



U.S. Department of Health and Human Services
Assistant Secretary for Planning and Evaluation
Office of Disability, Aging and Long-Term Care Policy



OPPORTUNITIES TO IMPROVE SURVEY MEASURES OF LATE-LIFE DISABILITY:

PART II--WORKSHOP SUMMARY

September 2006

Office of the Assistant Secretary for Planning and Evaluation

The Office of the Assistant Secretary for Planning and Evaluation (ASPE) is the principal advisor to the Secretary of the Department of Health and Human Services (HHS) on policy development issues, and is responsible for major activities in the areas of legislative and budget development, strategic planning, policy research and evaluation, and economic analysis.

ASPE develops or reviews issues from the viewpoint of the Secretary, providing a perspective that is broader in scope than the specific focus of the various operating agencies. ASPE also works closely with the HHS operating divisions. It assists these agencies in developing policies, and planning policy research, evaluation and data collection within broad HHS and administration initiatives. ASPE often serves a coordinating role for crosscutting policy and administrative activities.

ASPE plans and conducts evaluations and research--both in-house and through support of projects by external researchers--of current and proposed programs and topics of particular interest to the Secretary, the Administration and the Congress.

Office of Disability, Aging and Long-Term Care Policy

The Office of Disability, Aging and Long-Term Care Policy (DALTCP), within ASPE, is responsible for the development, coordination, analysis, research and evaluation of HHS policies and programs which support the independence, health and long-term care of persons with disabilities--children, working aging adults, and older persons. DALTCP is also responsible for policy coordination and research to promote the economic and social well-being of the elderly.

In particular, DALTCP addresses policies concerning: nursing home and community-based services, informal caregiving, the integration of acute and long-term care, Medicare post-acute services and home care, managed care for people with disabilities, long-term rehabilitation services, children's disability, and linkages between employment and health policies. These activities are carried out through policy planning, policy and program analysis, regulatory reviews, formulation of legislative proposals, policy research, evaluation and data planning.

This report was prepared under contract #HHS-100-03-0011 between HHS's ASPE/DALTCP and the Urban Institute. For additional information about this subject, you can visit the DALTCP home page at http://aspe.hhs.gov/_/office_specific/daltcp.cfm or contact the ASPE Project Officers, William Marton and Hakan Aykan, at HHS/ASPE/DALTCP, Room 424E, H.H. Humphrey Building, 200 Independence Avenue, S.W., Washington, D.C. 20201. Their e-mail addresses are: William.Marton@hhs.gov and Hakan.Aykan@hhs.gov.

OPPORTUNITIES TO IMPROVE SURVEY MEASURES OF LATE-LIFE DISABILITY: Part II--Workshop Summary

Vicki A. Freedman
University of Medicine and Dentistry of New Jersey

Timothy Waidmann
Brenda Spillman
Urban Institute

September 27, 2006

Prepared for
Office of Disability, Aging and Long-Term Care Policy
Office of the Assistant Secretary for Planning and Evaluation
U.S. Department of Health and Human Services
Contract #HHS-100-03-0011

Prepared for the Workshop on Improving Survey Measures of Late-Life Disability, held at the Urban Institute, Washington, DC, May 17, 2005, funded by the HHS Office of the Assistant Secretary for Planning and Evaluation. We thank workshop participants who provided summaries of their presentations for use in this report. The views in this overview reflect those of the authors alone and do not represent those of the authors' affiliations or funding agencies. The authors are solely responsible for any errors or omissions.

TABLE OF CONTENTS

WORKSHOP RATIONALE AND GOALS	1
OPPORTUNITIES FOR NEW MEASURES OF LATE-LIFE DISABILITY.....	2
INNOVATIONS IN MEASUREMENT	6
Current International and National Activities Related to Disability Measurement.....	6
Self-Reported Work Disability and the Use of Vignettes	8
Measures of Physical Functioning in Late-Life.....	10
Measuring Participation/Engagement	12
Linking Measures of Assistive Technology and Disability	14
Measures of Environmentally Determined Mobility Disability	16
CONSIDERATIONS FROM THE FIELD	19

WORKSHOP RATIONALE AND GOALS

Advances in conceptual thinking since the development of activities of daily living (ADLs) and instrumental activities of daily living (IADLs) provide new opportunities to expand the range of questions that might be answered with survey data. Standardization of disability measures, for example, may promote comparisons across surveys, groups, and countries. Distinguishing physiological, environmental, and social components of disability may help policy makers better target resources at interventions likely to have high impact on population disability rates. Moreover, such distinction can help researchers track and understand shifts in population-level disability trends. Widespread use of these new measures could also improve our understanding at the individual level of the physiology of functional loss and recovery, the accommodation process, and the effectiveness of interventions to enhance independence and participation. Many of the advances discussed here are quite recent and have not been routinely incorporated into most national surveys that address late-life disability.

The workshop on Improving Survey Measures of Late-Life Disability, held in May 2005 at the Urban Institute in Washington, DC, reviewed advances in our understanding and measurement of late-life disability with the aim of answering two fundamental questions:

- Are our current measures of late-life disability meeting the needs of researchers and policy makers?
- How can we improve measures of late-life disability within current surveys?

The meeting involved three sessions. The first session included a panel discussion to flesh out the opportunities that new measures of disability might provide to policy makers and researchers. The second session highlighted research by six speakers, each focusing on an innovation in disability measurement. The final session involved a panel discussion about the practical considerations in implementing new measurement techniques with the aim of identifying the most promising strategies. The complete workshop agenda and biographies of participants are included in the Appendix.

OPPORTUNITIES FOR NEW MEASURES OF LATE-LIFE DISABILITY

The first panel brought together researchers using disability data for both basic and policy studies and a representative of the Social Security Administration (SSA) for a discussion organized around four questions:

- What are important new research and policy questions relating to late-life disability? Do they imply different data needs for policy makers than for researchers?
- What are the shortcomings of current measures of ADL, IADL, and physical or mental limitations for addressing these questions?
- For what questions and purposes are current measures the “right” concept, and what modifications would improve how well they measure the concept?
- For what questions and purposes do we need new measures?

The panelist’s comments and subsequent discussion suggested the potential for common ground on how new measures and improvements to existing measures might meet the needs of both. From the researcher perspective, more emphasis was placed on better understanding the underlying processes that drive disability and disability trends and the implications for identifying persons at risk and predicting future trends. From the program perspective, more emphasis was placed on reliable, objective measures that predict program eligibility and participation.

Dr. Susan Allen and Dr. Robert Schoeni focused on the need to understand the causes of observed declines in old age disability in order to understand potential for effective interventions.

Dr. Allen noted that the causal pathways for disability are complex and different approaches may be needed to determine the role of socio-cultural and individual factors. We do not yet know the extent to which improvements in functioning may reflect healthier lifestyles among the older population, widespread changes at the societal level in how we perform the daily activities assessed in measures of IADLs and ADLs, or changing attitudes toward exercise and general physical activity or expectations about aging by both older persons and their informal network of informal caregivers. She suggested that measuring such individual factors as social involvement and personality characteristics (e.g., introvert/extrovert); potential early markers of decline such as level of fatigue and changes in activities or activity level; and characteristics of the physical environment may help identify those at risk for decline and the importance of these factors in preventing or deferring decline.

Dr. Schoeni identified a better understanding of differentials in disability decline across sociodemographic groups, across time and across geography as a key issue for both research and policy. He suggested that the underpinnings for better understanding might include more focus on identifying the factors and stage of life associated with higher disability rates. Examples are measures of very early life characteristics and events, for example birth weight, that may predict later life disability risks. He also noted the need for capturing the impacts of nonhealth factors--social, economic, and community policies--on health and disability. Focusing on ADLs and IADLs without sufficient information about accommodations used (including accommodations not conventionally thought of as assistive technology), and relevant aspects of the physical environment, household, and community may mask the effectiveness of policies that are "paying off." Such factors may affect both whether an individual has disability and whether the individual reports disability.

Dr. Schoeni also raised two issues that were a common theme among the panelists and in subsequent discussion: Measuring time costs of disability and the role of mental health and individual attitudes would provide insight into disability and its impact not supported by current measures. He gave the example of "daily reconstruction" using time diaries and eliciting respondent feelings about the time required to perform various activities being considered for the Panel Study of Income Dynamics (PSID).

Dr. Howard Iams noted that from SSA's perspective, a key issue is understanding the relationship between disability and "wellbeing," specifically factors that improve understanding of the timing of retirement and identify ability to work past traditional retirement ages as life expectancy continues to increase. From that perspective, ADLs and IADLs are less useful than actual measures of physical functioning or ability, such as "Nagi" items (e.g., walking short distances, climbing stairs, and lifting and carrying something weighing ten pounds) that determine degree of difficulty with components of more complex physical activities. Iams noted that better measurement of physical abilities, including more attention to measures that allow scaling of difficulty level, would provide an empirical underpinning for considering the potential for linking full benefits to life expectancy, as increasing longevity pushes Social Security costs upward.

A particular concern from the program perspective is that measures be operationally feasible, preferably in administrative data, and consistent across surveys and time. Specific measurement issues raised were the reliability of self-reports versus objective measures and the conditional format of disability measures on many surveys so that functioning is not assessed for the full survey sample. Iams noted the SSA's disappointing experience with efforts to use existing self-report survey measures, including ADL difficulty or a condition that limits work, to predict receipt of disability benefits. Like Schoeni, he advocated more and better measures of mental health, which is the basis of eligibility for 1/3 of disabled beneficiaries, and of time required to perform activities.

Dr. Lisa Iezzoni said that better information is needed on sensory impairments, which may result in disability and in reduced participation in all areas of life. Simple but effective treatments or accommodations that may delay or prevent disability resulting from such impairments (e.g., hearing aids) are not covered by Medicare.

She noted that shortcoming of existing ADL/IADL measures. For example, just as IADL questions may fail generic changes such as the advent of microwave ovens that make meal preparation less physically demanding, ADL questions fail to capture accommodative practices, such as wearing only elastic waist pants, that may allow independent dressing but are not commonly thought of as explicitly disability-related. More attention might be focused on how rather than whether activities are accomplished. This focus could also apply to improvements in measures of assistive device use, since a focus on whether particular devices are used for particular applications fails to capture the full array of devices and the multiple contexts in which they are used. Iezzoni also noted that the cross-sectional focus of most data collections (e.g., now, in the last week) fails to capture the waxing and waning of abilities that characterize many disability-related conditions and diseases. Finally, Iezzoni suggested improving the information collected on household structure, including who lives in the household and is available to assist with activities as a matter of course, which may affect disability reporting.

A number of additional observations and elaborations were made by meeting participants and panelists:

- A better understanding is needed of mental health and impairments beyond current memory and cognitive functioning measures and how they interact with ADL measures.
- Better ways are needed to control for endogeneity that limits the ability to demonstrate whether such medical inputs as restorative procedures and advances in medications are effective in reducing or delaying functional decline.
- Dependence on ADLs, in particular, as a disability measure may be too narrow to capture the broad range of disability across the age spectrum; better and more broadly applicable measures might focus on physical functioning necessary to perform activities rather than on the activities themselves.
- Distributional issues are crucial to consider in disability measurement: ethnic, racial, and socio-economic factors may affect biological aging and access to services and technology; geographic disparities in public benefits, infrastructure, and accommodative built environment also affect disability rates.

- ADLs, while perhaps limited as a measure of population disability, have value in terms of measuring long-term care use and needs and are the foundation for eligibility for both public and private insurance benefits for older persons.
- The difficulty in separating need from individual preferences pose a methodological problem for measuring time costs of disability.

INNOVATIONS IN MEASUREMENT

Presentations addressed efforts to standardize measures and underlying components of disability (physical functioning; the environment and assistive technology; and time use and participation).

Current International and National Activities Related to Disability Measurement

Dr. Jennifer Madans provided an overview of national and international efforts to create brief measures of disability. She began by noting that measures of ADLs and IADLs, while widely used, do not adequately reflect the full spectrum of the concept of disability. She then went on to provide a summary of the recent activities of the Washington Group, the inter-departmental workgroup tasked to review the disability questions on the American Community Survey (ACS), and a task force on health measurement formed by UNECE, WHO, and Eurostat.

Under the aegis of the United Nations Statistical Commission, City Groups are formed to address important problems in statistical methods. The Groups are composed of international experts, primarily from the national statistical authorities. The Washington Group, named after the location of its first meeting, was convened to promote the co-ordination of international cooperation in the area of health statistics by focusing on disability measures suitable for censuses and national surveys, which will provide basic necessary information on disability throughout the world.

The group's current objectives are to develop and test 2-3 sets of general questions for international use in censuses; to understand the limited choices associated with developing census questions; and to understand the product that results from census questions. A fourth objective, is to recommend an extended set of items (including functioning, participation, and the environment) that can be added to collect information on disability on population surveys of all kinds. The group will also address methodological issues associated with disability measurement, including measures for special populations (e.g., children, those living in shared residential care settings).

The Washington Group uses the International Classification of Functioning, Disability and Health (ICF) as a conceptual framework. The ICF depicts disability as the interaction between individual characteristics and the environment. The components of the model include body functions and structures, activities, participation, health conditions, environmental factors, and personal factors. The ICF provides the conceptual building blocks for users who wish to study different aspects of the disablement process, but it does not provide a way to measure the concepts. Thus the Washington Group's task is to operationalize the concepts for use in censuses around the world.

The group began by identifying three possible major purposes for measuring disability in a census: (1) to provide information around the provision of services; (2) to monitor the level of functioning in a population; and (3) to monitor equalization of opportunity. The characteristics of the measures to be developed will depend on the purpose chosen. For Censuses, equalization of opportunity was selected as the primary purpose of data collection. To meet this purpose, information is collected to identify those who would be at greater risk than the general population for limitations in activity or participation without accommodation. Questions would be designed to meet the criteria of cross-cultural comparability, suitability for self-report, parsimony, and validity across modes of data collection.

The Washington Group recommended that the Census questions should, at a minimum, cover three essential domains of functioning: walking, seeing, and cognitive functioning. Additional domains identified were hearing, upper body functioning and communication. In recommending response categories, the group recognized tradeoffs between the ease and comparability of yes/no outcomes versus the need for capturing degrees of difficulty with finer categories.

Washington Group Recommended Census Items

1. Do you have difficulty seeing even if wearing glasses?
2. Do you have difficulty hearing even if using a hearing aid?
3. Do you have difficulty walking or climbing stairs?
4. Do you have difficulty remembering or concentrating?
5. Do you have difficulty with (self-care such as) washing all over or dressing?
6. Because of a physical, mental, or emotional health condition, do you have difficulty communicating (for example, understanding others or others understanding you)?
 - a) No -- no difficulty
 - b) Yes -- some difficulty
 - c) Yes -- a lot of difficulty
 - d) Cannot do at all

These measures are now being tested in a variety of countries. The Washington Group is facilitating the process by training country representatives in cognitive testing, translation, and other relevant areas, and by developing an analysis plan for pre-test results.

In the United States, the disability questions that were on the 2000 Census and that are now on the ACS are going through a process of review with results similar to that of the Washington Group. The workgroup also chose “equalization of opportunity” as the main purpose of the measures and decided to focus on four key domains: hearing, seeing, walking and cognition. They also included one item to capture severe disability affecting the need for long-term care. Recommended items (with yes/no response categories) appear below. The first two items are asked for all ages, items 3-5 are asked for respondents ages 5 and older, and item 6 for ages 17 and older. The items have undergone extensive cognitive testing and will be included in the ACS content test.

Recommended American Community Survey Items

1. Is this person deaf or do they have serious difficulty hearing?
2. Is this person blind or do they have serious difficulty seeing, even when wearing glasses?
3. Because of a physical, mental, or emotional condition, does this person have serious difficulty concentrating, remembering, or making decisions?
4. Does this person have serious difficulty walking or climbing stairs?
5. Does this person have difficulty dressing or bathing?
6. Because of a physical, mental, or emotional condition, does this person have difficulty doing errands alone such as visiting a doctor's office or shopping?

Dr. Madans also briefly reviewed efforts by UNECE, WHO and Eurostat to develop a common instrument to measure health states. Their aim requires a multidimensional measurement approach, one which focuses on the capacity of the individual, and maximizes cross-population comparability. The collaborative effort is in the process of identifying a core and extended set of domains.

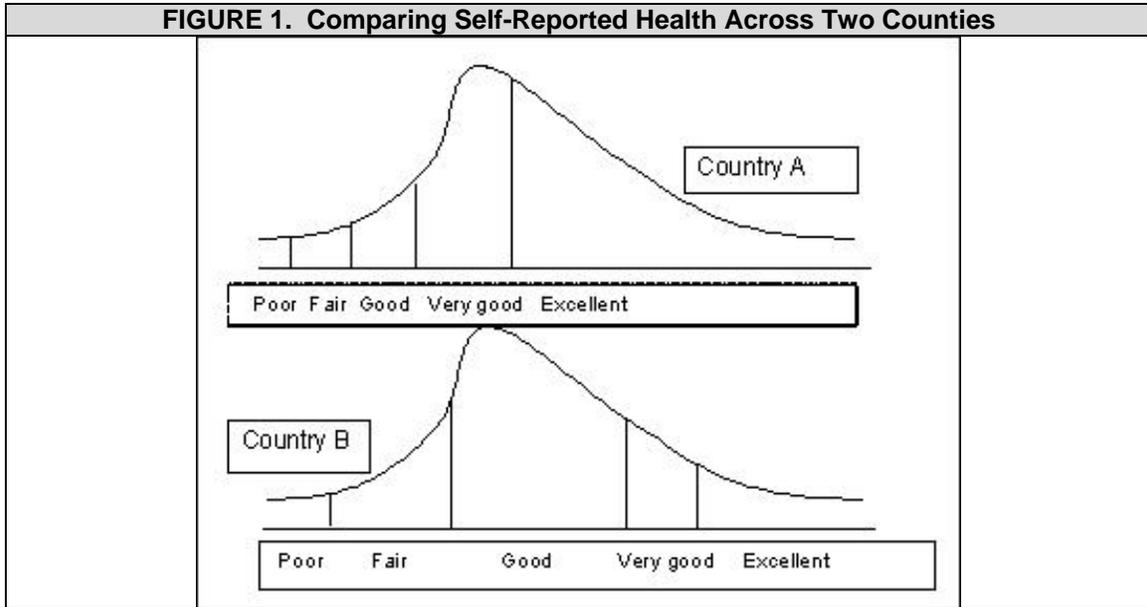
Dr. Madans concluded with a set of recommended next steps for measurement of late-life disability. She suggested that major surveys and censuses would benefit from less reliance on ADLs as the primary indicator of disability and more attention to measures of basic functioning, or precursors to ADL and IADL limitations. She recommends differentiating between capacity to perform and actual performance (with and without aids/assistance). In addition, she suggests reformulating the IADLs and expanding the range of participation domains measured and studies.

Self-Reported Work Disability and the Use of Vignettes

Dr. Smith provided an overview of the use of vignettes to standardize measurement of work disability. In motivating the usefulness of vignettes, he explained that disability insurance enrollment and work disability is much higher in the Netherlands than in the United States, but reports of chronic diseases are higher in the latter. In England, there are lower rates of diseases (based both on self-reports and administrative data), but Americans rate health better on a five-point general health status measure. The distributions of self-reported health may vary because the thresholds corresponding to poor, fair, good, very good, and excellent might differ across the two countries (see Figure 1). Vignettes have been used to understand differences in thresholds across populations and to correct for these differences.

Vignettes are a brief description of a hypothetical person. Each respondent is asked to self-report whether they have a health problem that limits the amount or type of work they can do, on a five-point scale from "not at all" to "extremely limited". They are then asked to rate individuals appearing in brief vignettes on the same scale. For example, a vignette for an individual with emotional problems follows:

[Tamara] has mood swings on the job. When she gets depressed, everything she does at work is an effort for her and she no longer enjoys her usual activities at work. These mood swings are unpredictable and occur two or three times during a month.



The vignettes vary in terms of severity and condition. Administration of the full set of vignettes takes approximately eight minutes. The information from the vignettes is then used to standardize rankings across groups so that comparisons are no longer biased.

Dr. Smith has been involved in efforts to add work disability vignettes to the CentERpanel in the Netherlands and the Health and Retirement Survey (HRS), PSID and the RAND Internet Panel in the United States. The internet panels provide the advantage of experiments in question wording and quick turnaround time.

Key findings from analyses by Dr. Smith and colleagues suggest that United States respondents have a stricter standard in identifying work disability than Dutch respondents. For example, 11.1% of Americans classified Tamara (above) as being extremely or severely limited whereas 17.6% of Dutch did so. Threshold differences were also found within countries by gender, education and health. For example, women and less educated respondents in the United States had stricter standards than men, those with emotional problems and those in pain were more likely to classify an illustrative case as having a limitation.

Evidence from experimental modules has provided some evidence on how best to administer the vignettes. For example, female respondents seemed to have strict standards for classifying emotional problems as causing work limitations; however, female respondents were given only female names in the vignettes whereas male

respondents were given male names. Further experiments showed that it is the gender of the vignette that matters and not the gender of the respondent.

In conclusion, vignettes can help to make reported work disability (or other measures of disability) comparable across populations. Large differences in reported work disability cannot be explained by differences in health between the two countries. Future work will extend these tools to additional countries, cross-walk them with objective measures like grip strength, and extend them to other measures of disability such as ADLs and IADLs.

Measures of Physical Functioning in Late-Life

Dr. Gill provided an overview of measurement of physical functioning in late-life. He explained that disability assessments can differ in several ways. First, different tasks may be included, including basic ADLs, IADLs, mobility, etc. Second, participants can be asked whether they have difficulty with a task or whether they need or get help for a task. Third, the help may be from another person or from special equipment and may or may not include supervision. Fourth, a preamble may be included in an attempt to narrow the scope of the questions (e.g., “Because of a health or physical problem, do you...”). Finally, the frame of reference may differ (i.e., participants can be asked whether they have disability at the present time, during the past month, usually, etc.). Estimates of disability prevalence and incidence can vary considerably based on these differences.

He recommended that when deciding between different disability assessments, investigators should consider whether the goal of the assessment is to detect change over time or to discriminate among individuals at a single point in time. Another important consideration is whether the specific questions can be administered and answered reliably, especially to proxy respondents. Because of the importance of the actual wording, the specific disability questions should be provided in published reports. Whenever possible, decisions to choose one disability assessment over another should be empirically based.

Dr. Gill went on to describe several empirical studies that may help guide decisions about disability measures. In one study, investigators at Yale investigated the distinctions between reports of difficulty and dependence in daily activities. The investigators used data from an NIA-sponsored, population-based cohort of 1,065 community living persons, aged 72 years and older. The investigators found that older persons who were independent but reported difficulty had functional profiles, physical performance scores, and rates of health care utilization and death that were intermediate to those of persons who were independent without difficulty and persons who were dependent. These findings suggest that questions about difficulty and dependence provide complementary information. Clinicians and investigators can depict the continuum of disability more fully by including questions about both difficulty and dependence in their clinical practice and epidemiologic studies, respectively.

The Yale Precipitating Events Project Study, an ongoing study of 754 initially nondisabled, community living persons, aged ≥ 70 years, with a median follow-up (to date) of 72 months, has also provided important insights into disability measurement issues. Participants have completed comprehensive home-based assessments at baseline and, subsequently, at 18-month intervals and have been followed monthly via telephone interviews to reassess their functional status and identify admissions to the nursing home and deaths. Dr. Gill and colleagues have evaluated whether assessing disability “at the present time” leads to underestimates of short-term disability. Among a subgroup of 186 participants who had no disability “at the present time” in bathing, walking, dressing and transferring, they found that only two (1.1%) reported disability “at any time during the last month”. Investigators have also recently added items for proxy respondents (from the National Mortality Followback Survey). This information, together with the prospective reports from respondents, can be used to determine whether proxy informants can accurately report the occurrence of disability among decedents in the last year of life.

The second half of Dr. Gill’s presentation focused on physical performance measures. A physical performance measure is an assessment in which an individual is asked to perform a specific task and is evaluated in an objective, standardized manner using predetermined criteria, which may include counting of repetitions or timing of the activity as appropriate. Physical performance measures usually assess functional limitations, rather than disability per se.

The Short Physical Performance Battery (SPPB) has become the most commonly employed physical performance test in epidemiologic studies and clinical trials. The SPPB has several attractive features, including: well validated, sensitive to clinically meaningful changes, relatively portable (i.e., can be performed in home or office), and a CD-ROM instructional manual is available. Components of the SPPB and the composite battery have high predictive validity for mortality, nursing home admission, and incident disability. Most of the predictive accuracy of the SPPB is attributable to the gait speed component, which according to numerous studies is the single best predictor of disability and functional decline. Data from the Iowa EPESE demonstrate that the change scores over four years for the SPPB are normally distributed. Changes in SPPB appear to be responsive to change. For example, increasing levels of depressive symptoms were found to be associated with greater physical decline and a randomized trial of a home-based exercise intervention resulted in changes in SPPB (but not the Physical Performance Test, an alternative measure of physical performance). Finally, data from the Women’s Health and Aging Study have demonstrated high test-retest reliability for the SPPB and its three components; data from the National Health and Nutrition Examination Survey (NHANES) III (1988-1994) suggests reliability for tandem stand was poor but acceptable for gait speed and chair stands.

Several additional upper extremity, lower extremity and mixed upper and lower extremity physical performance tests exist. Of these, grip strength probably has the strongest predictive validity for relevant outcomes, including death. Self-reports of

behavior change have also been shown to be predictive of subsequent changes in physical functioning, independent of physical performance tests.

Measuring Participation/Engagement

Dr. Janet Fast provided an overview of measurement of participation and engagement by older adults. A wide range of approaches have been used to measure whether, and the extent to which, an older person is participating in those activities believed to contribute to aging well. In many national surveys, as well as smaller scale studies, respondents are asked directly whether they participate in a particular activity or set of activities, usually within a specified time frame, such as the last week, month or year. For example, the following examples from Statistics Canada's General Social Survey illustrate typical question structure:

- In the past 12 months, did you help anyone by doing domestic work, home maintenance or outdoor work? (GSS 12)
- In the past 12 months, did you go to a cultural or artistic festival? (GSS 12)
- In the past 12 months, did you do unpaid volunteer work for any organization? (GSS 17)
- In the past 12 months, were you a member or participant in a political party or group? (GSS 17)

While such questions may provide a sense of the *rates* of participation in activities of interest, they are not good measures of the *intensity* of participation which is likely more relevant if one is interested in engagement in activities that are believed to promote aging well. Estimates of the amount of time spent participating in the activities are preferable for this purpose.

There has been a great deal of attention paid by those in the time use research community to how best to measure how people spend their time. Numerous approaches have been tested over the years--observation, shadowing, random time sampling, time diary and stylized methods among others. The latter two are the most common and the general consensus in the time use research community is that the diary method is preferred, especially when the purpose is to get an accurate picture of daily time allocation patterns of a population or sub-population.

Stylized estimates are used most often when the purpose is to determine how much time is usually spent on a single, specific activity. The following excerpt from the 1996 and 2001 Census of Canada illustrates the typical nature of stylized time use questions:

Last week how many hours did you spend doing the following activities:

- Doing unpaid housework, yard work or home maintenance for members of this household or others?
- Looking after one or more of your own children, or the children of others, without pay?
- Providing unpaid care or assistance to one or more seniors?

The stylized approach is likely simpler than a time use diary, but it has been criticized for its lack of accuracy. Research comparing estimates from diary and stylized methods suggest that stylized questions tend to produce under-estimates of time spent on frequent activities and over-estimates of time spent on infrequent or episodic ones. Accuracy also tends to deteriorate the longer the period over which the respondent is asked to recall. A modified version of the stylized approach was developed for use in Cycle 16 of the General Social Survey which focused on help provided to seniors. It comprised a series of questions intended to enhance the accuracy of respondents' recall of their caregiving work.

- In the past 12 months, have you assisted anyone with a health or physical limitation by [providing personal care such as assistance with bathing, toileting, care of toenails/fingernails, brushing teeth, shampooing and hair care or dressing]?
- What was the reason for providing assistance with these activities?
→ Because of their long-term health or physical limitations
- During the time that you assisted [care receiver], how often did you assist them with these tasks? Was it daily, at least once a week, at least once a month, less than once a month?
- What is the number of times [daily/weekly/monthly] that you assisted [care receiver] with these tasks?
- About how much time do you spend assisting [care receiver] with these tasks on each occasion?

Time diaries have become standard elements of the national statistics gathering systems in most developed, as well as many developing, countries around the world. They share common features but are implemented in slightly different ways in different countries. In Canada and the United States, for example, the data are collected over the telephone using a 24 hour recall diary. The following represents the typical structure of the interview in the Canadian survey:

- On [designated day] at 4 a.m. what were you doing?
- And then what did you do?
- When did you start?
- How long did you spend on this activity? When did this end?

In most countries, though, paper diaries are dropped off to the respondent who then records his/her activities, often for specified intervals (e.g., 15 minutes), during the designated period (from one to seven days), at the end of which the diaries are picked up. An interview in which basic demographic and other relevant information is obtained often is conducted when the diary is picked up or dropped off. Surveys also vary across countries with respect to the number of household members from whom a diary is obtained, and what contextual information is requested about the activity episodes (e.g., where they took place, who the respondent was with at the time, for whom the activity was done, secondary activities the respondent was doing simultaneously, etc.).

Time diaries are not without their problems, however. Both the incidence and duration of infrequent and episodic activities, such as volunteer work and caregiving, tend to be under-represented when diaries are collected for only one or two days. This

is what led to development of the modified stylized approach for the survey on social support described above.

Based on participation and engagement data, Dr. Fast also presented evidence that most Canadians are reasonably active and engaged, and becoming more so. She found, for example, that seniors remain engaged in productive activity into later life, indeed appearing to make a partial substitution of one form of productive engagement for another on retirement. Further, both passive and active leisure were found to increase across time for men and women of all birth cohorts and ages, with increases in passive leisure smaller for more recent cohorts and active leisure consistently greater than passive leisure.

Linking Measures of Assistive Technology and Disability

Dr. Agree presented results from a project funded by the U.S. Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation (ASPE) to develop and disseminate a set of questions on assistive technology use and the home environment for national surveys on health and aging.¹ The project involved an extensive review of existing measurement approaches, consultation with stakeholders in policy and national surveys, discussions with technical advisory group members, cognitive testing at the National Center for Health Statistics and a pilot study conducted by Westat of 360 people ages 50 and older. Roughly equal numbers of people 50-64, 65-79 and 80+ were administered a 25 minute computer assisted telephone interview. The sample represented persons at all levels of ability and approximately 20 individuals living in assisted living settings were included. Evaluation of the pilot instrument involved both an in depth interviewer debriefing, and coding of 150 taped interviews for key interviewer and respondent behaviors indicative of potential problems. Based on analysis of the pilot results, an eight minute version consisting of several modules was recommended. Few behavior problems were reported for these items (on average less than 3%, except for probing which was approximately 7%, typical for this age group).

Dr. Agree presented results from the pilot and recommendations in four areas: the home environment, assistive device use, the use of global vs. specific items, and effectiveness of assistive technology.

Home Environment. Existing measures of barriers in the home environment are generally found on long home assessment forms used by trained professionals. They rely on subjective assessments of whether features are causing problems. Surveys on the other hand have generally asked few questions about the home environment. Based on the cognitive testing and pilot results the research team recommended a series of questions (approximately 2½ minutes) to capture home features. The items distinguish the existence, acquisition, and use of these features. For example:

¹ The project was funded by ASPE in cooperation with the National Institute on Aging. The research team included Polisher Research Institute, Johns Hopkins University, the National Center for Health Statistics and Westat.

<p>1.2 [Whether or not you use them, does this bathroom have...] grab bars in the shower or tub area? BARBATH <input type="checkbox"/> YES (1) [GO TO 1.2 a] <input type="checkbox"/> NO (2) <input type="checkbox"/> DON'T KNOW (-8) <input type="checkbox"/> REFUSED (-7) } [GO TO 1.3]</p>	<p>1.2a Were these grab bars there when you moved in or were they added? BARADD <input type="checkbox"/> THERE WHEN MOVED IN (1) <input type="checkbox"/> ADDED (2) <input type="checkbox"/> DON'T KNOW (-8) <input type="checkbox"/> REFUSED (-7) [GO TO 1.2b]</p>	<p>1.2b In the last 30 days, when you bathed or showered, did you use the grab bars ... BAR30 <input type="checkbox"/> every time (1), <input type="checkbox"/> most times (2), <input type="checkbox"/> sometimes (3), <input type="checkbox"/> rarely (4), or <input type="checkbox"/> never (5)? <input type="checkbox"/> DON'T KNOW (-8) <input type="checkbox"/> REFUSED (-7) [Read intro "WHETHER OR NOT YOU USE THEM, DOES THIS BATHROOM HAVE ..." AND FILL IN BLANK WITH ITEM IN 1.3]</p>
---	--	---

Assistive Device Use. With respect to assistive device use, there has been a proliferation of questions, using diverse terminology and varying levels of detail. Based on the pilot test, the research team recommended avoiding the terms “special,” “assistive technology,” and “assistive devices,” and instead using the phrase “items that make your daily activities easy, safer, or so you can do them on your own”; naming specific devices and providing definitions as needed; giving specific time frames for use (e.g., in the last 30 days); and not restricting questions to people who report having difficulty. For nonmobility devices, they recommend asking “In the last 30 days have you used (name item).” For mobility devices, they recommended a series of questions to capture frequency of use by task and location. For example:

1. In the last 30 days, did you use ... [INTERVIEWER – IF ANSWER TO HOME 2=5 (NEVER GOES OUTSIDE), SKIP THE THIRD COLUMN OF QUESTIONS ABOUT USING ITEMS OUTSIDE]

<p>1.1 a cane? CANEUSE <input type="checkbox"/> YES (1) [GO TO 1.1 a] <input type="checkbox"/> NO (2) <input type="checkbox"/> DON'T KNOW (-8) <input type="checkbox"/> REFUSED (-7) } [GO TO 1.2]</p>	<p>1.1a. In the last 30 days, when you got out of a bed or chair, how often did you use your cane to help? Would you say ... CANEBED <input type="checkbox"/> every time (1), <input type="checkbox"/> most times (2), <input type="checkbox"/> sometimes (3), <input type="checkbox"/> rarely (4), or <input type="checkbox"/> never (5)? <input type="checkbox"/> DON'T KNOW (-8) <input type="checkbox"/> REFUSED (-7) []</p>	<p>1.1b. In the last 30 days, when you walked around <u>inside</u> your {home/ home or building}, how often did you use your cane? Would you say ... CANEWALK <input type="checkbox"/> every time (1), <input type="checkbox"/> most times (2), <input type="checkbox"/> sometimes (3), <input type="checkbox"/> rarely (4), or <input type="checkbox"/> never (5)? <input type="checkbox"/> DON'T KNOW (-8) <input type="checkbox"/> REFUSED (-7) If never goes out: [Read intro "IN THE LAST 30 DAYS, DID YOU USE ..." AND FILL IN BLANK WITH ITEM IN 44.2]</p>	<p>1.1c In the last 30 days, when you <u>left</u> your {home/ building} how often did you use your cane? CANEOUT <input type="checkbox"/> every time (1), <input type="checkbox"/> most times (2), <input type="checkbox"/> sometimes (3), <input type="checkbox"/> rarely (4), or <input type="checkbox"/> never (5)? <input type="checkbox"/> DON'T KNOW (-8) <input type="checkbox"/> REFUSED (-7) [Read intro "IN THE LAST 30 DAYS, DID YOU USE ..." AND FILL IN BLANK WITH ITEM IN 1.2]</p>
--	---	---	--

The entire assistive device use section takes approximately two minutes to administer.

Global Versus Specific Questions. The research team also investigated whether items that asked simultaneously about the use of multiple devices identified assistive technology users with the same accuracy as the full module. The pilot found that these did not do so consistently across types of devices. A global item for mobility related technology (“In the last 30 days have you used a cane, walker, wheelchair, or scooter?”) had high sensitivity and specificity compared to individual items (0.94 and

0.99, respectively), but global items that referred to home features had much lower predictive power.

Effectiveness of Assistive Technology. Finally, with respect to measures of effectiveness, existing tools are generally clinical assessments about a specific piece of equipment; survey questions have most often evaluated satisfaction with or need for more assistive technology. The research team pilot tested two sets of effectiveness questions. The first set asked about difficulty with specific tasks using the specific devices or features named earlier in the interview but without assistance from another person. For example:

4. [Using your: IF BATH 1.2b=1,2,3,4 DISPLAY "grab bars in the shower or tub"; IF BATH 1.3b=1,2,3,4 DISPLAY "seat in the shower or tub"] h/How much difficulty do you have showering or bathing by yourself? Would you say ...

BTHDIFF

- none (1),
- some (2),
- a lot (3), or
- are you unable to do this by yourself (4)?
- NEVER DOES WITHOUT HELP (5)
- DON'T KNOW (-8)
- REFUSED (-7)

The residual difficulty module takes approximately 1½ minutes. These items had far fewer behavior problems than well established functional limitation items. The items also scale extremely well (Cronbach's alpha=0.8).

A second set of effectiveness items asked individuals to report about specific outcomes related to their assistive technology. Three of the six items (which took approximately 30 seconds to administer) performed exceptionally well: Because you use these items how much safer do you feel when you do your daily activities? Because you use these items how much more control do you have over your daily activities? Because you use these items how much more often do you take part in activities that you enjoy? Preliminary structural equation models suggest that these three items scale well and correlate with the intensity of assistive technology use and the extent of functional limitations, but not with the amount of personal help.

Measures of Environmentally Determined Mobility Disability

Dr. Shumway Cook presented an overview of her research on measures of environmentally determined mobility disability. She explained that although the exploration of the environment as a determinant of behavior is not new and the concept is especially important in models of the disablement process. According to the ICF, for example, environmental factors are external to the individual and interact with health conditions to produce barriers to full participation in society. In the Institute of Medicine's publication "Enabling America" the disabling process occurs when the person's need enlarge relative to the existing environment.

Dr. Cook's research has focused on studying the environmental determinants of mobility disability in aging. Mobility is defined as the ability to move (walk) safely and independently in the environment. Mobility disability results from an interaction of the attributes of the individual (impairments and functional limitations) and attributes of the environment that constrain walking. In order to examine how attributes of the environment contribute to mobility disability in older adults, she developed with colleagues at the National Institute on Aging a model of the environment identifying features within the physical (natural or built) environment that affect the neural organization of a task related movement.

The model encompasses eight dimensions which represent the spectrum of external demand that have to be met for an individual to be fully mobile within a community context. These dimensions include: **distance**, **temporal characteristics** (the need to walk at a certain speed, as for example when crossing a street controlled by a traffic light), **ambient conditions** (light and weather conditions), **terrain characteristics** (both the geometry and surface features), **physical load** (including static loads such as when opening a heavy door, and dynamic loads such as when carrying packages that shift), **attentional demands** (walking while talking to a travel companion, navigating in an unfamiliar environment), **postural transitions** (having to stop, start, change directions, stoop, reach and turn), and **density** (number of people and object in the immediate environment requiring collision avoidance).

Using this model Shumway-Cook and colleagues developed a self-report survey, Environmental Analysis of Mobility Questionnaire (EAMQ), to determine the frequency with which specific features in the environment are encountered versus avoided during routine trips into the community. The instrument was pilot tested with 54 older adults (≥ 70 years), who were recruited from two sites (Waterloo, Ontario, Canada and Seattle, Washington) grouped according to level of physical function (elite, physically able and physically disabled). The self-report survey examined 24 features of the physical environment, grouped into the eight dimensions. Older adults were asked, "On routine trips into the community, how often do you encounter..." Each encounter question was paired with an avoidance question "On routine trips into the community, how often do you avoid..." Frequency of encounter/avoidance was reported using a five-point ordinal scale (never, rarely, sometimes, often, always). Frequency of encounter/avoidance was determined for each individual dimension, and in addition, a total encounter and avoidance score was calculated by summing across dimensions.

Results from the survey pilot study suggested that mobility disability was associated with reduced encounter and increased avoidance of physical features within the environment that impacted walking. However, not all features of the environment were avoided, instead both encounter and avoidance varied by environmental dimensions. Results supported the concept that mobility disability results from an interaction between the individual and the environment. Furthermore, it suggested that some features within the environment were more disabling to mobility than others.

Further research has begun to test the psychometric properties of the EAMQ. To test the reliability of the survey, older adults completed the survey twice, one week apart. The test-retest reliability was good (ICC = 0.81-1.0). The validity of the EAMQ was examined in several ways. Dr. Shumway-Cook and colleagues compared self-report responses to observed behavior in the community, and to level of ADL and IADL disability. As a part of the pilot study older adults were directly observed during six trips into the community to perform activities of daily life (grocery shopping, visit to medical practitioner, social participation, etc.). Frequency of encounters with environmental features within each of the eight dimensions was recorded. Results from the direct observation protocol identified environmental demands associated with mobility in the community. In addition, similar to the survey results, the direct observation study found that mobility disability was not associated with a uniform decrease in encounters across all dimensions, rather certain dimensions were encountered less frequently by the disabled compared to nondisabled, while features in other dimensions showed comparable levels of encounters among the groups. Results from the direct observation protocol supported the survey results in suggesting that certain environmental features may be more disabling than others to community mobility.

In summary, as a survey tool, the EAMQ has a number of strengths. It appears to be a reliable way to determine perceptions related to features within the environment that disable community mobility. It has demonstrated concurrent validity with community mobility directly observed, and is strongly associated with disability in activities of daily life. Current limitations of this survey are its length and the redundancy of questions within each dimension. In addition, it has been tested on only a small sample of older adults, and therefore requires further testing.

CONSIDERATIONS FROM THE FIELD

The final session involved discussion from the perspective of major federal and nonfederal national surveys on aging, disability, and long-term care. The discussion was organized around the following questions:

- What innovations in measurement are you exploring or have you explored already on your surveys?
- Which of the concepts discussed today seem relevant to your survey/national surveys in general? Feasible to include? Completely unworkable?
- What additional pilot/measurement work would you like to see done so new approaches can be readily adopted?
- What constraints are binding in surveys that need to be considered? Time? Subject matter? Format (performance/mode of interview/etc.)?
- What would it take administratively to modify an existing instrument? How much lead time is generally needed to make modifications? Can you give ballpark costs for different types of changes (e.g., add a module; modify existing questions; add to existing questions)?

Dr. Kenneth Manton provided an overview of the National Long Term Care Survey (NLTC). He explained that since its inception in 1982 the survey has emphasized stability of health conditions and disability measures. One of the main purposes of the survey is to examine change in human capital and disability in elderly. The focus is purposefully demographic in order to provide accurate estimates for policy makers of Medicare and Medicaid use and the need for long-term care services.

Dr. Manton provided a reaction to the use of vignettes. He wondered how they would fair during the Office of Management and Budget's (OMB) clearance process because of the potential respondent burden. He also questioned whether the approach could be used validly with individuals who have cognitive impairment. On the methodological note, he questioned whether results were sensitive to the choice of standardization and whether the uncertainty of the parameter estimates was taken into account when setting the thresholds.

In terms of adding innovations to the NLTC, Dr. Manton described three important areas. First, they have had a consistent set of measures over time, but have the ability to add supplemental questions so that new measurement approaches can be explored. Second, they have over-sampled the 95 and older population so that robust estimates can be made for this group. Third, they have linked to Medicare data so that transitions between survey years can be inferred.

D.E.B. Potter then provided an overview of the Medical Expenditure Panel Survey (MEPS). She explained that everyone in the household, not just older respondents, is interviewed. The main purpose of the MEPS is to provide calendar year estimates of medical expenditures and use. As such, the survey includes a variety of measures of health but not detailed disability measures (e.g., no questions about individual tasks or activities).

In the coming year MEPS will be changing its CAPI from a DOS environment to a Windows environment. They are going into the field with a split panel to pre-test the new environment. They are also developing new questions to measure caregiving in the household and to quantify the costs of caregiving.

Ms. Potter commented on some of the innovations presented earlier in the day. She echoed some of the concerns raised by Dr. Manton about whether the OMB clearance process might view vignettes as burdensome. She also suggested that Internet pilot testing might be of concern in older age groups because the mode might be sensitive to cognitive ability.

She suggested that in the future it would be important to investigate the influence of question order. For example, if physical functioning measures are included in a survey with self-reported items, how does order influence performance and/or reports? She suggested that items tapping unmet need are important as is obtaining a balance of physical and cognitive disability items.

Ms. Potter also reviewed a list of potential constraints in adding innovations to existing surveys. Budgetary constraints are clearly always of concern. For some agencies Congressional priorities drive how agency resources are allocated. Funds are needed not only for data collection but for development work as well. For the MEPS, there is a steering group that agrees to survey additions and modifications. Currently if items are added, other items must be dropped. A lead time of approximately two years is needed to introduce items at the beginning of a calendar year.

Dr. Julie Weeks elaborated upon some of the constraints raised by other panelists. She explained that for some surveys many agencies (with potentially competing interests) may be involved so it can be challenging to maintain consistency in items over time. In her experience with surveys of older adults, she has found fatigue to be a major constraint that may preclude efforts to cross-walking multiple measurement approaches. Time constraints, funding constraints, and limits as to what proxies can report about older respondents with disabilities were identified as issues.

Dr. Weeks also provided a brief overview of three federal review processes that govern federal surveys--the Institutional Review Board, which ensures human subjects protection; OMB, which ensures that burden is not excessive; and the Disclosure Review Board, which ensures that public use data are not identifiable. These are not constraints in and of themselves but may require additional time and resources.

Dr. Weir provided an overview of disability measures in the Health and Retirement Survey. David Weir. Core measures in telephone interviews every two years include functional limitations, ADLs, IADLs, work limitations, housework and other limitations. He also described several innovations. In 2004, for example, SSA funded an in-person interview to increase consent for record linkages and the National Institute on Aging provided funds to include physical performance measures (grip strength, puff test, and times walk) and drop off questionnaires that included work limitation vignettes. The latter had an 80% response rate and data will be released in summer 2005.

In the near future, the HRS plans to include vignettes in its internet sample. In addition, one of the experimental modules in 2006 will be devoted to measuring modifications to the built environment designed by Dr. Vicki Freedman and Dr. Emily Agree. Also proposed for 2006 through 2010 are plans for one-third of the HRS to be interviewed in person and for performance measures and biomarkers to be collected. The performance measures collected in 2004 had good response rates (74.8%) even though individuals were given several opportunities to refuse, although preliminary analysis of the data suggests that there may be some selection according to disability status.

A number of additional points were raised by audience members:

- There seems to be important differences in the ease with which innovations can be added to government and nongovernment surveys. The processes and constraints appear to be different. However, some of the limitations raised by the panel may be overstated. For example, NHANES is able to make changes to its CAPI and put data out every two years.
- Survey content is driven in part by the audience for the survey. Policy-relevant surveys serving Congressional needs will have different measures than surveys designed to meet basic and social science research needs.
- The question of whether innovations are being driven out by the need for consistency over time was raised. Audience and panel members suggested that this was not the case. One audience member suggested that surveys can be modified if the sample is large enough to undertake a “split sample” design, where the old and new questions are both asked for several years. Another audience member suggested that for measurement innovations to be accepted it would be important to demonstrate that the “new” items are more valid and reliable than the old. In making such changes it was suggested that the “old” questions appear first and newer questions placed at the end so that old questions would not be contaminated. Other creative ways to introduce and test new questions quickly include the experimental module (in which a random subset is administered the questions at the end of the main survey) and an Internet panel (in which respondents answer internet-based questionnaires).

OPPORTUNITIES TO IMPROVE SURVEY MEASURES OF LATE-LIFE DISABILITY

Files Available for This Report

Part I--Workshop Overview

HTML Version <http://aspe.hhs.gov/daltcp/reports/2006/survmeasI.htm>

PDF Version <http://aspe.hhs.gov/daltcp/reports/2006/survmeasI.pdf>

Part II--Workshop Summary

HTML Version <http://aspe.hhs.gov/daltcp/reports/2006/survmeasII.htm>

PDF Version <http://aspe.hhs.gov/daltcp/reports/2006/survmeasII.pdf>