

ISSUES IN DESIGNING THE MEDICAL EVALUATION STUDY FOR DISABILITY

Carla E. Maffeo, Graham Kalton, Leyla Mohadjer, Westat Inc.; Mitchell P. LaPlante, University of California; Martynas A. Ycas, Social Security Administration
Carla E. Maffeo, Westat, 1650 Research Blvd., Rockville, MD 20850

Key Words: Disability Measurement, Functional Status, Impairment

1. INTRODUCTION AND OBJECTIVES

The Social Security Administration (SSA), Office of Research and Statistics, is planning to conduct the Medical Evaluation Study (MES) to estimate the number and characteristics of persons in the working age population who have a disability that may qualify them for benefits under the Social Security Disability Insurance (SSDI) or Supplemental Security Income (SSI) Programs. The MES is being developed in response to substantial increases in the applications and awards for disability benefits since 1983, a corresponding decrease in the average age at time of award, and an increase in the duration of benefits.

The objectives of the MES include: (1) estimating the total number and proportions of selected subgroups of the U.S. working age population who are disabled, with particular attention to those who are not currently receiving disability benefits; (2) analyzing the impact of changes to the SSA disability definition; (3) identifying disability-related variables not currently included in the disability determination process that might be included in future definitions; and (4) examining differences in factors that might mediate the disability process among disabled persons who are and are not working.

A critical consideration for the survey is how to define and measure disability. Thus section 2 briefly reviews the literature on the disability process and measures of disability that have been developed in previous research. Section 3 describes the current SSA disability determination process, survey methodology, and operational issues that were also considered in the selection of disability measures for the survey.

Interest on the part of SSA in studying disabled nonbeneficiaries and the fact that disability is a relatively rare event in the population led to a proposal to use a large-scale population screening for disability. Section 4 explores the implications of this kind of approach for data collection.

2. MEASURING DISABILITY: AN OVERVIEW

Probably the two conceptual frameworks cited most frequently in current disability literature are those of Nagi (1991) and Wood (1980) for the World Health Organization.

While there are important differences in the many conceptual frameworks that have been offered, there are also commonalities and patterns that appear across them. Generally, disability is considered to include four stages: (1) disease, (2) impairment, (3) disability or functional limitation and (4) handicap.

Stage 1 describes a medical or mental health condition. Stage 2, impairment, describes a loss in an anatomical, physiological, mental, or emotional structure or function. Measures that have been developed for disease and impairment include symptoms, signs, and physiologic measures. The U.S. National Health Interview Survey (NHIS) uses condition and symptom checklists to measure prevalence of disease by self-report. Limitations of such measures, including the inaccuracy of self-reported medical conditions and the underestimation of the prevalence of medical/mental disorders, are discussed by Zola (1993). Studies, such as the Health and Nutrition Examination Surveys (National Center for Health Statistics, 1994) and others, use medical/physical/psychological examinations and physiologic measures, such as analyses of biologic or pathologic specimens, and medical tests, such as electrocardiograms, to measure the prevalence of disease and/or impairment. Recent measures of the presence of a mental health condition or impairment include the Structured Clinical Interview for DSM-III-R (Spitzer, 1990). Kessler et al. (1995), have developed a short set of screening scales (CIPI) for the DSM-III-R that can be administered by interviewers.

Impairments that cannot be or are not corrected may result in a restriction or limitation of an individual's ability to perform activities in a normal manner, Stage 3 in the conceptual models. Functional limitation is typically measured in one of two ways: (1) by determining capacity, that is, asking persons what they can do, or (2) by determining performance, asking persons what they do or actually asking them to perform a series of tasks. Self-reported measures of physical, emotional, and cognitive functioning include questions about inability to perform self-care activities (Katz, 1987); inability to perform activities necessary to live in the community (Lawton et al., 1969); and measures of mobility limitation or body movement limitation (Rosow and Breslau, 1966.)

Since the 1980's measurement of functional limitation has shifted from self-report measures to performance-based measures (Guralnik et al., 1989) that

are designed to test a subject's capacity to perform basic actions. Measures of muscle strength, standing, ambulation, fine motor skills, extension and flexion (upper/lower extremity), and manual dexterity have been developed.

Cognitive impairments generally manifest themselves as problems in a number of areas of cognitive functioning, including judgment or reasoning, impaired thought, memory, concentration problems, disorientation, attention disorders, impaired alertness, or the inability to carry out intellectual behaviors. The Mini-Mental State Examination (Folstein et al., 1975) and the Cognitive Capacity Screening Exam (Jacobs et al., 1977) are examples of cognitive test batteries.

The fourth and final stage in the disability conceptual model is handicap. This refers to the limitation on social roles within a specified sociocultural environment that results from impairment and functional limitation. The NHIS, the Survey of Income and Program Participation (SIPP), and the Current Population Survey (CPS) include questions about limitations in the kind or amount of work a person can do as well as the ability to work at all because of a health condition or impairment. The Functional Assessment Inventory (Crew and Athelston, 1981) contains measures of work history from which can be derived variables, such as frequency of tardiness, periods of unemployment, frequency of job changes, and periods of absence from work due to illness.

Verbrugge (1991) and others have noted that factors such as family/social support and an individual's ability to adapt to and cope with limitations are related to the extent of handicap that occurs. For the work perspective of the SSA, factors such as access to job opportunities, education and training, work experience, and social supports like the availability of transportation all mediate the disability process.

3. MEASURING DISABILITY IN THE MES

For the purposes of the MES, the term disabled persons refers to those who satisfy the SSA definition of disability, that is, the inability to engage in any substantial gainful activity by reason of any medically determinable physical or mental impairment that can be expected to result in death or has lasted or can be expected to last for a continuous period of not less than 12 months. Operationalizing the SSA definition and determination process for the MES requires taking a process developed for evaluating an individual's suitability for a program, which included data collection mechanisms, and converting it to a process that can be used in a research setting. It was important to

SSA that the survey replicate the current determination process, to the extent possible. It was also important, however, to consider the objectives of the survey that went beyond the current process and to incorporate measures into the survey that will allow SSA to examine the impact of changes in the definition of disability on prevalence and costs to the program.

Survey design issues also influenced decisions about the selection of survey measures and operationalizing the disability concept. First, the emphasis on studying nonbeneficiaries had to be considered. Additionally, since disability is a rare event in the population from a sampling perspective, a population screening approach will have to be used to identify disabled persons for the survey and a large number of persons will have to be screened. Since many of the MES study subjects will not be disability beneficiaries and may not be in regular medical treatment, accurate identification of individuals as disabled according to the SSA definition of disability will require the collection of medical data, in part by clinical examination, to establish the presence and/or severity of an impairment. The collection of medical data is quite costly. These factors lead to a two-phase approach to data collection.

The general research plan proposed for the MES is to conduct a large-scale first-phase screening (either in person or by telephone) to classify a sample of non-institutionalized, working age persons (18 to 69 years of age) into one of the following groups: all SSA beneficiaries, nonbeneficiaries who are severely disabled, and thus likely to meet the SSA definition, nonbeneficiaries who are less severely disabled, and nonbeneficiaries who are not disabled with significant disability. Then subsamples must be drawn from the screening sample to satisfy the MES objectives. The second phase involves extensive face-to-face interviews, physical performance testing, collection of medical records, and medical examinations.

The SSA disability definition includes physical disabilities and mental health disabilities, as well as assessments of the person's status in each stage of disability (impairment, functional loss, and handicap). A two-dimensional typology of type of disability by stage of disability was constructed to provide a convenient way to organize and select appropriate measures for the study. Table 1 lists the measures selected for each category in the typology. While the MES data collection protocol includes measures for all the categories in the typology, the following discussion of the rationale for the selection of measures is limited to the impairment and functional loss stages of disability. The discussion will first focus on measures that may

Table 1. Disability process measures proposed for the MES in the Screener (S) and Main Survey (M)

Disability Stage	Type of Disability		
	Physical Health	Mental Health	
		Cognitive	Psychiatric/Emotional
Condition/Impairment	<ul style="list-style-type: none"> • Condition list of SSA presumptive disabilities (S) • Condition list of chronic conditions and impairments (M) • Symptom checklist (M) • Medication usage (M) • Weight/Height (M) • Hospitalization (M) • Exposure to selected toxic substance (M) • Medical examination by physician (M) • Selected biochemical tests (M) • Selected medical tests (M) 	<ul style="list-style-type: none"> • Condition list of developmental disabilities (S) • Medication usage (M) • Cognitive status assessment by physician (M) • Medical examination by physician - Neurological (M) 	<ul style="list-style-type: none"> • Condition list of mental health problems (S) • CIDI (S) • Medication usage (M) • SCID (M) • Mental status assessment by physician (M)
Functional Loss	<ul style="list-style-type: none"> • ADL (S/M) • IADL (S/M) • Mobility and body movement limitations (S/M) • Lifting (S/M) • Physical performance tests (M) 	<ul style="list-style-type: none"> • ADL (S/M) • IADL (S/M) • Mini-Mental State Exam or TICS 	<ul style="list-style-type: none"> • ADL (S/M) • IADL (S/M) • SOFA (M)
Handicap/Disability	<ul style="list-style-type: none"> • Limitation in major/usual activities (S) • Work limitation questions (M) • Work history/ employment experience (M) 	<ul style="list-style-type: none"> • Limitation in major/usual activities (S) • Work limitation questions (M) • Work history/ employment experience (M) 	<ul style="list-style-type: none"> • Limitation in major/usual activities (S) • Work limitation questions (M) • Work history/ employment experience (M)
Exogenous Environmental Variables	<ul style="list-style-type: none"> • Family composition (M) • Health insurance coverage (M) • Education/ training (M) • Accommodations/interventions (M) • Barriers to work (M) • Income/assets (M) • Social support/social network (M) • Health habits (M) 		

be used in the first-phase screening and then the second-phase measures, which are more extensive and can be used to apply the SSA disability determination criteria.

Screener - Since the screening sample will be large, the screener instrument developed to identify disabled persons needs to be administered easily by survey interviewers and be based on respondents self-reports. Measures that have been used in population surveys, such as the CPS, the SIPP, and past NHIS, generally reflect some of the dimensions of the SSA disability definition. While the questionnaires for these surveys have been useful in identifying important measures of disability, they do not adequately represent the range of potential limitations resulting from disability.

More recently, NCHS has fielded the NHIS 1994-95 Disability Supplement, which includes a wide range of measures of the many dimensions of disability. The NHIS Disability Supplement, which forms the core of the MES screener questionnaire, defines various disability groups and includes measures of disability for each of the groups. The NHIS is con-

cerned with all levels (ranges of severity) of disability, and for the purposes of its Disability Followup Survey, includes all persons reporting any disability.

In reviewing the NHIS approach to defining disability with the MES consultants and expert panel, two shortcomings for its use on the MES were identified. First, while the NHIS classification system is useful for identifying policy relevant groups, it includes persons who are not (and are not likely to be) severely disabled, and thus must be modified for use in the MES. Second, the measures of cognitive impairment and mental illness were considered to be too gross, thus not permitting persons with these impairments to be placed in groups by severity of functional limitation as required by the MES objectives.

Identification of persons who were disabled in the physical domain in the screener should be based primarily on reports of functional loss rather than conditions or impairments because of the problems of self-report inaccuracies of medical conditions and the difficulties associated with assigning a measure of severity to conditions and impairments that would be necessary to classify persons for sampling purposes. The exception to this approach is the inclusion of a list of

conditions/impairments considered by SSA to be presumptive disabilities, for example, AIDS and blindness. Thus the screener includes ADL and IADL measures and mobility and body limitation measures as well as a condition list of the presumptive disabilities.

To address the problems in the cognitive and mental health arena, the Mini Mental State Exam (or the Telephone Interview for Cognitive Status, Brandt, 1988) and the Composite International Diagnostic Interview screening test for mental illness and substance abuse (Kessler et al., 1995), were included, as was a checklist of conditions/impairments related to severe developmental disabilities, for example, Down syndrome and autism.

Main Survey Data Collection - Condition/impairment data are collected in the interview and during the medical exam, tests, and from medical records. Functional loss data are collected primarily in the interview and physical performance testing; however, the physician is also asked to provide an assessment of the subject's functional capacity since this is the approach used by SSA in the determination process. Data collection and measurement issues for condition, impairment, and functional status measures are discussed in the remainder of this section.

Questionnaire information on the presence of diagnosed conditions and impairments will be collected using condition lists based on the SSA medical listings, which use a body system approach to categorizing impairment. Symptom checklists are also incorporated in an attempt to collect information on undiagnosed problems and to serve as a medical history that will assist the physician in conducting the medical exam. Condition and symptom questions were adapted from the NHIS, NHANES and the SSA data forms as applicable. Information on date of onset, duration, medications prescribed, and hospitalizations for the problem are captured in the questionnaire. Where feasible, a measure of severity is included.

The development of the medical examination content and process was driven largely by impairments considered in the SSA medical listings. Generally, a medical examination is structured by patient complaint and history. While this approach is included in the MES exam, a general exam of all body systems will be conducted for each study participant not only to ensure consistency but also to address the issues of undiagnosed disease and co-morbidity.

In reviewing the lists of medical tests required as evidence by the SSA medical listings, only a subset could be included in the survey, because some tests either required extensive preparation or carried more-

than minimal risk to the subject. Consequently, for some subjects with selected impairments it might be impossible to make a disability determination because the required medical tests cannot be performed. Two options for addressing this problem were considered: (1) sending subjects to a specialized health care facility for additional testing and (2) obtaining medical records from health care providers. Because the first option was likely to result in extremely high nonresponse and presented logistical problems, the collection of medical records, where necessary, was included in the data collection protocol.

Most measures of functional loss have been developed primarily in relation to research on aging and vocational rehabilitation, with aging studies focusing on functional loss as it affects independence and community living and vocational rehabilitation focusing on what someone can do so as to place them in a job. Little research has been conducted that explores the relationship between these functional measures and the inability to work. While the relationship is clearer with the vocational rehabilitation measures, the measures are often specific to one limited job type, require workstations and the use of expensive equipment to conduct the assessment, and usually take several hours (or days) to complete. Measures developed in the aging arena are more globally applicable, but virtually no research exploring their relationship to work inability exists.

Of concern in the measurement of functional loss is the phenomenon of compensation and the concepts of maximal effort and sustainability. These issues pertain to both self-reported and performance measures of functional loss. Compensation occurs when a subject reports that s/he can perform much better than would be the case under normal circumstances out of an interest in pleasing the tester. Adding phrases to questions, such as "without special equipment or assistance," helps focus the respondent on what s/he can realistically do. If an accurate measure of functional loss is to be obtained, compensation should be controlled in the data collection process.

Maximal effort and sustainability are dimensions of the measure of functional loss. Maximal effort refers to doing the best one can in performing a task and sustainability to determining how long a subject can perform a particular task. These concepts pertain most clearly to physical performance measures of functional loss, but sustainability is often also measured when self-reports of functioning are obtained using a questionnaire. Questions can include items about varying lengths of time performing a task. During performance tests, several trials of a test can be performed.

Physical performance measures are proposed as are ADL and IADL measures in order to capture severe levels of disability and ability to function in the community. The performance tests include tasks involving using the telephone, balancing a checkbook, and paying a bill.

The measures described will also be used to measure functional loss in persons who are cognitively impaired and those who are mentally ill. In addition, a number of dimensions of cognitive functioning (memory, attention span, construct ability, orientation in time and place, among others) will be assessed by the physician during the exam for persons cognitively impaired, and a score for the Social and Occupational Functioning Assessment Scale - SOFAS (Goldman et al., 1994) - will be obtained during the psychiatric exam for persons who are mentally ill.

4. SAMPLE DESIGN ISSUES

To satisfy the analytical goals of the MES, adequate sample sizes are required for the following four subgroups of the working age population (i.e., persons aged 18 to 69 years old): severely disabled nonbeneficiaries; SSA beneficiaries; less severely disabled nonbeneficiaries; and not disabled nonbeneficiaries. Of these subgroups, the one that dominates the determination of the screening sample size is that of the severely disabled nonbeneficiaries, for which a sample size of 3,000 is proposed. From the limited data currently available, the percentage of the working age population in this subgroup is tentatively estimated to be around 3.5 percent. Using this estimate, and allowing for nonresponse at both phases of data collection, around 130,000 to 170,000 working aged persons (depending on the mode of screening data collection) need to be screened to yield a sample size of 3,000 disabled nonbeneficiaries.

The high cost of screening such a large sample may be reduced if the screening can be attached to an existing large-scale survey. The NHIS is a natural candidate in this regard because in 1994 and 1995 it included a disability supplement to collect detailed information about the presence and degree of disabilities for all members of the sampled households. There are, however, a number of issues that need to be considered in such an application.

One issue concerns the contents of the NHIS Disability Supplement. The supplement probably collects the data needed for the MES screening with respect to physical disabilities, but it is less well suited with respect to mental health disorders and cognitive disabilities, where it contains only items relating to severe impairments.

Another issue concerns the time interval between the conduct of the NHIS and the MES. During that interval persons who were not disabled will become so, some persons identified as disabled will die or will leave the noninstitutionalized population (e.g., enter a nursing home), and some will move, thus necessitating a tracing operation. The long interval between the 1994 NHIS and the MES data collection probably rules out the 1994 NHIS as part of the screening for the MES.

A third issue concerns the wide geographic spread of the 1995 NHIS sample across 358 primary sampling units (PSUs). For operational, quality control and cost reasons, it is desirable to conduct the MES second phase data collection in a smaller number of PSUs, say around 100. While the 1995 NHIS sample has a 110 PSU subdesign, that subdesign would yield only about one-quarter of the required screening sample for the MES. The remaining three-quarters of the sample would then need to be obtained by additional screening interviews in the same PSUs.

A fourth issue concerns a NCHS follow-back interview survey of persons identified by the NHIS Disability Supplement as disabled. This survey raises concerns about respondent burden and about timing. Disabled NHIS participants will already have participated in a core and disability supplement interview and a follow-back interview when they are asked to participate in the extensive MES data collection. Moreover, the need to complete the follow-back interviews before the conduct of the MES further extends the time interval between the NHIS interview and the MES data collection, with the attendant problems of timing previously noted.

Finally, because of the time delay, the additional contacts required, and the heavy respondent burden, linking the MES to the NHIS would probably produce a lower response rate than would an independent area probability sample design. Thus, while there would be some cost savings associated with using the NHIS sample, there are a number of methodological problems that need to be addressed.

An alternative approach is to conduct the MES screening as an independent operation with a sample design tailored to the needs of the second phase of the MES data collection. This independent screening could be conducted either by face-to-face or telephone interviewing. In either case, the screening sample would need to be clustered into around 100 PSUs to facilitate the second phase data collection.

If face-to-face interviewing is chosen, standard area probability sampling procedures can be employed. The sample can be restricted to around 100

PSUs to facilitate the second phase data collection. With the face-to-face interviewing, all eligible persons in selected households can be screened. With this design, after allowance for nonresponse, around 88,000 dwelling units and 132,000 persons of working age would need to be screened to produce the required sample size of 3,000 disabled nonbeneficiaries.

Screening by telephone is a much less costly alternative to face-to-face screening. There are, however, a number of disadvantages to telephone screening for this purpose that must be balanced against the cost savings. In the proposed design for face-to-face screening, all adults in the household are included in the screening sample. The screening questionnaire, which takes about 35 minutes to complete, is administered to each of them. If this procedure were administered by telephone interviewing, the heavy household burden would result in a high nonresponse rate. One alternative is to sample only one adult per household. This would reduce the household response burden, but it would require a major increase (about 70%) in the number of sampled households. Another alternative is to structure the screener, where possible, as a household interview with one knowledgeable respondent for all family members, and additional short interviews with individual household members to collect information on mental health and cognitive function that can be obtained accurately only from the sampled individual.

Even with the knowledgeable respondent approach, nonresponse will be higher for telephone than for face-to-face screening, possibly by as much as 10 to 15 percent. There are two main consequences of the lower telephone screening response rate. The first is that the screening sample will have to be increased to compensate for the losses. A sample size of about 170,000 persons of working age is needed with telephone screening to yield the required 3,000 disabled nonbeneficiaries completing both phases of the MES data collection. More importantly, the increased nonresponse raises the risk of bias in the survey estimates. Nonresponse at the screening interview is particularly damaging because little is known about the characteristics of the nonrespondents.

Telephone noncoverage, which is typically not a major problem in surveys of the working age population (NHIS estimates that it is about 5.3%), is a more serious concern for the disabled population. In fact the NHIS indicates that the proportion of persons who are unable to work because of a health problem and do not have a telephone is about 9.8 percent.

Not only will telephone screening lead to a lower screening response rate, it will also probably lead to a lower response rate at the second phase. The face-to-

face model can combine screening and data collection, including the main interview and functional performance testing, into a single data collection contact. The telephone approach separates these into two distinct contacts and thus provides respondents with an additional opportunity to refuse to cooperate in the study.

REFERENCES:

- Brandt, J., Spencer, M., Folstein, M. (1988). The telephone interview for cognitive status. *Neuropsychiatry, Neuropsychology, Behavioral Neurology*, 1 (2), 111-17.
- Crew, N.M., Athelston, G.T. (1981). Functional assessment in vocational rehabilitation: A systematic approach to diagnosis and goal setting. *Arch Phys Med Rehabil*: 62: 299-305.
- Folstein, M. J., Folstein, S. E., McHugh, P. R. (1957). Minimal state. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatric Res*: 12 (3), 189-98.
- Goldman, H. H., et al. (1994). Social and occupational functioning assessment Scale (SOFAS). In *Diagnostic and Statistical Manual of Mental Disorders*, 4th Ed. Appendix B., American Psychiatric Association, Washington, D. C.
- Guralnik, J. M. et al. (1989). Physical performance measures in aging research. *J. Gerontol Med Sci.*, 44: M141-6.
- Jacobs, J. W., Bernhard, M. R., Delgado, A., Strain, J. (1977). Screening for organic mental syndromes in the mentally ill. *Annals of Internal Medicine*, 86, 40 - 46.
- Katz, S. (1987). The Portugal Conference: Measuring quality of life and functional status in clinical and epidemiological research. *Journal of Chronic Disorders*, 40 (6), 459-650.
- Kessler, R. C., Mroczek, D., Vitken, H. (1995). Short form screening scales for DSM - III - R. *Psychiatric Disorders* (in preparation). Ann Arbor, MI: Institute for Social Research, University of Michigan.
- Lawton, M. P., Brody, E. M. (1969). Assessment of older people: Self maintaining and instrumental activities of daily living. *Gerontologist*: 9, 179-86.
- Nagi, S. (1991). Disability concepts revisited: Implications for prevention. In A. Pope & A. Tarlov (Eds.). *Disability in America: Toward a national agenda for prevention*. Washington, D. C., National Academy Press.
- National Center for Health Statistics (1994). Plan and operation of the Third National Health and Nutrition Examination Survey, 1988 - 1994. *Vital Health Statistics* 1 (32).
- Rosow, I., Breslau, N. (1966). A Guttman Health Scale for the aged. *Journal of Gerontology*, 21 (4), 556-59.
- Spitzer, R. L., et al. (1990). *Structured clinical interview for DSM - III - R*. Washington, D. C. American Psychiatric Press.
- Verbrugge, L. (1991). Physical and social disability in adults, Conference Proceedings - Primary care research: Theory and methods, . *Agency for Health Care Policy and Research*.
- Wood, P. (1980). *International classification of impairments, disabilities and handicaps: A manual of classification relating to the consequences of disease*. Geneva: WHO.
- Zola, I. K. (1993). Disability statistics, what we count and what it tells us: A personal and political analysis. *Journal of Disability Policy Studies*. 4(2), 10-39.