of the CIDI-SF is about 10 minutes.

However, problems with validity exist that led to the exclusion of three syndromes from the full CIDI version. All three are of interest in the NSHA—bipolar disorder, dysthymia, and nonaffective psychosis. In the case of nonaffective psychosis, the basic problem is the relatively high rate of false positives resulting (Kendler

et al., 1996) from use of the CIDI, as compared to clinician diagnosis. We have attempted to address this problem by augmenting the CIDI-SF with a five-item psychosis screening scale developed by Bebbington. The five items include the following:

- Over the past year, have there been times when you felt happy without a break for days on end?
 - Was there an obvious reason for this?
 - Did your relatives or friends think it was strange or complain about it?
- Over the past year, have you ever felt that your thoughts were directly interfered with or controlled by some outside force or person?
 - Did this come about in a way that many people would find hard to believe, for instance, through telephathy?
- Over the past year, have there been times when you felt that people were against you?
 - Have there been times when you felt that people were deliberately acting to harm you or your interests?
 - Have there been times when you felt that a group of people was plotting to cause you harm or injury?
- Over the past year, have there been times when you felt that something strange was going on?
 - Did you feel it was so strange that other people would find it very hard to believe?
- Over the past year, have there been times when you heard or saw things other people couldn't?
 - Did you at any time hear voices saying quite a few words or sentences when there was no one around that might account for it?

Mini-Mental State Exam: The MMSE is a well-studied clinical and research tool with population-based norms (Crum, Anthony, Bassett, Folstein, 1993). It has been used extensively to provide a brief, standardized measure of global cognitive functioning. The MMSE offers the NSHA screen a brief method for ensuring that persons with moderate to severe cognitive impairment are accurately classified into the appropriate study groups. While the MMSE does not provide specific categorical diagnoses, performance has been associated with the many mental conditions identified in the Medical Listings, including mental retardation and dementia. Strong associations have also been linked to schizophrenia, depression, and various physical impairments.

A major shortcoming of the MMSE is its lack of sensitivity to certain types of cognitive deficit

(Anthony, LeResche, Niaz, Von Korff, Folstein, 1982). While the MMSE appears to reliably identify persons with dementia or mental retardation, it is less reliable in capturing the performance of persons with some types of memory or concentration losses. Thus, many persons with traumatic brain injury or some forms of amnesia may be missed by the screener and either lost to the study or, worse yet, misclassified into the borderline or nonimpaired categories. In an effort to address these concerns, the Mental Status Examination (Schretlen, 1998) is being considered to replace the MMSE.

Physical Performance Measures. The performance measures include a series of eight standardized tasks. All of the tasks have been used in numerous research studies. Each task was assessed to ensure its relevance to work disability. The performance measures provide a structured review of upper and lower extremity functioning. The measures include upper extremity abilities (manual dexterity, hand strength, hand and arm movement, lifting, and carrying) and lower extremity abilities (stooping, bending, and kneeling; movement from sitting to standing positions; standing; walking). With the exception of climbing, all physical Residual Functional Capacity (RFC) concepts used by SSA in disability determination are covered by the physical performance measures.

The CIDI-SF is the primary method employed by the NSHA to identify persons with mental illness. However, we also capture a small portion of persons through the combined measures of cognitive ability, lower caseness probability CIDI-SF scores, the psychosis screening items augmenting the CIDI-SF, and the performance measures.

A number of screener-related methodological issues are being examined using pilot study data. These include:

- A comparison of SP self-reports in the initial screener with SP self-reports in the followup screener;
- A comparison of household reporter reports in the initial screener with SP reports in the followup screener;
- An examination of initial screener sensitivity to disability;
- An examination of coverage of impairment groups;
- The ability of the IS and FS to delineate the impairment groups;

- A determination of whether the IS identifies the beneficiaries;
- A comparison of the self-report screen with diagnostic tests.

2. Sampling Issues

The DES requires a national probability sample of the noninstitutional population aged 18 to 69 years old in the coterminous US. The major issue for sample design is the rarity of the population of prime interest; that is, persons who would be eligible for the SSA disability insurance program (excluding considerations of work status) but who are not current beneficiaries. Existing data suggest that this group comprises about 3 to 4 percent of the population aged 18 to 69. A large-scale screening phase is needed to generate an adequate sample size for this group.

The DES will therefore employ a three-phase sample design. At the first phase, a large sample of households will be screened to classify household members aged 18 to 69 as either current beneficiaries or non-beneficiaries, and to subclassify the latter group according to the likelihood that they would qualify on health grounds for SSA benefits. The HS is administered to a single knowledgeable household member who reports for all household members in the specified age range. The subclassification is likely to comprise three groups that have been labeled for sampling purposes as severely disabled non-beneficiaries, moderately disabled nonbeneficiaries and nondisabled nonbeneficiaries; however, this subclassification may be collapsed to two groups in the final design, depending on the ability of the screening instrument to distinguish between the severely and moderately disabled groups. The development of a screening instrument that performs this classification effectively is a major challenge for the DES, given the multidimensional nature of disabilities (including both physical and mental disabilities) and the need to keep the instrument as short as possible (see below). Based on the screener classification, sample persons will be subsampled for the second phase, with all of the group of severely disabled non beneficiaries and subsamples of the other groups being included. The second phase consists of an in-person interview, comprising a follow-up screener on disabilities and a questionnaire about economic background, work history, etc. Current plans call for a reclassification of disability status based on the followup screener, with subsampling at differential rates for the third phase, the medical examinations

conducted in the MECs and the collection of medical records.

Both face-to-face and telephone interviewing are being considered for the first phase screener and have been tested in pilot study. If face-to-face interviewing is used, standard multistage area probability sampling methods will be employed. The major difference from most household area probability sample designs is that the primary sampling units (PSUs) need to be smaller in geographic area than those generally used. This requirement occurs because sample persons have to travel to the MECs for their medical examinations, and the time they spend traveling must be kept acceptably short. For this reason, like the National Health and Nutrition Examination Survey, the DES will use individual counties as the PSUs (with a few exceptions), not combinations of counties as is usual practice. A standard multistage stratified sample of PSUs will be selected using probability proportional to size (PPS) sampling, where the measures of size will be based on the latest population estimates and may incorporate an allowance for variation in disability rates across areas. Within each sampled PSU, a sample of segments will be selected by PPS sampling using data from the 2000 Census. Lists of dwellings will be compiled for the selected segments and dwellings will be sampled for the first phase screening. In view of the rarity of the population of primary concern, large samples of dwellings will be taken from selected segments. A possible modification to the above design is that group homes may be sampled separately within the selected PSUs because of the concentration of disabled persons in many of these homes. With this approach, group homes would be selected from lists of homes for the sampled PSUs where available, and homes on those lists would be excluded from the area sample.

If telephone screening is employed, the household screening sample will be selected by list-assisted RDD methods. Unlike the sample designs for most telephone surveys, the DES screener sample will need to be restricted geographically to allow for the subsequent phases of data collection. Also, the noncoverage of a telephone sample needs to be addressed, since a higher proportion of disabled persons than of the general population live in non-telephone households. With telephone screening, the sample design for DES would involve a sample of PSUs as for the area sample described above, the linking of telephone exchanges to those PSUs in general on a plurality basis, and the selection of households from those exchanges by list-assisted RDD methods. A

supplement sample of non-telephone households would be selected by area sampling for face-to-face screening in the sampled PSUs. Some telephone households would also be screened from the area sample to provide evidence on the comparability of telephone and face-to-face screening.

3. Improving Response Rates

A number of methods to maximize response rates were tested in the pilot study for the NSHA. Response rates among three alternative data collection methods were compared. In the first method a sample was selected using random-digit dial procedures (RDD). Initial screening was done by telephone with nonresponse followup in-person for SPs for whom addresses could be obtained. The followup screener, interview, and exam were conducted primarily in a site office and the medical exam center (MEC). The second method consisted of selecting a sample using RDD procedures, screening by phone with nonresponse followup in person where possible, followup screener and interview in the home and the exam conducted in the MEC or in the home, if desired. The third method involved selecting an area probability sample, conducting the initial screener by telephone for cases where telephone numbers could be obtained and in-person for the others, conducting the followup screener interview in the home and the exam in the MEC or in the home, if requested by the SP.

Various incentive experiments were conducted with SPs and medical care providers to assess the impact of varying incentives on response rates.

Four PSUs were selected and assigned to one of two sampled types, Random Digit Dial (RDD) and Area Probability (AP). In order to improve response in the RDD sample, addresses were obtained for about 50% of the sample by matching to Axcion, a database of names, addresses and telephone numbers. Refusal conversion letters that included \$5.00 were sent to households where addresses were obtained and in-person refusal conversion was also attempted. In the AP sample, telephone numbers were obtained for about 50 percent of the sample, and initial screening was attempted by phone for this group.

In order to provide diversity in the data collection experience, two rural PSUs, one suburban/urban, and one suburban PSU were included in the pilot sample. Two different incentive levels for study respondents were tested: in two PSUs respondents

received \$30 at the time of interview and \$30 at the completion of the exam; in the other two PSUs, respondents received \$50 at interview and \$50 at exam. Advance letters were sent to all households in the AP sample and to the portion of the RDD sample for which addresses could be obtained.

Finally medical providers were contacted for subjects' medical records. This provided an opportunity to test three levels of remuneration to providers: \$0, \$36, and \$100. Providers were randomly allocated to one of these treatments within each PSU.

Table 1 is provided to demonstrate the magnitude of the data collection. It does not provide all necessary information to compute response rates. Pilot study data collection (screening, interviewing and exams) was conducted in four PSUs from March, 2000 through September, 2000. Medical record retrieval and simulated decisions will be conducted from August 2000 through November 2000. Table 1 shows the number of cases completed for each data collection activity.

Table 1. Number of Cases Completed at Various Stages of Data Collection²

Number of household screeners completed	5,189
Number of SP interviews ³	791
Number of exams completed ²	409
Number of simulation folders sent for review	409

Preliminary Findings from the Pilot Study. Since data processing, medical record retrieval and simulated decisions are ongoing, only very preliminary results are available for some response rate questions. Screener, interview and exam data are available from only two PSUs. However, some preliminary general conclusions can be made.

The screener response rate in the RDD samples was low, despite the use of both standard and non-standard followup techniques. Besides, standard followup phone calls, an attempt was made to obtain addresses for the respondents and send a refusal conversion letter that included a \$5.00 incentive. Because addresses could be obtained for only about

² Includes eligible and ineligible households.

³ After sub-sampling and non-response.

one-half the sample, the improvement in the response rate was small, less than 10 percent.

The initial screener response rate was substantially higher (about 25 percentage points) in the area sample. Because we had addresses for the entire sample, refusal conversion efforts could be conducted in-person with the entire sample. While this is a more costly approach, the gain in response is quite large.

Interestingly, response rates for the followup screener and interview and the exam were about the same in the RDD and area probability samples, although the overall response rate was much higher in the area sample because of the higher initial screener rates.

Our initial examination of response rates by rural/urban status shows higher rates in the rural areas than in the suburban or urban areas. While this is not unexpected since other interview surveys have shown the same thing, it should be noted that we only had four PSUs to compare, and so caution should be exercised in generalizing.

As described earlier, initial screening was conducted by telephone when numbers could be obtained and in-person for cases where no telephone numbers could be obtained in the area probability sample. There was some concern that follow-up screener/interview and exam response rates might be lower for the group where initial screening was done by phone since refusal conversion might be more difficult without initial "personal contact." Preliminary examination of follow-up screener/interview and exam response rates does not show a difference between the group initially screened by phone and the group initially screened in person, however. Given this, the main study will likely employ this approach to reduce costs.

4. Medical Examination Issues

The purpose of the medical examination was twofold. First, it had to yield information that was useful to the disability examiners making the simulated disability decision; and secondly, it had to yield information that would be useful in projecting the number of persons likely to become disabled over time. In the true application and decision process, the applicant is evaluated in a series of steps, one of which involves examining medical information for the applicant to see whether s/he fits the levels of disability described in the Social Security "listings" of medical

impairments. Westat reviewed the listings and developed the exam so as to provide the necessary information.

The exam had to be directed or targeted in some way to make it manageable. The range of tests, measures and exams that could be performed on a study participant was broad since the participant did not really present with a medical complaint. Conceivably it could cover all body systems and any disease that results in impairment. In order to limit the exam appropriately, information collected in the medical history and a brief core exam conducted on each participant was used to select additional tests and procedures for a particular body system. For example, if someone reported shortness of breath on climbing stairs, he or she received the cardiovascular module and the pulmonary module but not the musculoskeletal module.

There were a few tests that could not be performed in the exam center because either the equipment was too expensive or they required a specialist to conduct. These included extensive vision, hearing and pulmonary function testing. Persons falling below a designated level in the preliminary test conducted in the exam center were asked to go for further testing. Blood specimens, x-rays and ECGs conducted in the exam center were sent to labs for reading and a notification/referral procedure was established for persons in cases where a medical problem was found during the exam.

Exams were conducted in specially equipped Mobile Exam Centers (MEC) which traveled from PSU to PSU. It was decided to use MECs rather than local physicians or medical centers because this approach afforded more control over the process both from a quality and consistency of data perspective and from a survey operations perspective, the timeliness of completing the work. If subjects preferred, exams could be conducted in the home using portable equipment. The exam team consisted of a physician, nurse practitioner and x-ray technologist.

During the development of the exam there was concern about the implications of several of the methods that were used in the data collection. Questions were raised about the reliability/comparability of the nurse practitioner and physician's data and about the home and MEC exams. Experiments were imbedded into the pilot study to answer these questions and the findings from these experiments will guide our thinking for the main survey.

There was also concern about how to set the test limit results to trigger further testing because the "listings" required medical evidence that could not be obtained with testing procedures in the MEC. For example, vision tests were performed in the MEC. If results showed a corrected loss of 20/40 or greater, the participant was sent to a referral ophthalmologist for a fundus exam that required dilation of the eye and extensive visual field testing. At issue was whether we were referring too many participants and consequently finding that few of the results would warrant an allowance by the disability program or too few participants and therefore missing necessary information to make the decision.

Another exam issue that is being examined concerns participants who were classified as unimpaired in the screener yet reported something in the questionnaire or had a score in a performance test that triggered tests in the MEC beyond the core exam. We believe the algorithm used to determine further testing may be too sensitive or, perhaps, the questionnaire items are too sensitive.

5. Building Simulated Disability Folders

A primary objective of the DES is to determine the number of individuals in the general population who meet the Social Security definition of disability. The principal means of accomplishing this objective is to simulate the SSA disability decision-making process using data collected in the DES. Consequently, it is important that the simulated process is consistent with the process used by SSA.

The process is called "simulated" since in the DES, there are no true applicants for disability benefits only survey participants. In the true process, an individual who has a severe disability that keeps him from working files an application with SSA alleging a disability. He also provides the necessary medical documentation from treatment sources as well as vocational, financial, education and employment information. If the information that is supplied is incomplete, the Disability Determination Service (DDS) claims examiners acquire the additional missing information from medical consultants, former employees, schools, and other sources over what is often a several month period of time.

Given that the DES was to be a cross-sectional survey that did not contemplate any recontact of participants or other data sources, it was necessary to

design a data collection process that included forms and procedures that collected information usually collected by the DDS claims examiners. This one-time survey would have to provide all the necessary data usually collected in an iterative fashion over time.

In order to make the disability decision, examiners need several kinds of information. There must be evidence of a medically determinable impairment that limits an individual from working and information on the duration and severity of the impairment. Other information on vocational training, employment, finances and functioning must also be available. The data collection design called for information about vocational training, finances and employment to be collected by interview. Physical and cognitive performance tests were designed to measure functioning. Information on diagnosis, impairment, severity and prognosis (duration) were collected by medical exams, other medical tests, and laboratory tests and by obtaining, where appropriate, participant medical records from medical providers.

Using data from the interviews, performance tests, exams, and medical records, disability folders were developed for each examined study participant. The folder contained the forms typically completed during the process of application and acquisition of evidence by DDS staff. Folders were sent to the DDS offices where a decision was made.

Prior to sending folders to the DDS examiners several steps were taken to test the completeness of the folder design and its usability and to ensure that the necessary information was being captured. Two cognitive laboratories were held with claims examiners from several DDS offices to review examples of "mock" folders. Revisions were made, new "mock" folders developed and reviewed by DDS "experts." Finally, DDS staff who would be making the decisions were trained on the study procedures and differences in the process were discussed. For example, DDS examiners were told that current employment should not be considered in the decision for the study.

An evaluation of the simulated process is on-going. Of concern is whether the simulated process is as reliable as the actual process and whether the simulated decisions are accurate. In order to assess the reliability of the process, a second examiner in the DDS is evaluating a sample of folders. SSA has data on agreement between DDS examiners for the actual determination process and these will be compared with

findings from this reliability study. This work is in process.

Validity is much more difficult to assess. The optimal approach would be to include a group of recent applicants—both allowed and denied cases—in the study and then compare the study decisions with actual decisions. However, SSA felt it was not possible to “salt” the study sample with denied cases. Somewhat less rigorous, but what is turning out to be very informative, tests of validity are being employed instead.

The study sample has been salted to include recent, current beneficiaries, which allows study decisions to be compared with the actual decisions for this group. DDS examiners have been asked to complete a brief questionnaire for each case which asks, among other things, what further information they would have needed to make a decision, what information was most useful, and whether the folder format was problematic. SSA’s quality control unit will review a sample of the study decisions and indicate whether or not they concur with the decision.

Several problems have already been identified from the review of beneficiaries’ folders and from the questionnaires being completed by the DDS staff. Information that documents the diagnosis and medical impairment is not always available either because the exam did not include the appropriate tests or the medical records were not obtained from the appropriate treatment source. Similarly duration or prognosis information was not always available. Finally, information that permits the linking of the impairment and the functional loss was not always available.

These problems did lead to rethinking the medical exam and obtaining medical records. It is clear that the medical history, formally taken by the interviewers would need to be elaborated on by the physician to obtain more information on impairments and how they limit the participant. The exam will focus more on medical impairments and tests necessary to provide the evidence for presence of and severity needed by the examiners. Finally the process for obtaining medical records will be re-designed to collect records that provide the diagnosis of the impairment and duration information, even if this means going back in time a number of years and to more treatment sources. This will likely require a review by an exam staff member to be sure that contact information and permission forms have been collected for all relevant medical providers.

Conclusions

Since data are still being examined from the pilot study, any conclusions made at this point must be considered preliminary and may change. However, several design changes are under consideration subject to pilot study findings, including:

- Selection of an Area Probability sample rather than an RDD sample;
- Initial screening done by telephone for cases where telephone numbers can be obtained;
- A one stage, rather than two stage, screening process;
- Eliminating the physical portion of the medical exam for persons who report only mental health problems;
- Revising the criteria for sending someone for a referral exam; and
- Substantial shortening of data collection instruments.

References

- Anthony, J.C., LeResche, L., Niaz, U., VonKorff, M.R., Folstein, M.F. (1982). “Limits of the ‘Mini-Mental State’ as a screening test for dementia and delirium among hospital patients,” *Psychological Medicine* 12, 397-408.
- Barker, P.R., Manderscheid, R.W., Hendershot, G.E., Jack, S.S., Schoenborn, C.A., Goldstrom, I., (1992) *Serious Mental Illness and Disability in the Adult Household Population: United States 1989*. Advance data from vital and health statistics; no 218. Hyattsville, Md: National Center for Health Statistics.
- Bassett, S.S., Chase, G.A., Folstein, M.F., and Regier, D.A. (1998) “Disability and psychiatric disorders in an urban community: measurement, prevalence and outcomes,” *Psychological Medicine* 28: 509-517.
- Crum, R., Anthony, J., Bassett, S., and Folstein, M.F., (1993). “Population-based norms for the Mini-Mental State Examination by age and educational level,” *Norms of Mini-Mental State Examination*. 269, (18) 2386-2391.
- Kendler, K.S., Gallagher, T.J., Abelson, J.M., and Kessler, R.C., (1997). “Lifetime prevalence, demographic risk factors, and diagnostic validity of nonaffective psychosis as assessed in a U.S. community

sample: The National Comorbidity Survey,” *Archives of General Psychiatry* 51, 8-19.

Kessler, R.C., and Frank, R.G., (1997). “The impact of psychiatric disorders on work loss days,” *Psychological Medicine* 27: 861-873.

Ormel, J., VonKorff, M., Ustun, T.B., Pini, S., Korten, A., Oldehinkel, T. (1994). “Common mental disorders and disability across cultures: Results from the WHO collaborative study on psychological problems in general health care,” *Journal of the American Medical Association* 272:1741-1748