
Assessing Functional Status and Disability in Epidemiologic Studies

7

Jack M. Guralnik, Kushang Patel,
and Luigi Ferrucci

Abstract

In epidemiological studies of older populations, the assessment of functional status and disability plays an important role in comprehensively characterizing the population. Epidemiological research in aging has used functional measures to examine the overall health status of older adults, understand the consequences of chronic diseases and behavioral risk factors, and identify individuals at high risk of a variety of adverse outcomes. Important domains are personal roles, social roles and physical capacity. The selection of instruments and questions is guided by the specific population in which they will be used and the objectives of the functional assessment. A variety of approaches have been used to meaningfully summarize multiple disability items. Objective physical performance measures, most of which assess physical capacity, offer a means of assessing functioning in a standardized environment and have found many uses in aging research. Describing trajectories of functioning over time, including both improvement and decline, is crucial for understanding the disablement process. Population trends in disability are a critical measure of health status changes in the older population. Active life expectancy, a

J.M. Guralnik, M.D., Ph.D (✉)
Department of Epidemiology and Public Health
University of Maryland School of Medicine,
Howard Hall, Room 200, 660 W. Redwood Street,
Baltimore, MD 21201, USA
e-mail: jguralnik@epi.umaryland.edu

K. Patel, Ph.D.
Laboratory of epidemiology, Demography and biometry,
National Institute on Aging, 7201 Wisconsin Ave., Room
3C-309, Bethesda, MD 20892, USA
e-mail: patelku@mail.nih.gov

L. Ferrucci, M.D., Ph.D.
Intramural Research Program, National Institute
on Aging, 251 Bayview Blvd, Suite 04C228
Baltimore, MD 21224, USA
e-mail: ferruccilu@grc.nia.nih.gov

measure of the remaining years of life from a specific age that are free of disability, reflects both survival and length of high-quality life. Its extension is an important public health goal.

Keywords

Aging • Epidemiology • Geriatrics • Older adults • Longevity • Functional status • Disability • Epidemiologic studies • Assessment • Physical performance • Measurement • Proxies • Self-report • Mobility disorders

Abbreviations

| | |
|--------|---|
| ADL | Activity of daily living |
| ANCOVA | Analysis of covariance |
| CAT | Computer adaptive testing |
| CI | Confidence interval |
| IADL | Instrumental activity of daily living |
| IL | Interleukin |
| IRT | Item response theory |
| LIFE-P | Lifestyle Interventions and Independence for Elders Pilot |
| NHANES | National Health and Nutrition Examination Survey |
| NHATS | National Health and Aging Trends Study |
| PEP | Precipitating Events Project |
| PF-10 | Short Form 36 Health Survey physical functioning subscale |
| PPT | Physical Performance Test |
| SF-36 | Short Form 36 Health Survey |
| SPPB | Short Physical Performance Battery |
| US | United States |
| WHO | World Health Organization |

social and environmental factors in the older population can best be understood by evaluating the functional status of the individual. Assessing functional status in older adults is useful because it is:

- Prognostic of adverse medical outcomes
- Prognostic of further decline in function and loss of independence
- An indicator of level of independence
- An indicator of need for formal and informal care
- A global measure of the burden of disease
- A relevant measure of quality of life.

A large body of epidemiologic research over the past three decades has treated disability as a condition that can be studied in much the same way as is a well-defined chronic disease: by using epidemiologic tools to assess prevalence, incidence and a wide range of risk factors. This work has led to a greater understanding of the occurrence, determinants and consequences of disability in the older population and has provided insights into strategies for the prevention of disability.

7.1 Introduction

While much epidemiologic research is targeted at understanding the distribution and determinants of a particular disease or pathological process, a hallmark of epidemiologic research in aging is the study of the functional consequences of diseases and physiologic changes that occur in aging. Although the epidemiologic study of individual diseases that are important in aging is a high priority, the joint impact of diseases and multiple other physiological, psychological,

7.2 Models of Disablement and Domains of Function

7.2.1 Models of Disablement

Epidemiologic research has found a long list of diseases to be associated with disability onset. The most important disease categories include musculoskeletal, cardiovascular and neurologic conditions. In certain cases, such as stroke or hip fracture, it is obvious how disability results from the disease. In others, however, the onset of

disability may occur over years and the effects of a single or multiple diseases, health behaviors and other risk factors (e.g., obesity, social isolation) are more difficult to understand. In these cases, epidemiologic research into the mechanisms and pathways that lead to disability can be substantially enhanced by modeling steps along this pathway that represent the consequences of disease on impairments and body functions. Ultimately, disability is best conceptualized as the gap between an individual's physical/cognitive capabilities and the demands of the environment, which leads to an inability or difficulty in fulfilling one's social or role functions.

The most prominently employed model of the disablement process was originally developed by Nagi [1] and updated to include environmental influences by Verbrugge and Jette [2]. In 1980, the World Health Organization (WHO) introduced a model that included disability and handicap. More recently, the WHO updated this disability model with a framework that introduces new terminology and is less focused on the disablement process or pathway. Instead, it proposes that disability represents a complex interrelationship of factors [3]. A new model, based on both the Nagi and WHO approaches, was begun in 2011 [4]. It was developed for the National Health and Aging Trends Study (NHATS) and funded by the United States (US) National Institute on Aging. NHATS is a new panel study of individuals ≥ 65 years of age that will support research on national trends in late-life disability and factors that influence these trends. The steps in these three models are shown in Fig. 7.1.

Most empiric research that addresses the pathway from disease to disability has relied on the Nagi framework, which operationalizes the steps so that specific assessments can clearly be classified as to where they fit in the pathway. In the Nagi model, impairments are defined as dysfunction and structural abnormalities in specific body systems (e.g., cardiovascular system, musculoskeletal system), functional limitations are defined as restrictions in basic physical and mental actions (e.g., ambulation, reaching, grasping, climbing stairs, speaking) and disability is defined as difficulty in or inability to perform life activities

(e.g., self-care, household management, jobs, hobbies) [2]. Functional limitations, which are often measured using objective tests of physical performance, are a critical link between impairments, such as poor strength, and disability, such as difficulty transferring from bed to chair or shopping.

The WHO model does not specifically use the term disability, but considers it a general term for the whole process. Instead, it introduced the term "participation" to indicate the degree to which an individual can engage in the activities that are important and relevant to their life. The strengths of the WHO framework include:

- It changes the language we use in approaching disability, which may have a strong effect on perceptions and reactions to disability
- Main domains are framed in a neutral way so that each has positive as well as negative aspects and can represent individuals across the functional spectrum
- The concept of participation provides a strong focus on the ultimate goal of maximizing the opportunities for all individuals with limitations
- There is clear emphasis on the effect of the environment and society on the experience of disability.

However, the WHO model has not been used to any degree by epidemiologists who study the disablement process because it does not outline a clear pathway to be investigated; does not clearly operationalize the components of the model; and while it defines activities and participation differently, it does not discriminate between them in its combined, very-detailed list of activities and participation, so further guidance is needed as to how to discriminate these two domains.

The NHATS model provides clear steps in the pathway to disability, which meet the purposes of the study by operationalizing the concepts so that underlying factors that are responsible for trends in disability can be easily understood. Measurements of both capacity (the same as functional limitations in the Nagi model) and accommodations are important for understanding trends. For example, if improvements are seen over time in self-report of ability to carry out self-care

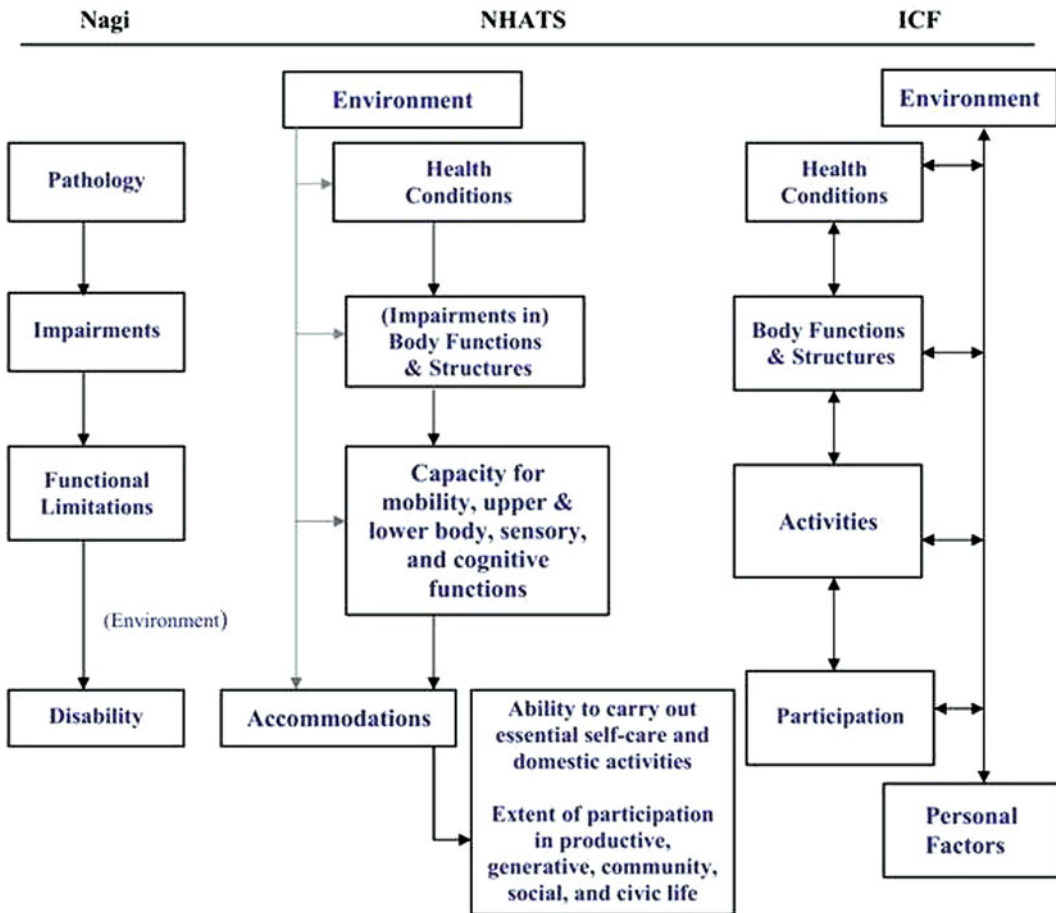


Fig. 7.1 The Nagi International Classification of Functioning, Disability and Health (ICF) and National Health and Aging Trends (NHATS) disability frameworks [4]

or domestic activities, capacity measures can be assessed for changes and—if they have not improved—measures of accommodation may reveal improvement, indicating that accommodations rather than better physical function are responsible for the decline in disability. This model is similar to the Nagi model, but it has not been specifically evaluated with empirical data, which will be possible when the study has collected data on a representative US population.

7.2.2 A General Approach to Domains of Function

Whatever disability framework may be used, there are basic domains of functioning that are

relevant in the older population. Function itself has multiple domains, including physical, cognitive, sensory, psychological and social. Within physical functioning, three general domains are highly relevant for the older population: *personal roles*, *social roles* and *physical capacity*. The most commonly assessed disabilities related to *personal role* are self-care tasks, known as activities of daily living (ADLs). The items assessed usually include the following:

- Eating
- Dressing
- Bathing
- Transferring from bed to chair
- Using the toilet.

Also critical for personal roles are tasks that are required for independent living in the

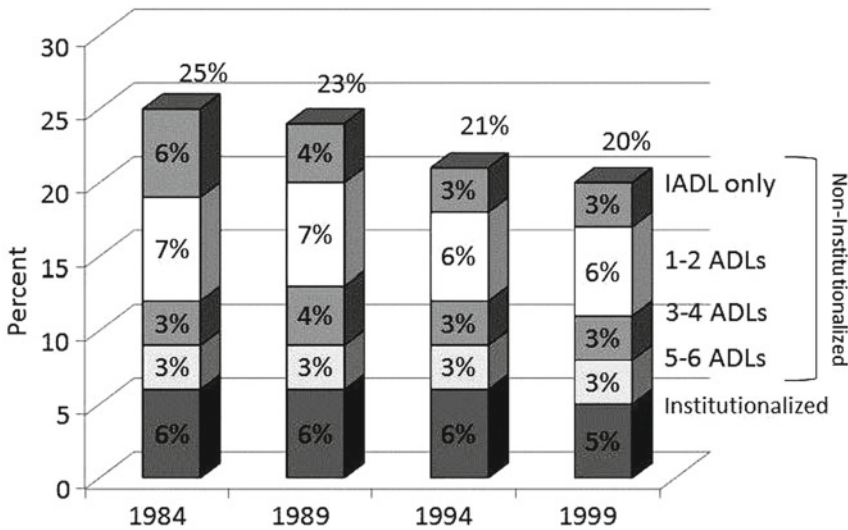


Fig. 7.2 Age-adjusted percentage of Medicare enrollees ≥ 65 years of age who are chronically disabled (receiving help, needing supervision, using equipment or not able),

by level and category of disability: 1984, 1989, 1994, 1999 [5]. Abbreviations: ADL activity of daily living; IADL instrumental activity of daily living

community. These are termed instrumental activities of daily living (IADLs) and generally include the following:

- Preparing meals
- Shopping
- Housekeeping
- Managing money
- Taking medications
- Using the telephone.

There is generally a hierarchical relationship between ADLs and IADLs, such that nearly everyone with disability in ADLs also has IADL disability. National and local disability statistics often take this into account when presenting rates of disability in these areas. Figure 7.2 shows US data that illustrates the percentage of the US population with disability in specific numbers of ADLs (who generally have IADL disability as well), the percentage with IADL disability only, and the percentage who live in long-term care facilities.

Social roles are an important component of a satisfying and full life, and restrictions in participation in these roles due to illness are an important aspect of disability. There has been less research and less assessment of these roles in national surveys than for the ADLs and IADLs. An instrument that illustrates these kinds of roles is the Late Life Disability Instrument, which

includes ADL- and IADL-type measures but also has a strong emphasis on social roles [6]. These social role functions include:

- Visit friends
- Travel out of town
- Go out to public places
- Work at a volunteer job
- Keep in touch with others
- Participate in social activities
- Invite family and friends into home
- Participate in active recreation
- Provide assistance to others.

The third general domain of function is *physical capacity*. In developing his model, Nagi saw this as representing functional limitations and he developed questionnaire items that have been extensively used to represent this domain. These include:

- Pulling or pushing a large object, such as a living room chair
- Stooping, crouching or kneeling
- Lifting or carrying weights under 10 pounds
- Lifting or carrying weights over 10 pounds
- Reaching arms above shoulder level
- Handling small objects.

Mobility is also an important aspect of physical capacity. Numerous questionnaire batteries and objective performance tests have been used to

assess walking and lower extremity functioning. The Short Form 36 Health Survey (SF-36) [7] is a commonly used scale of overall health status, but a large part of the SF-36 is directed toward physical functioning (physical functioning subscale [PF-10]). Most of the questions on the PF-10 assess capacity, with a majority of those aimed at mobility assessment. The PF-10 includes:

- Vigorous activities, such as running, lifting heavy objects or participating in strenuous sports
- Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling or playing golf
- Lifting or carrying groceries
- Climbing several flights of stairs
- Climbing one flight of stairs
- Bending, kneeling or stooping
- Walking more than a mile
- Walking several hundred yards
- Walking 100 yards
- Bathing or dressing oneself.

Many self-report items related to capacity were evaluated in the development of the Late Life Disability Instrument's function domain [8]. It was found that they aggregated into three specific domains, which include:

- Upper extremity function (e.g., unscrew lid without assistive device, pour from a large pitcher, use common utensils, reach behind back)
- Basic lower extremity function (get into and out of a car, get up and down from a curb, get on and off a step stool, stand up from a low soft couch)
- Advanced lower extremity function (carry while climbing stairs, walk a brisk mile, run to catch bus, walk on slippery surface).

Many aspects of capacity are addressed using measures of physical performance that are assessed using standardized tests in a controlled setting. A wide range of performance measures have been developed, and the reason they work well is probably because they likely capture the impact of multiple chronic conditions and physiologic changes that are related to aging and a sedentary lifestyle. Physical performance tests have

been found to predict multiple adverse outcomes, including mortality. Even a simple measure, such as usual gait speed tested over a short course, has a strong relationship with multiple characteristics of the study participant and is highly predictive of adverse outcomes. A meta-analysis that included gait speed data from nine cohorts showed a clear and consistent relationship with mortality risk (Table 7.1) [9]. There was a steep gradient of 5-year survival across baseline gait speed for men and women in each of three age groups, the greater baseline gait speed being associated with greater 5-year survival. For example, for age group 75–84, comparing those with the slowest gait speed (<0.4 m/s) to those walking at 1.4 m/s or faster, the proportion surviving for 5 years was 60 and 93%, respectively, in men and 69 and 95%, respectively in women. Physical performance measures will be further described later in the chapter.

7.3 Assessment by Self- and Proxy-Report

Disability is most often assessed using a self-report format in which the patient or study participant is asked to respond to a series of closed-ended questions which have specific response categories. If a study participant is too ill or too cognitively impaired to provide responses, a proxy may be used for certain types of questions. Disability assessment was originally used in the clinical setting, most often to follow-up patients in rehabilitation programs. However, over the past few decades, its value has developed in assessing the health of the population as a way to characterize older adults in epidemiologic studies and, eventually, as an outcome in clinical trials. Various methods of assessing disability were created for use in different types of studies, but standard sets of items and instructions on how to ask questions regarding these items have never established. The basic ADLs were proposed by Katz et al. in 1963 [10], but even his short list has been modified over time (e.g., his incontinence item is generally not included in current ADL disability definitions).

Table 7.1 Percentage (95% confidence interval) of men and women surviving 5 years by gait speed and age group

| Gait speed m/s | Men | | | Women | | |
|-----------------|------------|------------|------------|------------|------------|------------|
| | Age 65–74 | Age 75–84 | Age ≥85 | Age 65–74 | Age 75–84 | Age ≥85 |
| <0.4 | 68 (47–82) | 60 (38–76) | 25 (15–36) | 80 (71–86) | 69 (58–78) | 47 (40–54) |
| ≥0.4 to <0.6 | 77 (72–81) | 57 (49–64) | 31 (24–39) | 88 (85–90) | 75 (68–80) | 61 (50–70) |
| ≥0.6 to <0.8 | 79 (74–83) | 65 (57–71) | 49 (35–61) | 91 (89–93) | 82 (78–86) | 74 (69–78) |
| ≥0.8 to <1.0 | 85 (82–88) | 75 (69–79) | 54 (43–64) | 93 (91–95) | 89 (86–91) | 73 (59–83) |
| ≥1.0 to <1.2 | 90 (85–93) | 83 (76–87) | 68 (57–77) | 96 (94–98) | 91 (87–94) | 61 (35–79) |
| ≥1.2 to <1.4 | 93 (86–96) | 85 (79–89) | 62 (46–74) | 96 (94–97) | 93 (87–96) | 67 (5–95) |
| ≥1.4 | 95 (89–97) | 93 (86–96) | 91 (51–99) | 97 (94–99) | 95 (72–99) | NE |
| All gait speeds | 87 (82–91) | 74 (65–81) | 46 (39–53) | 93 (91–94) | 84 (80–87) | 64 (58–70) |

Source: Adapted from Studenski et al. [9]. Abbreviations: NE not estimable due to small numbers of participants in categories

In addition to item selection, choices on how to ask the questions and categories of response were often newly created for each survey or epidemiologic study, limiting the ability to compare disability rates across studies. It is certainly important to use measures that meet the requirements of a study, but even in a small study with a narrow focus, the use of a battery that is used for national surveys enables one to evaluate how the study sample compares to the general population of older adults. In the following subsections, we will discuss a number of issues that exist regarding decision-making for the selection of self-reported disability assessments.

7.3.1 Instrument and Item Selection

When planning a new study, an initial decision must be made whether to utilize an existing battery of items or to pick and choose a specific set of items to meet the study requirements. As in any measurement decision, is advantageous to retain a full battery of items when the battery has had extensive methodological work on validity, reliability and sensitivity to change because selected components of the battery will not have these characteristics. A number of considerations (outlined in Table 7.2) go into determining what kinds of disability items are appropriate for a particular situation. At the most basic level, items used in a clinical setting may be different than those that are most appropriate for a research setting that is evaluating a specific population

group. A clinical setting may have the advantage of a trained medical professional, such as a nurse or physical therapist, compared to an epidemiologic research setting that uses assessors with little medical training and in which assessor decision-making about the study participant needs to be minimized. In studies that have many assessors, especially multi-site studies, full training and quality control should be implemented to maximize inter-rater reliability and avoid site-specific deviations in assessment that can result in biases.

A key consideration in item selection is the general health status of the population under study. For example, although it is extremely important to identify ADL disability, only about 10–15% of the total population ≥65 years of age will have ADL disability. Therefore, if no other items are selected, there will be no way to discriminate across the functional spectrum for the remaining 85%. While ADL assessment alone may be adequate for individuals who live in a nursing home or assisted living facility, more information is critical for understanding the functional level and needs of older adults who live in the community. Other characteristics of the study population (e.g., a specific disease that is the focus of the research, impairments of the study sample, educational level) may play a role in the selection of items, construction of questions and mode of administration. A study of a population with arthritis and a study of a population with heart disease may use the same general measures of disability, but each study may choose to add

Table 7.2 Selecting measures of functioning

| Measure specifics | Information |
|---|--|
| Individual, in clinical setting | Minimize false positives and false negatives, even at the expense of ease and economy of administration |
| Group or population, in research setting | Often administered by multiple assessors with lower levels of professional training Requirements: <ul style="list-style-type: none"> • High levels of inter-rater reliability • Minimize dependence on professional judgment • Be able to administer in reasonable time, with minimal burden on the respondent |
| Must consider general health status of population to be studied | May differ by use – describe population vs. study functional consequences of specific disease or intervention Nursing home Disabled in community Healthy aging Community-dwelling, heterogeneous |
| Specific characteristics of population | population with a specific disease Impairments that affect functional evaluation (hearing impairment, cognitive impairment) Educational status |
| Cross-sectional vs. longitudinal measurement | For longitudinal studies, must be sensitive to change Continuous measures may be more sensitive to change the dichotomous measures |

items that assess aspects of functional loss that are specific to the respective diseases. Finally, the selection of disability measures should reflect whether they will be used only in cross-sectional analyses or whether assessment of longitudinal change will be an important study goal. For example, it is important to identify the subset of a population that has ADL disability at the baseline of a study, but in the remainder of the population there can be a lot of functional change over time that will not be severe enough to be detected by ADL measures.

7.3.2 Formulation of Questions

Three general issues must be considered when determining the formulation of disability questions:

- The actual performance of a task vs. the capacity to perform
- The time interval for the respondent to consider
- Query difficulty, the need for help, or both.

For basic self-care tasks, virtually everyone will need to perform or get help to perform all of the tasks. More difficult tasks may be discretionary. Getting a negative response to a question about whether an individual has walked a quarter mile in the past month may give concrete evidence that the activity was not performed, but it gives little indication as to whether this individual simply didn't choose to do the task, didn't have the opportunity to do the task, or was incapable of doing the task. On the other hand, asking whether the individual is able to do the task may force them to make a decision about their potential ability or capacity to do the task when they have not actually attempted it recently. There is no simple solution to this dilemma. Some surveys query both aspects by asking those who report not doing a task if they believe they could do it if they tried. When asking about discretionary kinds of tasks, it is advantageous to add to the question a phrase about whether they don't do it "for health reasons", particularly when individuals report that they are unable to do tasks such as IADLs, which they may have never done.

Research has shown that there is a great deal of short-term change in functioning, so the time interval that the study participant is asked to consider may have a substantial impact on their response. Many disability instruments simply query the current time (e.g., “do you now have difficulty...”), which is easiest for the respondent. However, more disability will be identified if a time interval is attached to the question, and asking about “in the past year” will give higher prevalence rates than asking “in the past month.”

Response categories that are used in disability questions vary across studies and types of questions. The simplest outcome to assess is whether the participant needs help to perform a task, with the option of clarifying whether this means help from a person, help from equipment or both. An alternative to asking whether help is needed is to ask whether help is received, but it should be kept in mind that there are individuals who need help who don’t receive it, and they will not be identified if the question is worded to ask about receiving help. Needing or receiving help is at the more severe end of the disability spectrum and there is some advantage to using questions about level of difficulty to characterize the remaining population that does not need or receive help. Questions about difficulty should include a response option to identify that the respondent is unable to do the task in order to not miss this important subgroup that is not able to report on difficulty (response categories to questions about difficulty can be: none, a little, some, a lot, unable to do).

Prevalence estimates of disability in epidemiologic studies and population surveys can be substantially affected by the way that questions are asked and the possible response categories provided. This is illustrated in Table 7.3, which shows percentages of those with disability in a survey for three different assessment outcomes of ADL and mobility items [11]. Looking at a summary of these measures, the percentage with a positive response for 1 or 2 items was 15.0% for difficulty, 9.0% for human assistance and 12.6% for human or mechanical assistance. The percentages for a positive response on three or more items were 8.7, 4.2 and 9.4%, respectively.

7.3.3 Innovative Approaches to Self-Report

In addition to the traditional assessments of the need for help and difficulty with disability items, recent years have seen the development of innovative approaches to assessment. For example, it has been shown that individuals who don’t report disability by traditional criteria may report that they have modified the way they do a task or how often they do it. In fact, a substantial percentage of individuals who do not report difficulty will say that they have changed the way they do an activity (e.g., climbing stairs, doing housework) or that they do the activity less often. The predictive validity of this approach has been demonstrated by showing that this subgroup has intermediate rates of adverse outcomes compared to those with difficulty and those without difficulty or modifications [12]. An additional approach to assessment in individuals who report no difficulty is to query how easy it is to do a task, such as walking a quarter mile. Table 7.4 shows objective performance test results for non-disabled individuals (report no difficulty walking ¼ mile) that report whether walking ¼ mile is not so easy, somewhat easy or very easy, and whether or not the participant is doing the task less often [13]. It is clear that for all of these performance tests, there is a gradient of function that increases as the task is perceived to be easier and if the individual has not resorted to doing the task less often.

The use of video images to demonstrate an activity solves some of the difficulties individuals may have in deciding how to answer questions about activities that they have not recently performed [14]. In this approach, stick figure images are shown doing multiple mobility tasks, calibrated for specific speeds of walking and stair climbing. The animations also include certain environmental challenges, such as carrying bags of groceries, stepping across obstacles and walking over rocky terrain. This approach still relies on someone’s self-perception of their capacity, but it standardizes the tasks on which participants are reporting.

Table 7.3 Disability estimates for New England states using three different scaling methods

| ADL activity | Difficulty/does not do | | Human assistance/does not do | | Human or mechanical assistance/does not do | |
|--------------------|------------------------|--------------|------------------------------|------------|--|--------------|
| | % | 95% CI | % | 95% CI | % | 95% CI |
| Walking | 8.1 | (6.9, 9.4) | 1.6 | (1.0, 2.2) | 10.0 | (8.9, 11.7) |
| Bathing | 12.2 | (10.7, 13.7) | 8.6 | (7.3, 9.9) | 12.5 | (11.0, 14.0) |
| Dressing | 7.3 | (6.1, 8.5) | 5.0 | (4.0, 6.0) | 5.3 | (4.2, 6.3) |
| Eating | 3.2 | (2.4, 4.1) | 1.2 | (0.7, 1.7) | 1.2 | (0.22, 2.2) |
| Getting to toilet | 4.3 | (3.3, 5.2) | 2.0 | (1.4, 2.7) | 7.5 | (6.3, 8.7) |
| Bed/chair transfer | 13.4 | (11.8, 15.0) | 3.0 | (2.2, 3.8) | 5.5 | (4.4, 6.5) |
| Getting outside | 9.7 | (8.3, 11.0) | 7.8 | (6.6, 9.0) | 15.4 | (13.7, 17.1) |

Source: Adapted from Jette [11]. Abbreviations: ADL activity of daily living; CI confidence interval

Table 7.4 Mean performance by category of reported functioning in walking ¼ mile among persons reporting no difficulty. Health ABC Study

| Difficulty/modification | 6 m usual walk (m/s) | SPPB score | Completed 400 m walk (%) |
|-------------------------|----------------------|------------|--------------------------|
| <i>How easy?</i> | | | |
| Not so easy | 0.94 | 8.69 | 41.7 |
| Somewhat easy | 1.09 | 9.48 | 62.5 |
| Very easy | 1.21 | 10.28 | 81.9 |
| <i>Do less often?</i> | | | |
| Yes | 1.07 | 9.39 | 64.2 |
| No | 1.19 | 10.13 | 77.4 |

Source: Adapted from Simonsick et al. [13]. Abbreviation: Health ABC health, ageing and body composition; SPPB short physical performance battery

A particular challenge in using self-report measures is comparing rates across countries, ethnic groups and racial groups. This is due to differences in the way that individuals perceive the questions and relate them to the kinds of activities in which they engage. In the section on physical performance measures (Sect. 7.5), there is a discussion about using objective measures to calibrate responses to self-report questions. Researchers have also used vignettes to anchor assessments of health states. In this approach, a hypothetical person with a specific problem is presented and the respondent is asked to evaluate that hypothetical person using the same format grading scale used to assess respondents. The vignette standardizes the specific nature and severity of a condition, and enables the detection of different reporting styles across countries or cultures. These differences can then be used to adjust observed differences in rates of self-reported disability. This technique and others are

described in a comprehensive overview of innovative methods for approaching the measurement of disability in older adults in population surveys. The overview is published by the National Research Council and based on a meeting it convened with leaders in the field [15].

7.3.4 Mode of Administration

Various modes of administration are available for ascertaining self-reports of functioning and disability in survey research, epidemiologic studies and clinical research. Most assessment tools were developed for interviewer administration, but many have been shown to be easily adapted for telephone administration and for self-administration using a paper form at the study site or at the participant's home. Questionnaires that are completed by the participant have a number of advantages [16]: participants can complete the form

when convenient to them; their use reduces staff time spent in administering interviews and reduces the burden on participants when they are seen by the staff; participants can obtain information that they need, such as dates of hospitalization and contact information for physicians and hospitals; the instructions that are written into the form are standardized; and the privacy of the situation encourages more honest responses to uncomfortable questions. A methodological study in the Netherlands demonstrated greater reporting of disability in a self-administered questionnaire compared to a face-to-face interview [17]. On the other hand, interviews have advantages that should be considered when deciding the mode of administration for a study [16]. An interview is easier for the participant in many ways and may avoid difficulty with vision problems or low literacy. An interview may be more enjoyable, ensuring the completion of the full protocol. An interviewer can clarify questions to some degree, according to the guidelines set forth by the study. Interviewers can collect more complex data and can minimize missing and inappropriate responses. Finally, the respondents' appearance and behavior can be directly observed by the interviewer.

7.3.5 Proxy Report

Proxy respondents may be required for research on older populations if the study participant is too ill to respond or has dementia. If a study has a longitudinal component, then even if all participants can participate fully in the study at baseline, a certain percentage of participants will eventually need a proxy respondent to provide information for them. It is of critical importance to construct proxy interviews to maximize the validity of proxy responses related to functioning and disability, as these deficits are often the reason that an individual cannot participate and there can be substantial bias in a study that misclassifies these individuals. A full chapter in this book deals with issues related to the use of proxies (Chap. 6). These issues are vital for obtaining proxy assessments of functional status.

7.4 Summarizing Multiple Disability Items

Disability is most often assessed using multiple self-report items in a battery or formal scale. Often these items are all scored in the same manner, such as yes or no to questions about the ability to perform without the help of another person or equipment, or selection from a standard set of ordinal responses regarding level of difficulty. Occasionally, a battery may have different response frameworks for different items. Regardless of what types and patterns of response categories are utilized for a set of items, it is usually necessary to aggregate responses to multiple items of functioning and disability in order to represent the true nature of functional status in the individual. This may not be ideal in clinical care, where it is important to understand the specific functions in which someone is having difficulty or needs help, but is important in research settings to be able to summarize an individual's functional deficits for analytic purposes. Common approaches to summarizing multiple items that assess disability include the following:

- Difficulty/inability in one or more items
- Summated scale
- Hierarchical scale
- Computer-adaptive testing.

When utilizing a large number of items, it is necessary to understand whether the items represent a single underlying concept or multiple concepts, which items should be aggregated and which should be deleted. A variety of statistical techniques (e.g., factor analysis, Rasch modeling) are used to develop scales that represent underlying or latent constructs that validly represent a domain of functioning (for an example of this process, see [8]).

7.4.1 Disability in One or More Items

The most common summarizing technique for frequently used batteries such as ADLs and IADLs is to classify someone as disabled in the domain if they have difficulty or are unable to

perform one or more items in that domain. A great deal of publicly available data are produced in this way, showing the percentage of the population with disability in one or more ADL items or one or more IADL items, stratified by various demographic variables. For certain purposes, a cutpoint that requires a specific number of ADL items may be set (e.g., eligibility for long-term care that requires inability or the need for help in performing three or more ADLs).

7.4.2 Summated Scale

Summated scales are more complex than the simple dichotomy of disability vs. no disability. They can be constructed in a variety of ways. The simplest is to count up the number of items in which disability is present. In the presentation of public data, these counts are often aggregated into a few categories, such as 1 ADL, 2 or 3 ADLs, and >3 ADLs (Fig. 7.2). More complex summated scales may add up the difficulty score for a number of items to get a continuous summated score of disability. This has been the case in a number of large clinical trials with functional outcomes. An important trial of both aerobic and resistance exercise in individuals who have osteoarthritis of the knees utilized a comprehensive summated scale of lower extremity functioning [18]. This scale contained 23 items on ambulation and stair climbing, transfer, upper extremity functioning, and basic and complex ADLs. Possible responses to each item were no difficulty, a little difficulty, some difficulty, a lot of difficulty, and unable. Each was scored from 0 to 5, and the composite score was an average of the items. The trial demonstrated a significant difference in the two exercise groups compared to an education control group. A continuous variable such as this offers more statistical power to show benefit in a clinical trial, but a limitation is difficulty in understanding what a change of a certain number of points means in terms of real improvement in the disability that result from this intervention. The section on change in function (Sect. 7.6) will describe how clinically meaningful change can be assessed for summated scales and continuous variables.

Another example of a summated scale that worked well in a clinical trial is from a study that was done to evaluate the benefits of a home physical therapy program in frail older adults which focused on improving underlying impairments [19]. The ADLs were scored as (0) having no difficulty and needing no help in the past month, (1) having difficulty but not needing help and (2) needing help. The ADLs were walking, bathing, upper- and lower-body dressing, transferring from a chair to a standing position, using the toilet, eating, and grooming. The scores on these eight items were then summed. The intervention group had an increase in disability from 2.3 at baseline to 2.7 at 12 months, but the control group had a significantly greater increase in their disability scores, from an average of 2.8 at baseline to 4.2 at follow-up.

7.4.3 Hierarchical Scale

Many self-report items have a hierarchical response pattern that is very useful in scoring the items. In the example above, ADLs were scored according to a hierarchy that went from “no difficulty, no help required” to “difficulty but not needing help” to “needing help”. It is important to recognize that the scoring of items in a hierarchical manner is distinct from a hierarchical scale, which has separate items that are themselves hierarchical. A simple example of the latter is a scale that queries the ability to walk across a room, walk 50 ft, and walk one-half mile. Knowing that the study participant can walk a half mile indicates that they are also able to do the two easier tasks.

A simple hierarchical scale that has been useful in epidemiologic studies includes ADLs and higher mobility, often defined as the ability to walk ¼ mile and climb a flight of stairs. This scale is hierarchical because nearly everyone who has difficulty performing basic ADLs also has difficulty with one or both of the higher mobility items. Figure 7.3 illustrates research that demonstrates the predictive validity of this hierarchical scale. It shows the joint contribution of serum albumin level and the three-item hierarchical

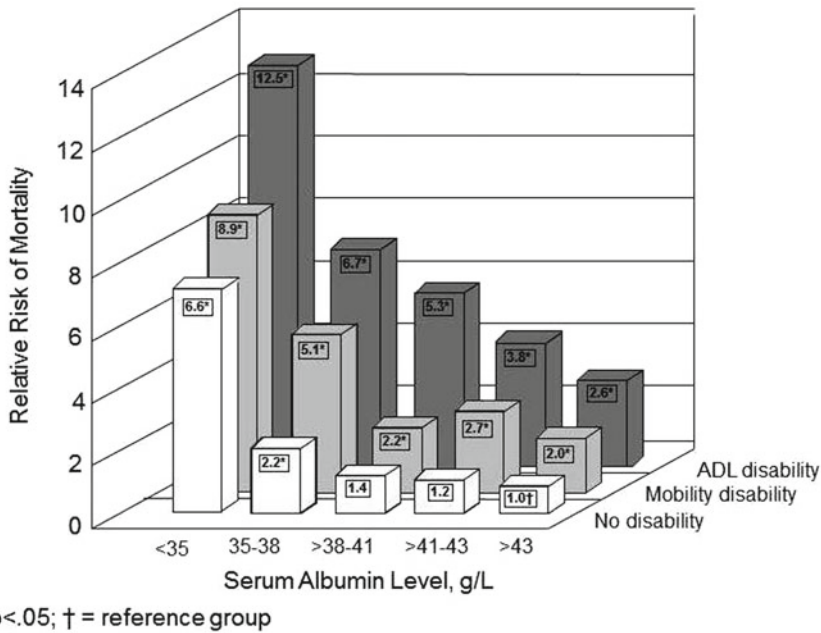


Fig. 7.3 Four year age-adjusted risk of death according to serum albumin in disability status, Women, EPESI 1988–1992 [20]. *Abbreviation:* ADL activity of daily living

disability scale in predicting mortality. At each level of serum albumin, there is a gradient of risk for mortality that goes from no disability to mobility disability to ADL disability [20]. In addition to being a predictor of adverse outcomes, this hierarchical scale has also been utilized as an outcome measure in studying predictors of disability [21].

7.4.4 Item Response Theory (IRT) and Computer Adaptive Testing (CAT)

An innovative approach to using a large number of items to efficiently assess the full range of a domain is using IRT to develop an item bank that is then utilized by CAT to score an individual [22]. IRT was originally developed for educational testing and assumes an underlying latent trait that is manifested by responses to a wide range of items, with the level of the latent trait related to the probability that a particular item will be answered in a specific way. This implies a

unidimensionality to the underlying trait and a range of items that cover the full spectrum of functioning represented by the trait. In practice, a large number of items that are hypothesized to be related to the trait under study are administered and an item bank of appropriate items is developed using IRT. Once this has been accomplished, CAT can be used to efficiently hone in on the level of functioning of an individual by using the knowledge gained from IRT to sequentially select relevant questions from the item bank until a pre-determined level of precision is obtained regarding where the individual stands on the scale of functioning. Using this approach, individuals on each end of the functional spectrum may be asked very different questions from the item bank, but their position on the functional spectrum can be determined to similar degrees of precision and with a minimal number of items utilized.

This approach is illustrated in Table 7.5, which shows the correlations between a CAT-based approach and criterion physical function scores that come from the full-item bank of 124 items. These correlations are compared with

Table 7.5 Intraclass correlation coefficients between CAT-based and random-based scores with IRT-criterion physical functioning scores after administering 5, 10 and 20 items, for three score ranges

| Items administered | Low-range (N=131) | Mid-range (N=521) | High-range (N=150) |
|--------------------|-------------------|-------------------|--------------------|
| CAT – 5 items | 0.87 | 0.88 | 0.83 |
| Random – 5 items | 0.64 | 0.77 | 0.53 |
| CAT – 10 items | 0.92 | 0.93 | 0.91 |
| Random – 10 items | 0.80 | 0.89 | 0.76 |
| CAT – 20 items | 0.97 | 0.97 | 0.95 |
| Random – 20 items | 0.92 | 0.96 | 0.93 |

Source: Adapted from Haley et al. [22]; CAT computer adaptive testing; IRT item response theory

correlations between randomly selected sets of questions from the data bank and the criterion physical function score. It is impressive that in the middle range of functioning, a CAT approach that used only five items has a correlation as high as 0.88 with the full bank of items, with correlations nearly as high at the high and low ends of function. This is in comparison with correlations of ≤ 0.77 if the items were selected randomly rather than chosen using the CAT process. Using ten items selected by CAT increases the correlations to above 0.90. It is thus clear that CAT is a very efficient method of determining where an individual is on the spectrum of functioning.

While this approach has many advantages, it should be kept in mind that it is very different than using a standard set of assessment items in an evaluation. Individuals at different ends of the functional spectrum may start out with the same question, but will then be presented with a very different set of questions. This can efficiently provide a precise measure of their level of functioning. However, in longitudinal studies or clinical trials, a participant who has a change in functioning may get completely different questions when returning for a follow-up visit. Theoretically, this is the way CAT is supposed to work, but the research community will have to become more familiar with this approach and accept the results of CAT. Researchers who design clinical trials will have to develop methods to assess the power of a randomized trial that uses this type of outcome.

7.5 Objective Physical Performance Measures

Objective measures of physical performance have received increasing attention as assessments that can measure functioning in a standardized manner in both the research and clinical settings. These measures can be defined as assessment instruments 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 the counting of repetitions or the timing of the activity as appropriate. As per the Nagi model of the disability framework that was discussed earlier in the chapter, these measures can be utilized to represent impairments, functional limitations or actual disability. However, most are indicators of functional limitations and can be thought of as representing the building blocks of functioning [23]; discrete actions that can be readily tested and that are necessary components of the ability to function independently in the community.

The development of performance testing was in response to concerns that self-report of functioning and disability may not be accurate. Additionally, disability is meant to be a measure of how well people are functioning in their own environments and, while this is critical to know in understanding their abilities to function independently, it is highly related to the environmental challenges that are specific to their situation. For example, the ADL that assesses bathing may

have very different responses if the participant has a walk-in shower vs. a bathtub that is difficult to climb into. Standardized performance tests avoid this limitation. Self-reported measures of disability identify a small fraction of the population that is disabled, but the remainder of the population is not characterized according to their functional abilities in most self-report measures. Performance testing has the potential to expand the spectrum across which people can be characterized, and is therefore able to identify both low and high functioning. While there are clear benefits with performance testing, understanding an older adult's ability to function independently in his or her own environment is best described using disability assessment. Research has shown that rather than replace disability assessment, the use of performance measures appears to complement disability assessment in characterizing an individual, especially with regard to individuals who are not severely disabled.

7.5.1 Examples of Performance Tests

A wide range of performance tests have been developed for use in the older population.

Examples of performance tests that assess both upper and lower extremity function include:

- Pegboard test
- Picking up object
- Lifting 10 pounds
- Gait speed
- Timed up and go test
- Chair rise – single and repeated
- Stair climb.

These tests all evaluate a single task except for the timed up and go test, which combines chair rise and gait speed, asking the participant to rise from a chair, walk 10 ft and then return to the chair [24]. The tests listed above are commonly employed, but there are others that have been developed for specific studies or for use in specific populations. This list does not include tests of aerobic capacity, which also assess physical functioning and may be valuable in identifying problems in older adults, but are more targeted toward

endurance. The most commonly used of these tests include the 6-min walk and the 400 m walk.

In addition to individual items used for performance testing, several batteries of performance tests have been developed to give a broader picture of functional abilities. The most commonly used of these batteries are the Short Physical Performance Battery (SPPB) [25] and the Physical Performance Test (PPT) [26]. Items used in these tests are as follows:

SPPB

- Side-by-side, semi-tandem and tandem stands, each held for 10 s
- Four-meter walk at usual pace
- Single chair stand and if successful, five timed chair stands as quickly as possible.

PPT

- Writing a sentence
- Simulated eating
- Turning 360°
- Putting on and removing a jacket
- Lifting a book and putting it on a shelf
- Picking up a penny from the floor
- 50-foot walk test
- Climbing stairs (scored as two items).

The SPPB is more purely a lower extremity function test, while the PPT is multidimensional. Both batteries have been used in observational studies and as outcomes in randomized controlled trials. Each of the three components of the SPPB is scored categorically from 0 to 4, and a total SPPB score of 0–12 is created by summing the three components. The SPPB has been found to predict mortality, the need for nursing home admission, and health care utilization in the overall older population. Furthermore, in a population that had no disability at the time the performance battery was administered, the score was found to be highly predictive of those who developed ADL and mobility disability 1 and 4 years later (Fig. 7.4). These findings have been replicated in other populations and with other, similar performance measures, and indicate that there is a state of preclinical disability—expressed as impairments and functional limitations—that indicates a high risk of proceeding to full-blown disability. This finding also provides a way of identifying

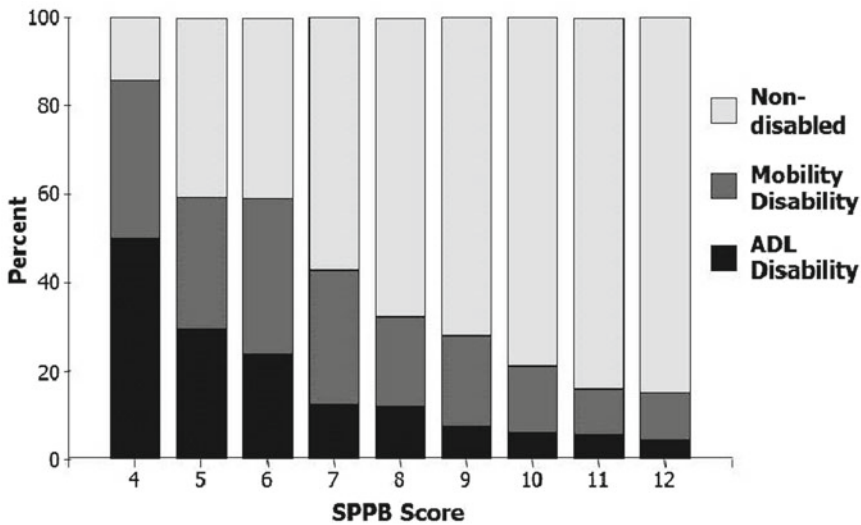


Fig. 7.4 Disability status at 4 years among individuals who were not disabled at baseline, according to baseline Short Physical Performance Battery (SPPB) score [21]

high-risk older adults for whom preventive interventions may be highly effective.

7.5.2 Overview of Uses

Physical performance measures have the following potential applications:

- Identifying high levels of functioning
- Identifying non-disabled individuals at risk of disability
- Clinical “vital sign”
- Evaluating change in functioning and health status
- Intervention studies
- Cross-national and cross-cultural studies.

The ability of performance tests to describe the full spectrum of functioning, including the high end of function, has already been described. This is not an inherent quality of performance measures, and only certain measures will actually identify the highest level of functioning. For example, the balance tests in the SPPB were designed to be done safely in a population starting at age 70 and with no upper age limit. Younger, healthy individuals will generally be able to hold all three stands for 10 s, and the highest performers

cannot be identified. This ceiling effect would be useful if identifying and characterizing very high performers were an important part of a research project. For this reason, in some instances investigators have added more difficult balance tests, including walking on a narrow course, a single leg stand with eyes open for up to 30 s and, if the participant is successful at that, a single leg stand for up to 30 s with eyes closed, a task that few people can fully accomplish [13, 27].

The identification of non-disabled individuals at risk of disability is related to the ability of performance measures to tap into the higher end of the functional spectrum, but it is not necessary to identify the very highest level of functioning in order to accomplish this. For example, the SPPB is scored from 0 to 12, with 12 being the best functioning. While additional testing might be able to stratify individuals with scores of 12 into a further hierarchy of very high functioning, the 0–12 range works very well in generally classifying the full older population. Individuals with scores >8 generally self-report no disability, but study of those with scores of 9–12 reveals a clear graded risk of multiple adverse outcomes (e.g., disability, mortality, nursing home admission, hospital use), indicating that in these non-disabled

individuals it is possible to use a performance test to characterize level of risk.

Even in individuals who have lower SPPB scores, there will be a subset that reports no disability, but these individuals will have a higher risk of adverse outcomes than those who report no disability and have a score >8 . This is illustrated in Fig. 7.4, which is restricted to non-disabled individuals at baseline (when the SPPB was measured) and demonstrates the graded relationship between SPPB scores and the future risk of ADL and mobility disability. This kind of relationship between performance in non-disabled individuals and adverse outcomes has been examined in many studies that extensively adjust for multiple measures of demographics, health status, behavioral and other risk factors, with virtually no weakening of the performance–adverse outcome relationship. Cooper et al. [28] reviewed the associations of multiple physical performance measures and a variety of adverse outcomes. In the Health, Aging and Body Composition (Health ABC) study, which included only non-disabled individuals who have no difficulty walking a quarter mile or climbing stairs, individuals who had a gait speed slower than 1.0 m/s had greater risk of future persistent mobility limitation, death and hospitalization than did individuals who had a gait speed of 1.0 m/s or faster [29]. What is not yet clear is what it is about the performance tests that make them independent predictors of aging outcomes. It is likely that they are capturing diverse aspects of health in older adults—including comorbidity severity, physiologic decline and motivation—that are not represented by the usual kinds of potential confounders that are studied.

Objective performance testing is beginning to be applied in the clinical setting, though research in this area is sparse. Busy clinicians rarely observe the functional abilities of their older patients but could gain much by knowing whether the patient had functional declines over time, which may occur even while their patient’s disease status apparently remains stable. Studenski et al. [30] integrated the SPPB into two geriatric outpatient settings and found that after training the staff, the battery could be performed efficiently

and was well accepted by those who administered the tests and by the patients. It showed predictive features similar to what was found in epidemiologic studies of representative community-dwelling populations. In a study in Italy, the SPPB was performed within 24 h of hospitalization and before discharge in geriatric patients who were hospitalized for several different diagnoses [31]. It was demonstrated that the testing was feasible and safe, and that the SPPB score at admission was a significant predictor of length of stay, even after adjustment for a comprehensive measure of comorbidity. Furthermore, a poor SPPB score in the test done just prior to discharge predicted greater rehospitalization, mortality and decline in ADLs over the year subsequent to discharge [32]. There is thus evidence for the potential advantages of utilizing physical performance testing in both the outpatient and hospital settings, but further work will be necessary to examine just how this might influence clinical decision-making and, ultimately, patient outcomes.

Observing change in function and disability are an important part of epidemiologic studies on aging. Transitions in states of self-reported disability are a critical part of this research, but performance measures offer a way of measuring change in a setting that is standardized rather than within the context of an environment that itself may be changing and affecting a disability outcome. Having measures that are precise, reliable and sensitive to clinically meaningful change is of particular importance in clinical trials, and trials that use performance tests as primary and secondary outcomes are becoming increasingly common. Figure 7.6a, b show results from randomized controlled trials that used the SPPB and PPT as primary outcomes.

One limitation of physical performance measures in longitudinal studies is that, over time, individuals with the most disability are less likely to return to the clinic for evaluations. Obtaining high follow-up rates often requires the use of home visits, telephone contacts and proxy interviews [33]. Certain performance measures, such as the SPPB and usual gait speed, have been successfully performed in the home setting. However if only telephone contacts can be used, then it

may be necessary to follow-up individuals using self-report or proxy.

Objective performance measures also provide a means of comparing functional status across countries or cultures, where disability measures may lose comparability due to environmental differences or differential access to assistive devices. Figure 7.5 shows results from the English Longitudinal Study on Aging, a nationally representative sample of older adults in England. It demonstrates the prevalence—by age and sex—of poor physical performance, documented as an SPPB score of ≤ 8 and gait speed of < 0.5 m/s. Longitudinal studies have shown performance below these cutpoints to be strongly associated with multiple adverse outcomes. Poor performance affects only about 10% of individuals in their 60s, but the prevalence rises rapidly in the 70s and attains very high levels in those ≥ 80 years of age. Women have higher rates of poor performance than do men at all ages. When a US national study that is doing these same tests (the National Health and Aging Trends Study [4]) is completed, results across the US and England will be readily comparable, as will results from other countries that are using these tests.

It is not always practical to utilize performance measures of functioning in large surveys, but methodological work with performance measures may help in the interpretation of differences in self-report that are observed in the self-report of functioning and disability. This approach, first developed by the WHO, uses performance measures of functioning to calibrate responses to self-report items. Using the modeling technique Hierarchical Probit Modelling, Iburg et al. [36] used performance tests from the National Health and Nutrition Examination Survey (NHANES) III to create a vector of performance that represented a latent variable which indicated the true underlying level of performance. They then looked at how different subsets of the population reported disability at different levels of this background latent variable. Utilizing this approach to assess differences in self-report across countries, Melzer et al. [37] compared disability self-reports from the US with the Longitudinal Aging Study Amsterdam. A lower prevalence of dis-

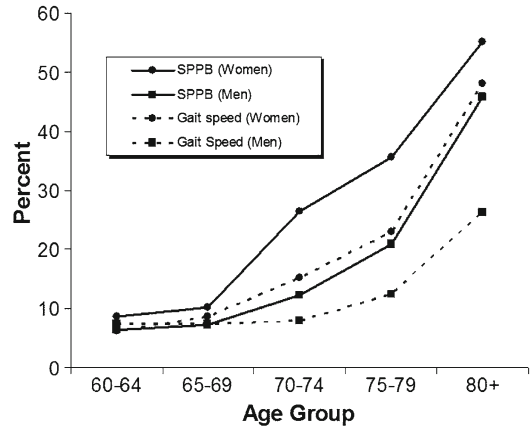


Fig. 7.5 Percentages of men and women with a SPPB ≤ 8 and gait speed ≤ 0.5 m/s [39]. *Abbreviation: SPPB* Short Physical Performance Battery

ability was observed in the Netherlands, but this approach revealed that individuals in the Netherlands did not report disability until they had more reduced levels of background performance. This indicated that differential reporting of disability across countries might be at least partially explained by cultural differences in how individuals perceive and report their disability level. This calibration technique can thus be used to adjust disability rates to make them comparable across countries, or across cultural or ethnic groups.

7.5.3 Psychometric Properties

Psychometric properties of many physical performance tests have been studied and they have generally been found to have excellent validity and reliability. Predictive validity has been extensively examined by demonstrating the high predictive power of these tests for important adverse outcomes in the older population, as previously described in this chapter. Test-retest reliability is very high when two examiners observe the same test, and intra-observer reliability—when tests are done 1 or 2 weeks apart—has also been found to be high. For example, in the Women’s Health and Aging Study, a subset of participants was seen at home weekly and administered the SPPB

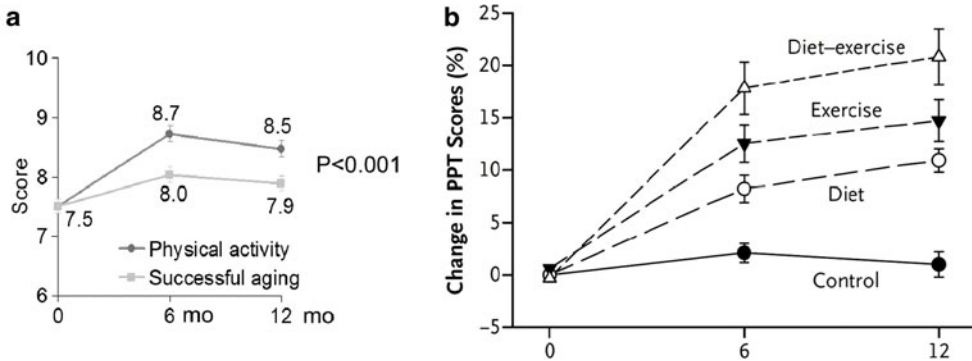


Fig. 7.6 Results from randomized controlled trials that used the SPPB and PPT as primary outcomes. (a) LIFE-P SPPB scores at baseline, 6 and 12 months. Means estimated from repeated measures of ANCOVA adjusted for gender, field center and baseline values [34]. *Abbreviations:*

ANCOVA analysis of covariance; *LIFE-P* Lifestyle Interventions and Independence for Elders Pilot; *SPPB* short physical performance battery. (b). Mean percentage changes in Physical Performance Test (PPT) during 1 year intervention [35]

[38]. The test-retest reliability was examined for 1-week intervals at the beginning, middle and end of the 24-week substudy. The total SPPB score had the highest intraclass correlations for these three 1-week intervals, ranging from 0.88 to 0.92. The walk and multiple chair rise tests were nearly as high, ranging from 0.80 to 0.89 and 0.76 to 0.90, respectively. The reliability of the balance tests was somewhat lower, ranging from 0.70 to 0.82.

In addition to validity and reliability, it is critical to examine sensitivity to change in any measurement that is used for longitudinal studies or clinical trials. Figures 7.6a, b indicate that the SPPB and PPT respond to behavioral interventions. In the Women's Health and Aging Study, in which women were followed-up every 6 months, it was found that the SPPB was very responsive to clinical events that occurred sometime within the 6-month interval between which SPPB scores were measured [38]. This is illustrated in Figure 7.7, which shows the change in SPPB in women who were admitted to the hospital with specific diseases. The largest decline was for hip fracture, followed by stroke and myocardial infarction. These declines were much larger than the decline observed in participants who had none of these events over 6 months.

7.6 Change in Function at the Individual and Population Levels

A key aspect of physical function and disability measures in older adults is to represent the change that occurs—both decline and improvement—in response to aging, chronic disease, acute disease events, psychosocial and behavioral factors, and therapeutic interventions. All of the assessment strategies discussed to this point have been used to assess change, and formal evaluations of sensitivity to change have been made for some of them. It was previously described how interpreting change in a summated scale (e.g., the self-reported disability scale used for the FAST trial) can be difficult. A more obvious kind of change that is easy to understand is the onset of incident disability (e.g., ADL, mobility disability) in individuals who were previously not disabled. This approach is useful in prospective epidemiologic studies, but investigation of change across the total population may also require the use of function and disability scales. The interpretation of change is particularly challenging in characterizing physical performance measures, which, due to their excellent psychometric properties, are ideal for capturing change with high precision,

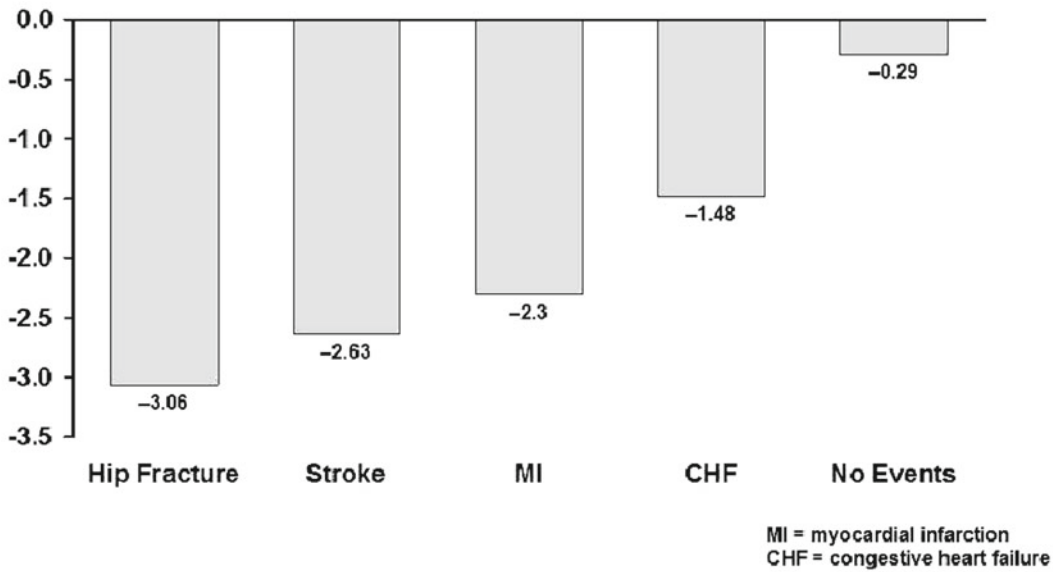


Fig. 7.7 Change in summary performance score over interval during which participants were hospitalized for specific events [38]. Abbreviations: *CHF* congestive heart failure; *MI* myocardial infarction

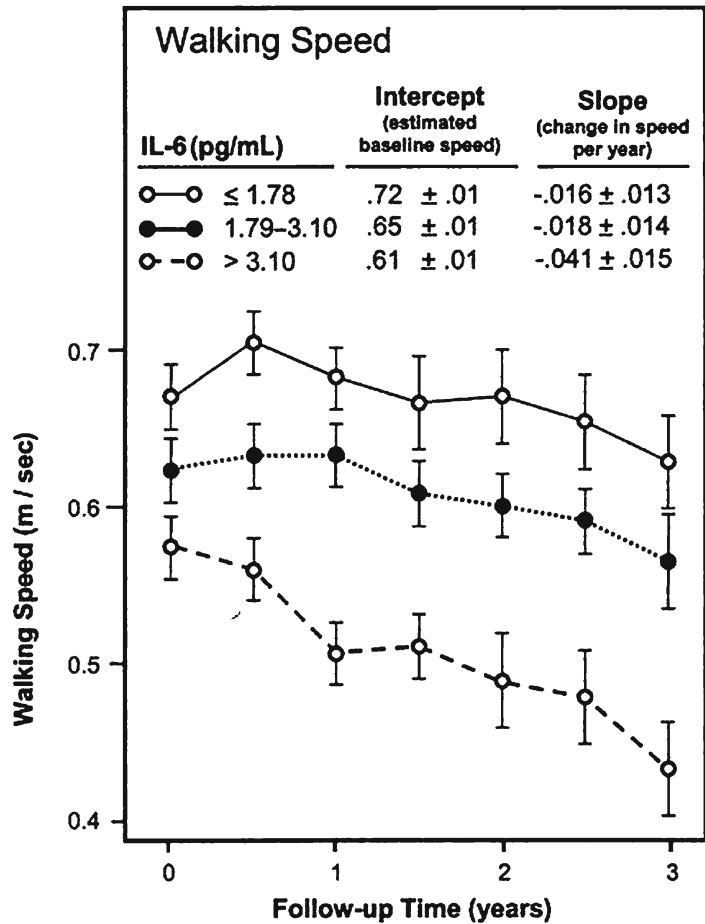
but for which it is not inherently obvious what the change means.

An important contribution in this area was the work on clinically meaningful change done by Perera et al. [40]. They used both of the techniques that are commonly used to define change, distribution-based and anchor-based analyses, and found similar results for both approaches. Using data from observational studies and a clinical trial, they evaluated both usual gait speed and the SPPB and found generally consistent results for both distribution-based and anchor-based analyses that determine small meaningful and substantial change. Small meaningful change was found to be 0.05 m/s for gait speed and 0.5 points for SPPB score. Substantial change was 0.1 m/s for gait speed and 1 point for the SPPB. In the example shown in Fig. 7.5a (Lifestyle Interventions and Independence for Elders Pilot [LIFE-P] study), the physical activity intervention group had a 1-point improvement in the SPPB score from baseline to 12 months, indicating substantial change, while the successful aging health education group had improvement that could be classified as small meaningful change [34].

7.6.1 Trajectories of Functioning

Many longitudinal studies measure functioning and disability repeatedly over time, and these studies provide data for examining trajectories of functioning. In particular, it is important to understand how various risk factors, both those measured at baseline and those that may change over time, affect trajectory of functioning. An excellent example of research aimed at examining trajectories of gait speed and how they relate to the inflammatory cytokine interleukin (IL)-6 is shown in Fig. 7.8 [41]. These data are from the Women's Health and Aging Study, and the population is stratified by tertiles of IL-6. There is a graded relationship between IL-6 tertile and baseline gait speed, with slower gait speed related to higher IL-6. Women in all strata of IL-6 had a decline in gait speed over a 3-year follow-up but there is a significantly steeper decline in those who had higher IL-6. This difference in trajectories was further examined to evaluate whether muscle strength was a mediator of the IL-6–gait speed association. In models that added muscle strength data over time, there was a substantial attenuation of the IL-6–gait speed relationship,

Fig. 7.8 Change over time in walking speed according to interleukin (IL)-6 tertiles [41]



giving evidence that IL-6 may affect gait speed through its impact on loss of muscle strength.

Trajectory of change in measures of functioning may be more powerful predictors of subsequent outcomes than are baseline measures alone. For example, in the Women’s Health and Aging Study II, the trajectory of decline of handgrip and hip flexion strength in up to six subsequent assessments was a significant predictor of subsequent mortality in a sample of non-disabled women who were 70–79 years of age at baseline, even after adjusting for baseline values of strength [42]. Trajectories of functioning can also be quite valuable as outcome measures in studies of predictors of functional status as they may be more accurate measures of change than a single outcome assessment. In studying patients with

peripheral artery diseases, McDermott et al. [43] evaluated changes in a number of functional outcomes and found that study participants who were more sedentary at baseline (operationalized by asking about number of hours spent sitting per day) had a steeper decline (over up to 4 years) in usual gait speed, fast gait speed and distance covered in a 6-min walk test.

7.6.2 Progression: Catastrophic and Progressive Disability

Longitudinal data provide a means of examining differences in trajectories of functioning or disability in those who develop substantial disability. This is illustrated in the concepts of

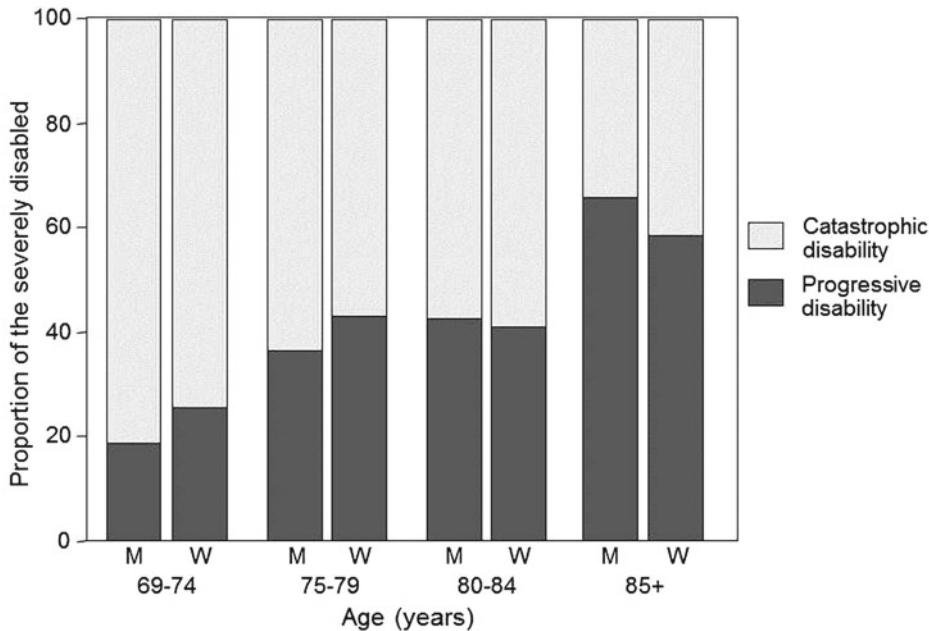


Fig. 7.9 Proportion of individuals with catastrophic and progressive disability among those who developed severe disability during the follow-up period, by age group and gender [44]. Abbreviations: *M* men; *W* women

catastrophic and progressive disability. These concepts were developed by examining trajectories of disability in individuals who eventually developed severe ADL disability, defined as the need for help or inability to perform three or more of six ADLs [44]. Using longitudinal data collected annually prior to the development of severe disability, catastrophic severe disability was defined as having no ADL disability in the 2 years prior to the onset of severe disability, while progressive severe disability was defined as having disability in one or two ADLs in the year prior to the onset of severe ADL disability. Among individuals who developed severe disability, the distribution of catastrophic and progressive severe disability by age group and sex is illustrated in Fig. 7.9. Catastrophic disability is the main type of disability in men and women 65–74 years of age, but this pattern is reversed in the ≥ 85 -years-of-age group, where 60% of severe disability is progressive. The prevention of disability is quite different for these two patterns of disability onset. The main way of reducing catastrophic disability is the prevention of diseases that lead to catastrophic

disability (e.g., stroke, hip fracture), whereas an important means of preventing progressive disability is the reduction in the impairments and functional limitations that accompany many chronic diseases of aging and lead to steady functional decline, the so-called geriatric model of functional decline.

7.6.3 Recovery, Interval of Recall

As experience with longitudinal studies of aging was acquired, it became evident that the transition to disability is not an absorbing state and that recovery from disability is common. This was first demonstrated in prospective studies that had intervals of at least 1 year between assessments, so the true dynamics of recovery and then potential return to disability in the short term was unclear. Recovery has best been characterized in the Precipitating Events Project (PEP) [45], a cohort study of 754 community-dwelling adults ≥ 70 years of age who were not disabled in ADLs at baseline and were followed-up with monthly

telephone interviews over several years. Individuals who were termed frail (determined by slow gait speed) were oversampled. During a follow-up period of >4 years, 56% reported disability (the need for help or instability in one or more of four ADLs: bathing, dressing, walking, transferring). Overall, 81% of these individuals recovered, becoming non-disabled in all four ADLs over the subsequent year, and among those who recovered, a full 57% remained non-disabled for at least 6 months. Recovery was less likely in individuals who had cognitive impairment, physical frailty or severe disability (three or four ADLs at the onset of their disability).

The PEP study also enabled the exploration of recall of disability in individuals who were not currently reporting disability, but who had a month or more of disability reporting in the previous year [46]. Only about half of individuals who had reported in the monthly phone assessments that they needed help with one or more of the four ADLs recalled that they had this disability when queried at the end of the year following the year in which they had recovered. Since even a short disability episode has been found to put older adults at higher risk of long-term adverse outcomes compared to those who never had disability, it is important to note that half of disability episodes are not recalled when participants are interviewed on an annual basis. Accordingly, with longer time intervals between assessments, there will be greater under-ascertainment of incident disability cases.

7.6.4 Joint Effect of Disability Transitions and Death: Active Life Expectancy

There is interplay among age at transition to disability, duration of disability, and length of life that determines the number of years that older adults live in the disability-free state (termed active life expectancy) and the number of years spent in the disabled state. Life table analyses have been used to partition total life expectancy from a specific age into active and disabled life expectancy, utilizing data from population-based

longitudinal studies on transitions from the nondisabled state to disability and death, and from the disabled state to nondisability and death. This approach is important for understanding how factors that affect both death and disability have their impact on active and disabled life expectancy. The approach is also critical for determining whether population increases in total life expectancy that are occurring throughout the world are due to the prolongation of years spent in the disabled state, or to increases in active life expectancy, years free of disability and with higher quality of life. The reduction of years spent with disability has been termed compression of morbidity, and it represents a decrease in disabled life expectancy that results from compressing chronic disease and disability into a smaller number of years between disease and/or disability onset and mortality.

There has been little nationally representative longitudinal disability data collected using identical instruments over time to allow for the examination of trends in active and disabled life expectancy. One example of where this has been possible is in the Current Beneficiary Survey, a multistage, longitudinal survey of the US Medicare population that was sponsored by the US Centers for Medicare and Medicaid Services. Figure 7.10 shows data from samples whose evaluations started in 1992 and 2002, and went on for up to 4 years [47]. All 1-year transitions for which data were available contributed to the estimates. Women had a modest increase in total life expectancy of 0.3 years, which resulted from an increase in active life expectancy of 0.7 years and a decline in disabled life expectancy of 0.4 years. Disability in this case was defined as having difficulty in any one of six IADLs and six ADLs. The analyses partitioned disabled life expectancy into three categories, years lived with IADL disability only, moderate ADL disability and severe ADL disability (difficulty with three or more ADLs). In women, 0.3 of the 0.4 years of decline in disabled life expectancy resulted from a decline in severe ADL disability years. Men had a substantial gain in total life expectancy of 0.8 years, which was due entirely to increases in active life expectancy. Overall disabled life

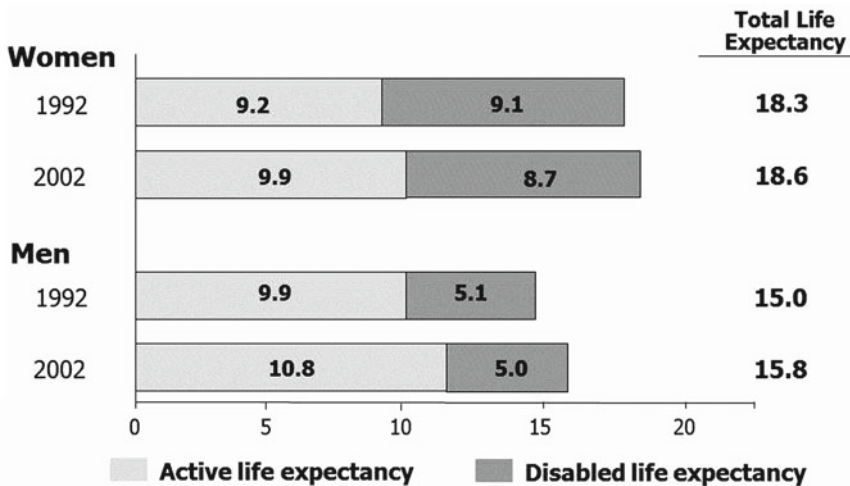


Fig. 7.10 Total active and disabled life expectancy, 1992 and 2002, from the Medicare Current Beneficiary Study [47]

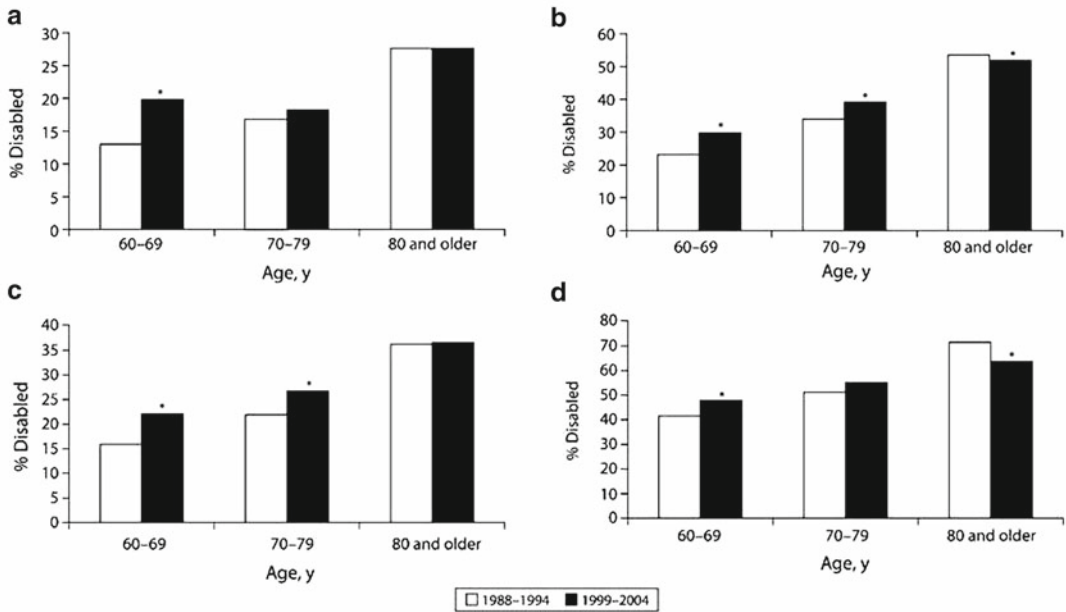
expectancy was virtually unchanged, with no change in life expectancy with IADL disability, a 0.1 year increase in moderate ADL disability and a 0.2 year decrease in severe ADL disability. On the whole, this presents an optimistic picture of improvements over time in the older population, with substantial increases in active life expectancy and moderate to slight decreases in disabled life expectancy.

7.6.5 Trends in Disability

Because disability status is a good way of representing overall health status in older adults who have complex patterns of disease, and because disability also has direct implications for the long-term care needs of an older adult, there has been much interest in evaluating disability trends over time. Although a number of cross-sectional national surveys now assess disability, uniform disability assessment done over time has been available only since the mid-1980s in just a few studies that have nationally representative samples. Although these studies use different assessment instruments, a convincing decline in age- and gender-specific rates of disability was observed from the mid-1980s through the 1990s [48]. The National Long Term Care Survey has similar

assessments of ADL and IADL disability available from 1982 through 2005, and recent findings indicate that the decline in disability observed for the first 12 years of the study continued and actually accelerated from 1994 through 2005. In another study that utilized reports of functional limitations—including lifting and carrying 10 lbs, climbing stairs, walking ¼ mile and seeing words in a newspaper—changes in prevalence were evaluated between 1984 and 1993. Declines were seen in the inability to perform all four of these tasks in the ≥ 65 -years-of-age population and in the ≥ 80 -years-of-age population. The functional limitations evaluated in this study, which assess more basic tasks than disability, are an excellent way to follow trends over time because they are less influenced by changing roles that can affect disability assessment (more men cooking in more recent surveys and more women managing money).

The observed declines in disability and functional limitations through the 1980s and 1990s were attributed to a number of factors. Educational status is strongly associated with disability, and it has been estimated that in more recent cohorts, from 25 to 75% of the observed functional declines are related to a higher educational level in more recent cohorts. Improving educational level and socioeconomic status in general likely



* $P \leq 0.05$ for difference between NHANES cohorts.

Fig. 7.11 Time trends by age group between National Health and Nutrition Examination Surveys (NHANES) in 1988–1994 and 1999–2004 [49]. (a) Prevalence of basic activities of daily living disability. (b) Prevalence of

instrumental activities of daily living disability. (c) Prevalence of mobility disability (d) Prevalence of functional limitations

have their impact on disability improvement through changes in behavioral risk factors, reductions in the prevalence of several chronic diseases, utilization of medical care, improved cognition and other less well understood factors. Other proposed explanations for the decline in disability include changes in nutrition and public health at the time when these cohorts were young, improved health promotion and medical therapy in more recent cohorts, and better utilization of assistive devices and technology.

Several recent studies have offered evidence that these positive trends in disability prevalence decline may not be sustained in the future. In Fig. 7.11, data are shown from the NHANES study regarding prevalence rates of disability from the periods of 1988–1994 and 1999–2004 [49]. It appears that in the oldest subset (those ≥ 80 of age), the more recent cohort has identical rates for ADL and IADL disability and lower disability rates for mobility disability and functional limitations (e.g., difficulty stooping,

crouching or kneeling; lifting or carrying 10 pounds; and standing from an armless chair). This is perhaps a vestige of this cohort doing better since mid-life. However, among the two younger age groups (60–69 and 70–79 years of age), the more recent cohort had higher disability rates for all of the measures, with most of them being statistically significant increases. This poorer picture for the “young old” in this national data set—which reflects recent increases in obesity and diabetes, both of which have major impacts on disability—may be a warning sign that in the future, the disability and health status of the older population may not be as good as in the current population. This argues strongly for the careful assessment of trends in the older population, with the institution of better strategies—in both disease prevention and therapy, as well as in improved behavioral risk factors—to reduce functional decline and disability in this rapidly growing segment of the population.

7.7 Summary

The assessment of physical functioning and disability is a component of nearly all epidemiological studies of older adults. There is a wide variety of measures to choose from and many decisions must be made when determining which assessments best fit a study that is being developed. Many aspects that are vital in making these choices have been discussed here. It is typically necessary to customize assessments for a particular study. Even the various large national studies that are supported by the US government often use different approaches to measuring disability. However, there are also advantages to having consistency in the way disability is measured. Having a small core of items that are asked in the same way on all studies would enable direct comparison of the populations being studied. A simple performance test, such as a gait speed test done in a standardized way, would also be an excellent assessment for comparing across studies. It would provide a simple measure of capacity that is free of the environmental challenges that differ across subsets of the population and have a strong impact on disability. Overall, there is evidence that self-report and performance measures complement each other and that both will add useful information in a comprehensive assessment.

References

- Nagi SZ (1965) Some conceptual issues in disability and rehabilitation. In: Sussman M (ed) *Sociology and rehabilitation*. American Sociological Association, Washington, DC
- Verbrugge LM, Jette AM (1994) The disablement process. *Soc Sci Med* 38:1–14
- World Health Organization (2001) *International classification of functioning, disability and health ICF*. World Health Organization, Geneva
- Freedman VA (2009) Adopting the ICF Language for studying late-life disability: a field of dreams? *J Gerontol Med Sci* 64:M1172–M1174
- Federal Interagency Forum on Aging Related Statistics (2004) *Older Americans 2004: key indicators of well-being*, p 28. Federal Interagency Forum on Aging Related Statistics Web site. http://www.agingstats.gov/agingstatsdotnet/Main_Site/Data/2004_Documents/entire_report.pdf. Accessed 27 Dec 2011
- Jette AM, Haley SM, Coster WJ et al (2002) Late life function and disability instrument: I. Development and evaluation of the disability component. *J Gerontol Med Sci* 57:M209–M216
- Ware J, Sherbourne C (1992) The MOS 36-item Short Form Health Survey (SF-36). *Med Care* 30(6):473–483
- Haley SM, Jette AM, Coster WJ et al (2002) Late life function and disability instrument: II. Development and evaluation of the function component. *J Gerontol Med Sci* 57:M217–M222
- Studenski S, Perera S, Patel K et al (2011) Gait speed and survival in older adults. *JAMA* 305:50–58
- Katz S, Ford AB, Moskowitz RW et al (1963) Studies of illness in the aged: the index of ADL, a standardized measure of biological and psychosocial function. *JAMA* 185:914–919
- Jette AM (1994) How measurement techniques influence estimates of disability in older populations. *Soc Sci Med* 38:937–942
- Fried LP, Bandeen-roche K, Chaves PH et al (2000) Preclinical mobility disability predicts incident mobility disability in older women. *J Gerontol Med Sci* 55:M43–M52
- Simonsick EM, Newman AB, Nevitt MC et al (2001) Measuring higher level physical function in well-functioning older adults: expanding familiar approaches in the Health ABC study. *J Gerontol Med Sci* 56:M644–M649
- Marsh AP, Ip EH, Barnard RT et al (2011) Using video animation to assess mobility in older adults. *J Gerontol Med Sci* 66:217–227
- Wunderlich GS (2009) Improving the measurement of late-life disability in population surveys: beyond ADLs and IADLs, summary of a workshop. Rapporteur. Committee on National Statistics and Committee on Population. Division of Behavioral and Social Sciences and Education. The National Academies Press, Washington, DC. (https://download.nap.edu/catalog.php?record_id=12740)
- Hulley SB, Cummings SR (1988) *Designing clinical research*. Williams & Wilkins, Baltimore
- Picavet H, van den Bos GAM (1996) Comparing survey data on functional disability: the impact of some methodological differences. *J Epidemiol Community Health* 50:86–93
- Ettlinger WH, Burns R, Messier SP et al (1997) A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis. The Fitness Arthritis and Seniors Trial (FAST). *JAMA* 277:25–31
- Gill TM, Baker DI, Gottschalk M et al (2002) A program to prevent functional decline in physically frail, elderly persons who live at home. *New Engl J Med* 347:1068–1074
- Corti M, Guralnik JM, Salive ME et al (1994) Serum albumin and physical disability as predictors of mortality in older persons. *JAMA* 272:1036–1042

21. Guralnik JM, Ferrucci L, Simonsick EM et al (1995) Lower extremity function in persons over the age of 70 years as a predictor of subsequent disability. *N Engl J Med* 332:556–561
22. Haley SM, Ni P, Hambleton RK et al (2006) Computer adaptive testing improved accuracy and precision of scores over random item selection in a physical functioning item bank. *J Clin Epidemiol* 59:1174–1182
23. Guralnik JM, Ferrucci L (2003) Assessing the building blocks of function: utilizing measures of functional limitation. *Am J Prev Med* 25:112–121
24. Podsiadlo D, Richardson S (1991) The timed “up and go”: a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 39:142–148
25. Guralnik JM, Ferrucci F, Pieper CF et al (2000) Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared to the Short Physical Performance Battery. *J Gerontol Med Sci* 55:M221–M231
26. Reuben DB, Siu AL (1990) An objective measure of physical function of elderly outpatients. The Physical Performance Test. *J Am Geriatr Soc* 38:1105–1112
27. Curb JD, Ceria-Ulep CD, Rodriguez BL et al (2006) Performance-based measures of physical function for high-function populations. *J Am Geriatr Soc* 54:737–742
28. Cooper R, Kuh D, Cooper C et al (2011) Objective measures of physical capability and subsequent health: a systematic review. *Age Ageing* 40:14–23
29. Cesari M, Kritchevsky SB, Penninx BW et al (2005) Prognostic value of usual gait speed in well-functioning older people – results from the Health, Aging and Body Composition Study. *J Am Geriatr Soc* 53:1675–1680
30. Studenski S, Perera S, Wallace D et al (2003) Physical performance measures in the clinical setting. *J Am Geriatr Soc* 51:314–322
31. Volpato S, Cavalieri M, Guerra G et al (2008) Performance-based functional assessment in older hospitalized patients: feasibility and clinical correlates. *J Gerontol Med Sci* 63:1393–1398
32. Volpato S, Cavalieri M, Sioulis F et al (2011) Predictive value of the Short Physical Performance Battery following hospitalization in older patients. *J Gerontol Med Sci* 66:89–96
33. Strotmeyer ES, Arnold AM, Boudreau RM et al (2010) Long-term retention of older adults in the Cardiovascular Health Study: implications for studies of the oldest old. *J Am Geriatr Soc* 58:696–701
34. LIFE Study Investigators, Pahor M, Blair SN et al (2006) Effects of a physical activity intervention on measures of physical performance: results of the Lifestyle Interventions and Independence for Elders Pilot (LIFE-P) Study. *J Gerontol Med Sci* 61:M1157–M1175
35. Villareal DT, Chode S, Parimi N et al (2011) Weight loss, exercise or both and physical function in obese older adults. *N Engl J Med* 364:1218–1229
36. Iburg K, Salomon J, Tandon A et al (2001) Cross-population comparability of self-reported and physician-assisted mobility levels: evidence from the Third National Health and Nutrition Examination Survey. Global Programme on Evidence for Health Policy discussion paper no. 14. World Health Organization, Geneva
37. Melzer D, Lan TY, Tom BD et al (2004) Variation in thresholds for reporting mobility disability between national population subgroups and studies. *J Gerontol Med Sci* 59:1295–1303
38. Ostir GV, Volpato S, Fried LP et al (2002) Reliability and sensitivity to change assessed for a summary measure of lower body function: Results from the Women’s Health and Aging Study. *J Clin Epidemiol* 55:916–921
39. Banks J, Breeze E, Lessof C et al (2011) English longitudinal study of ageing: retirement, health and relationships of the older population in England: THE 2004 ENGLISH LONGITUDINAL STUDY OF AGEING (Wave 2). Annex 6.1: Tables on measured physical performance Institute for Fiscal Studies Web site. <http://www.ifs.org.uk/elsa/report06/app6.pdf>. Accessed 27 Dec 2011
40. Perera S, Mody SH, Woodman RC et al (2006) Meaningful change and responsiveness in common physical performance measures in older adults. *J Am Geriatr Soc* 54:743–749
41. Ferrucci L, Penninx BW, Volpato S et al (2002) Change in muscle strength explains accelerated decline of physical function in older women with high interleukin-6 serum levels. *J Am Geriatr Soc* 50:1947–1954
42. Xue QL, Beamer BA, Chaves PH et al (2010) Heterogeneity in rate of decline in grip, hip, and knee strength and the risk of all-cause mortality: the Women’s Health and Aging Study II. *J Am Geriatr Soc* 58:2076–2084
43. McDermott MM, Liu K, Ferrucci L et al (2011) Greater sedentary hours and slower walking speed outside the home predict faster declines in functioning and adverse calf muscle changes in peripheral arterial disease. *J Am Coll Cardiol* 57:2356–2364
44. Ferrucci L, Guralnik JM, Simonsick E et al (1996) Progressive versus catastrophic disability: a longitudinal view of the disablement process. *J Gerontol Med Sci* 51:M123–M130
45. Hardy SE, Gill TM (2004) Recovery from disability among community-dwelling older persons. *JAMA* 291:1596–1602
46. Gill TM, Van Ness PH, Gahbauer EA (2009) Factors associated with accurate recall of prior disability in older persons. *J Am Geriatr Soc* 57:1897–1901
47. Cai L, Lubitz J (2007) Was there compression of disability for older Americans from 1992 to 2003? *Demography* 44:479–495
48. Freedman VA, Martin LG, Schoeni RF (2002) Recent trends in disability and functioning among older adults in the United States: a systematic review. *JAMA* 288:3137–3146
49. Seeman TE, Merkin SS, Crimmins EM et al (2010) Disability trends among older Americans: National Health and Nutrition Examination Surveys, 1988–1994 and 1999–2004. *Am J Public Health* 100:100–107