

Validation of New Measures of Disability and Functioning in the National Health and Aging Trends Study

Vicki A. Freedman,¹ Judith D. Kasper,² Jennifer C. Cornman,³ Emily M. Agree,⁴ Karen Bandeen-Roche,⁵ Vincent Mor,⁶ Brenda C. Spillman,⁷ Robert Wallace,⁸ and Douglas A. Wolf⁹

¹Institute for Social Research, University of Michigan, Ann Arbor.

²Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland.

³Jennifer C. Cornman Consulting, Granville, Ohio.

⁴Department of Population, Family, and Reproductive Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland.

⁵Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland.

⁶Department of Community Health, Brown University, Providence, Rhode Island.

⁷The Urban Institute, Washington, DC.

⁸Department of Epidemiology, University of Iowa, Iowa City.

⁹Maxwell School Center for Policy Research, Syracuse University, New York.

Address correspondence to Vicki A. Freedman, PhD, Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, MI 48106. Email: vfreedma@umich.edu

Background. Measurement gaps continue to hamper fuller understanding of late-life disability trends and dynamics. This article reports findings that validate the self-reported components of the disability protocol to be used in the new National Health and Aging Trends Study. The protocol was designed to redress existing measures by attending to environmental aspects of disability, capturing a broader range of capacity to perform tasks and including participation restriction items.

Methods. We undertook an in-person validation study to determine the reliability, validity, and initial measurement properties of the National Health and Aging Trends Study self-reported disability protocol ($n = 326$). A random subset ($n = 111$) was readministered the protocol within 2–4 weeks. The interview and reinterview included new self-reported measures of physical capacity, activity limitations, and participation restrictions, as well as established performance and cognitive tests. We calculated percent agreement and kappa between interviews for all self-reported items and summary measures. We also assessed the construct validity of summary measures through correlations with demographic characteristics, frailty, memory, and performance-based mobility and confirmed whether activity limitations and participation restrictions were distinct domains.

Results. New items and derived summary measures demonstrate robustness over a short time period, with kappas for retained/recommended items in the .60–.80 range. The summary measures correlate as expected with age, sex, residential status, and established performance-based constructs. Two factors, representing activity limitations and participation restrictions, were confirmed.

Conclusions. The National Health and Aging Trends Study protocol preserves the ability to examine more traditional measures of functioning while offering new insights into how activities are performed and preserving key conceptual distinctions.

Key Words: Disability—Functioning—Survey methods.

Received December 16, 2010; Accepted April 12, 2011

Decision Editor: Luigi Ferrucci, MD, PhD

MUCH has been learned over the past several decades about late-life disability—broadly defined to include deficits in capacity, limitations in activity, and restrictions in participation—at both the population and individual levels; yet, measurement gaps continue to hamper fuller understanding of these issues. Most studies of late-life disability continue to focus on limitations in activities measured by activities of daily living, instrumental activities of daily living, and functional limitations, all originating over 40 years ago outside the survey context (1–3). Such measures vary across studies in wording and in specific tasks queried but often ask individuals to report on levels of difficulty or

dependence with activities in the absence of supportive features of the environment, such as help from another person or use of assistive devices. Although the reliability and predictive validity of these measures have been established, questions of this type do not offer insights into how tasks are being carried out that may influence reports of difficulty and levels of dependence (4). Consequently, such items are valuable for tracking population trends but less so for sorting out what may be driving trends in independent functioning such as changes in the nature of activities (eg, electronic banking, shopping online), in the use of assistive devices (eg, especially for mobility- and bathing-related tasks), in

gender roles and division of labor in the household (eg, more husbands making meals or shopping), or in the prevalence of supportive residential care and retirement-oriented living environments. Similarly, the value of established measurement approaches for understanding the stages of disablement and whether there are identifiable signature trajectories in the adoption of accommodations—for example, first changing behavior, then adopting devices, finally relying on help—has been limited. Current items also focus mainly on deficits at the lower end of functioning (eg, difficulty picking up a small object) rather than ability to carry out somewhat more challenging tasks (eg, ability to open a sealed jar; see Simonsick and colleagues (5) for an exception). Moreover, measures of participation and restrictions in valued activities, although of growing interest internationally, have largely been absent from national U.S. studies of health and aging.

The National Health and Aging Trends Study (NHATS), funded by the National Institute on Aging as the successor to the National Long Term Care Survey (NLTC), is a new panel study of persons aged 65 years and older that will support analyses of both late-life disability trends and trajectories. A centerpiece of NHATS is a new disability measurement protocol, which includes both performance-based assessments and self-reports designed to measure key concepts in the disablement process including physical, sensory and cognitive capacity, accommodations, the ability to carry out essential activities independently, and participation and restrictions in valued activities (6). This article reports findings from a methods study designed to validate the self-report component. Specifically, we highlight the reliability and validity of self-reported measures of activity limitations that take into account environmental context, broader measures of physical capacity than typically available, new items on participation and restrictions in valued activities, and a set of derived summary measures. We also undertake an initial investigation of data structure to confirm whether participation restrictions form a distinct domain from activity limitations.

METHODS

We undertook an in-person validation study to determine the reliability, validity, and statistical properties of the NHATS self-reported disability protocol. An initial interview ($n = 326$) and identical reinterview of a randomly selected subsample ($n = 111$) were administered via computer-assisted in-person interviews.

Sample Selection

A purposive sample was recruited from four U.S. cities according to targets by broad age groups, self-care limitations and residential care status (see Table 1 for sample characteristics for both the initial and reinterview samples). The sample included 90 individuals with self-care limitations

Table 1. Sample Characteristics: NHATS Validation Study

Characteristic	Full Interview Sample ($n = 326$)	Reinterview Sample ($n = 111$)
	Percent (M)	Percent (M)
Age		
65–74*	42.0	44.1
75–84	39.3	36.9
85+	18.7	18.9
Gender		
Male	43.6	50.5
Female	56.4	49.5
Screened in based on needing assistance with personal care (bathing, dressing, getting around)	27.6	24.3
Lives in residential care setting	15.6	14.4
Frailty*		
Not frail (0–2 frailty criteria met)	73.9	73.0
Frail (3–5 frailty criteria met)	26.1	27.0
Word recall*		
10–17	35.0	35.1
7–9	31.6	31.5
0–6	33.4	33.3
Mean word recall	(7.9)	(8.1)
Mobility score* [†]		
10–12 (best performance)	31.3	30.6
6–9	44.5	41.4
0 [‡] –5 (worst performance)	24.2	27.9
Mean mobility score	(7.5)	(7.3)

Notes: *Measured at initial interview.

[†]Measured with the short physical performance battery.

[‡]0 = unable to perform any of the tests.

(needing help with eating, bathing, dressing, or getting around inside the home) and 51 persons living in residential care facilities. Proxy respondents were included in the initial interview ($n = 11$) but not eligible for reinterview.

We set a target of 100 reinterviews in order to assure reasonable precision for estimating reliability (eg, confidence intervals of width ± 0.1 for $\kappa = 0.9$ and ± 0.2 for $\kappa = 0.6$, for a relatively rare outcome of 0.20). Of the 111 reinterviews conducted, 70% were administered within 2 weeks of the first interview (92% within 3 weeks; all within 4 weeks). We examined reliability using both the full reinterview sample and only those within 2 weeks. Because results did not vary, we report results for the entire reinterview sample.

Measures

In devising NHATS measures, our aim was to address several gaps in currently available items. First, traditional activities of daily living measures along the lines of “Without help or special equipment, how much difficulty do you have . . .?” do not allow analysts to track independent functioning or the contribution of assistive technologies to promoting independence. Second, interpretation of instrumental activity of daily living changes over time, using traditional measures, has proven problematic as changes have occurred in the nature of how household activities are accomplished (eg, technology that allows banking or shopping from

home) and who does them (eg, shifting gender roles). Third, the self-report physical capacity measures commonly used, such as Nagi items, focus on identifying those at the lower end of functional capacity, so provide no means of differentiating higher functioning individuals (5). In addition, item wording typically does not recognize the measurement problems posed by use of assistive devices or lack of opportunity to do activities (eg, persons who always use a walker are asked a hypothetical question about difficulty walking three blocks without special equipment). Fourth, despite the importance of engagement in productive, social, and organizational activities for older people's wellbeing, items to measure participation and restrictions in valued activities because of health or functioning have not been part of efforts to assess late-life disability in a national context.

Over a 2-year process that involved input from the scientific community and several rounds of cognitive testing, the NHATS team has reengineered self-reported measures of physical capacity and activity limitations and developed new items to measure participation and restrictions in valued activities. For all domains, we focus on experience in the last month because when asking about how activities are carried out, a time frame is critical for focusing the respondent's answers and minimizing qualification of answers (7). Moreover, when combined with information to be obtained at planned annual follow-up interviews, this approach will allow researchers to explore fine-grained analyses of the dynamics of disablement (ie, monthly onset and recovery).

Physical capacity.—We ascertained whether a respondent was able to do 12 tasks by her/himself and without equipment, if used. Rather than using the phrase “special equipment” where relevant, we followed the convention of first ascertaining assistive technology use and then tailoring the items accordingly (eg, someone who uses a cane would be asked “without using your cane” rather than “without special equipment”) (7). In order to identify individuals across a wider range of functional abilities, we included six pairs of less and more challenging tasks—walking three and six blocks, going up 10 and 20 stairs, lifting and carrying 10 and 20 pounds, bending over paired with kneeling down, reaching up over head without holding on paired with putting a heavy book on a shelf overhead, and using fingers to grasp small objects paired with opening a sealed jar. For the more challenging items, we also allowed respondents to volunteer “no opportunity.” For the less challenging items, we also assessed how much difficulty the respondent had carrying out the task.

For each pair of less and more challenging tasks, we divided respondents into three groups: unable to do less challenging, able to do less challenging but not more challenging, and able to do the more challenging task. Respondents who indicated for a more challenging task, no opportunity to do it (ranging from 14 for kneeling down to 51 for going up 20 stairs) or do not know ($n = 7$), were grouped with the

middle category. Respondents able to do the more challenging task but not the less challenging task were relatively rare ($n = 19$; most common was “grasping a small object” and “able to open a sealed jar”). These cases were categorized as able to do the more challenging task. Summary measures for physical capacity focused on the lower (unable 1+ less challenging tasks) and upper categories (able to do all challenging tasks).

In addition, we created more traditional (Nagi) hierarchies based solely on the less challenging tasks. For each item, we divided respondents into unable to do, able but with difficulty, and unable. We built Nagi summary measures (unable 1+ less challenging tasks and no difficulty with any less challenging tasks) to facilitate contrast with the NHATS physical capacity summary measures.

Ability to independently perform self-care and mobility tasks, given assistive technologies, if used.—NHATS measures of self-care and mobility limitations focus on the ability in the last month to perform activities by oneself (independent functioning), with whatever aids or devices or environments the respondent reports having used. Respondents first reported whether they used specific assistive devices while doing the activity, then whether help was received with the activity, and then (if relevant) how much difficulty was experienced when doing the activity by oneself, with the specific devices indicated. We followed the same tailored approach described above, specifying, for example, for someone who uses a cane to get out of bed “In the last month, when you used your cane, how much difficulty did you have getting out of bed by yourself?”

Activities included eating, bathing or showering (according to the respondent's preferred method if more than one occurred in the last month), toileting, dressing, getting out of bed, getting around inside one's home (or building if in a multiunit setting), and leaving one's home (or building if in a multiunit setting). Self-care devices included adapted utensils, grab bars in the bath or shower, a bath or shower seat, grab bars around the toilet, a raised toilet, and items to help with dressing, such as a button hook or clothing that is easy to get on or off. Mobility devices included cane, walker, wheelchair, and scooter. For each activity, we constructed two measures and also summarized these measures across all activities. Both measures identified those who had help or any difficulty (a little, some, or a lot) with an activity when performed independently (with assistive devices, if used). The second measure also included individuals who used assistive devices who did not report any help or difficulty.

We also included in the NHATS protocol the NLTC self-care and mobility screener items, which were of the form “Do you have any problem . . . without the help of another person or special equipment?” We combined answers to eating, bathing, getting to the bathroom and using the toilet, dressing, getting in or out of chairs, getting in and

out of bed, walking around inside, and going outside to create a summary measure reflecting any problem.

Ability to carry out household activities independently.—The NHATS protocol first assessed how household activities were carried out (in the last month: always by yourself, always with someone else, someone else always does it with you or for you, or it varies). Follow-up questions were tailored to this initial response. If the respondent ever did the activity by her/himself, the level of difficulty was assessed. A person who indicated someone else did the activity with or for them was asked whether the reason was related to their health or functioning. Although not used in the current analysis, we also collected information on the use of computers for selected activities, such as banking and shopping.

We focused on six household activities: doing laundry, doing light housekeeping, shopping for groceries, making hot meals, paying bills and banking, and keeping track of medications. We created a measure for each activity and a summary measure for all household activities that identified respondents who either had difficulty (a little, some, or a lot) with the activity when carried out independently or had another person do the activity with or for them because of their health or functioning.

Participation and restriction in valued activities.—For each activity, we asked frequency of participation in the last month, whether the respondent's health or functioning limited how often they did the activity and how important it was to the respondent to be able to participate (a lot, some, not at all). Components were assessed for eight types of activities: talking on the phone, e-mailing or texting, socializing in person, attending religious services, attending organized club meetings, going out for enjoyment, caring for another person, and volunteering. (Work was not included because a separate NHATS labor force module provides detail on current work and lifetime employment.)

We analyzed participation in two different ways: as frequency in the last month (with categories of every day, most days, some days, rarely, and never) and as any participation in the last month. We also constructed for each activity an indicator of participation restriction—whether health or functioning limited how often or kept person from doing an activity with some/a lot of importance to them—and also summarized any restriction across all activities.

Validation Measures

To assess convergent and construct validity of the NHATS self-reported disability measures, we analyzed key demographic characteristics, as well as physical and cognitive performance constructs (see Table 1 for sample distributions). Demographic characteristics included age (65–74, 75–84, and 85+), sex, and whether the respondent lived in a

residential care facility. A frailty measure was constructed using guidance from previously published work (8, 9). Scoring above the a priori criteria on 3 or more of five items was considered frail: unintentional weight loss of 10 or more pounds in the last 12 months; lowest 25th percentile of grip strength (within gender); self-report of low energy; usual walking speed of 0.6 m/second or more; and low physical activity as evidenced by self-reports of never walking for exercise and never doing vigorous activities in the last month. A memory score (0–20) was constructed by summing the number of words correctly recalled from a list of 10 nouns immediately after being read and again after a 5-minute delay. We then divided the sample into tertiles: high (10+), medium (7–9), and low (0–6). Finally, we calculated the short physical performance battery score from three physical performance batteries: a 4-m usual walking speed test, a rapid chair stand, and several balance tests. Each test was coded from 0 (unable) to 4 (best performance) and total scores ranged from 0 to 12. A small number of cases did not have room to do the walking test ($n = 22$) or a suitable chair for the stand test ($n = 9$) and were assigned the mean walking speed for their age and gender. High (10–12), medium (6–9), and low (0–5) groupings were formed.

Analytic Approach

Reliability of NHATS measures was assessed in two ways: percent agreement between responses from the initial and reinterviews and a kappa coefficient. Kappa, which ordinarily ranges from 0 to 1.0, indicates the amount of correspondence between measures adjusted for chance association. Because some amount of change even over a short period may be real, there is no expectation that kappa will be 1.0. Indeed, by convention, acceptable to good levels of kappa are between .6 and .8 and higher than .8 is excellent. Because kappa is sensitive to underlying prevalence, test–retest disagreement for relatively rare or infrequent conditions results in markedly low kappa statistics. For ordinal variables, we calculated a weighted kappa that takes into account distance between categories. As a check on the weighting scheme, we also calculated an interclass correlation coefficient for ordinal variables and found substantially similar or higher values (data not shown), suggesting that the weighted kappas are appropriate (10).

We assessed the reliability of summary measures for each domain (none vs any limitation or restriction) with the test–retest kappa statistic and assessed the internal consistency of items contributing to each summary measure with Cronbach's alpha, a measure of inter-item correlation.

Analysis of construct validity focused on correlations between the summary measures and key demographic and established performance-based measures of functioning. We also tabulated means (percentages) of external measures for each disability measure to facilitate interpretation. We focused on age and gender to replicate existing relationships

Table 2. Frequency and Reliability of Physical Capacity Measures in NHATS Validation Study*

	Percent (<i>n</i> = 326)			Percent Agree	Kappa
				(<i>n</i> = 111)	
Ability with less/more challenging tasks	1: Unable, less challenging	2: Able less, unable more challenging	3: Able more challenging		
Walk 3/6 blocks	31.3	20.9	47.9	84.7	.75
Walk up 10/20 steps	18.7	30.4	50.9	69.4	.49
Lift and carry 10/20 lbs	21.8	29.8	48.5	82.9	.72
Bend down/get down on knees and back up	22.1	42.9	35.0	73.9	.59
Reach over head/put a heavy book on shelf over head	12.9	18.4	68.7	73.0	.43
Using fingers to grasp small objects/open a sealed jar	2.5	25.2	72.4	90.1	.72
Inability/difficulty with less challenging (Nagi) tasks	1: Unable	2: Able with difficulty	3: Able no difficulty		
Walking 3 blocks	31.9	23.9	44.2	84.7	.75
Walking up 10 steps	19.0	36.2	44.8	77.5	.64
Lifting and carrying 10 lbs	22.1	29.5	48.5	75.7	.59
Bending over	23.9	26.1	50.0	74.8	.57
Reaching over head	17.2	22.4	60.4	76.6	.57
Using fingers to grasp small objects	7.7	27.3	65.0	78.4	.53

Note: *All items ask about ability/difficulty without help or equipment, if used.

in the literature, namely higher rates of disability among women and older adults. Known group validity was assessed with measures of whether a respondent lived in a residential care facility and frailty status (yes/no). We expected much higher prevalences of disability among those living in residential care and those considered frail, with one exception: We did not expect higher participation restrictions among persons in residential care settings because by design, these settings provide opportunities for socialization and leisure. Finally, convergent validity was assessed by examining correlations with a 10-word immediate and delayed recall score (ranging from 0 to 17) and a mobility score (ranging from 0 to 12) based on the short physical performance battery (11). We expected correlations to be negative and stronger for the mobility than for the cognitive score.

To serve as a point of comparison, we also examined the reliability and validity of summary measures for traditional Nagi items, which distinguish between difficulty and inability, and activities of daily living items from the NLTCSS screener, which ask about any problem carrying out the given activity without help from another person or special equipment.

Finally, as an initial exploration into the structure of the data, we ran a confirmatory factor analysis for all items comprising three of the NHATS summary measures (self-care/mobility limitations, household activity limitations, and participation restrictions, all defined below). We adopted an oblique rotation using promax, which allows the factors to be correlated. Based on the literature (12), we expected to find two factors: one reflecting activity limitations and a second for participation restrictions.

For all items, missing data were rare (4 or fewer cases, $\leq 1.2\%$) and, with the exception of more challenging capacity items, were combined with the most common (modal) response category. Sensitivity analyses in which we omitted

cases with missing responses produced nearly identical results (data not shown).

RESULTS

Reliability of Individual Measures

Reliabilities for physical capacity measures reflecting less-challenging/more challenging hierarchy ranged from .43 to .75 and performed as well or better than the inability/difficulty hierarchy (less challenging tasks only) for four of the six pairs (Table 2). However, pairs for climbing stairs (10 and 20) and reaching (overhead/put book on shelf) had reliabilities of .49 and .43, respectively.

For self-care activities, kappas ranged from .43 to .55 for getting help or having difficulty, depending on the activity, whereas the broader definition (including equipment use) yielded kappas in the .54–.81 range (Table 3). A similar pattern was observed for mobility-related activities, where kappas improved for individual items from the .51–.59 range to the .60–.78 range when individuals using equipment were included. Household activity limitations also demonstrated generally high reliability, with kappas in the .47–.77 range, and all but light housekeeping and paying bills and banking reached .70 or higher.

Test-retest reliability of measures of participation based on frequency of an activity in the last month (every day, most days, etc.) did not rise to adequate levels of reliability (range .33–.55, with one exception; Table 4). Weighted kappas were marginally better. However, dichotomous indicators of whether the respondent ever did the activity in the last month exhibited noticeably improved kappas (range of .56–.91; above .60 for all but volunteer activities). Although percent agreement for participation restrictions was uniformly high (>90% for all but volunteering), kappas were generally low, ranging from less than .30 (socializing and volunteering)

Table 3. Frequency and Reliability of Self-care, Mobility, and Household Activities in NHATS Validation Study

	Percent (n = 326)	Percent Agree (n = 111)	Kappa
Self-care			
Gets help or has difficulty*			
Eating	9.5	93.7	.55
Bathing	22.7	84.7	.43
Toileting	9.8	92.8	.52
Dressing	25.8	84.7	.53
Gets help, has difficulty*, or uses equipment			
Eating [†]	10.1	93.7	.55
Bathing	54.0	86.5	.72
Toileting	34.7	91.9	.81
Dressing	30.7	83.8	.54
Mobility			
Gets help or has difficulty*			
Getting around outside	27.9	82.0	.53
Getting around inside	20.6	84.7	.51
Getting out of bed	19.3	90.1	.59
Gets help, has difficulty*, or uses equipment			
Getting around outside	39.0	86.5	.70
Getting around inside	34.7	91.0	.78
Getting out of bed	23.3	87.4	.60
Household activities[‡]			
"Help" for health or functioning reason or difficulty			
Doing laundry	27.3	90.1	.71
Doing light housework	32.8	81.1	.52
Shopping for groceries	32.8	91.0	.77
Making hot meals	31.6	89.2	.70
Paying bills and banking	18.7	84.7	.47
Keeping track of medication	24.5	92.8	.73

Notes: *Difficulty (a little, some, or a lot) by oneself, with equipment if used.
[†] No one in the re-interview sample reported using equipment for eating.
[‡] Respondent has difficulty by him/herself or another person does the activity with or for the respondent because of a health- or functioning-related reason.

to .60 (club meetings or group activities) because of the sensitivity to disagreement for these low-prevalence items.

Reliability of Summary Measures

The reliability of NHATS derived summary measures for physical capacity fell in an acceptable range (.84 and .79 for alpha and .77 and .63 for kappa, see Table 5) and was as good as or marginally better than the Nagi summary measures. The measures also identified the top 16% of the sample as higher functioning (able to carry out all tasks) compared with 22% having no difficulty with the traditional (less challenging) Nagi tasks.

For self-care and mobility, inter-item correlations were strong (alpha = .81) and kappas were above .60 for both summary measures, an acceptable range. Reliability indicators for the NLTCS measures were higher (alpha = .86, kappa = .80), although the percentage of sample identified with the NLTCS measures was much lower than with the NHATS measures. For household activity and participation, reliabilities were in a similarly acceptable range. We also explored whether dropping items with lower reliability would improve kappa and found that excluding light housekeeping

Table 4. Frequency and Reliability of Participation Measures in NHATS Validation Study

	Percent (n = 326)	Percent Agree (n = 111)	Kappa	Weighted Kappa
How often respondent participates (every day, most days, some days, rarely, never)				
Talking on the phone	—	64.9	.47	.61
Using e-mail or texting	—	82.9	.72	.84
Socializing with friends or family	—	51.4	.33	.49
Attending religious services	—	71.2	.55	.55
Attending club meetings or group activities	—	64.0	.49	.58
Going out for enjoyment	—	66.7	.52	.56
Providing care to another person	—	64.9	.44	.59
Volunteering	—	62.2	.43	.51
Any participation in activities				
Talking on the phone	98.5	99.1	n/a*	—
Using e-mail or texting	39.9	95.5	.91	—
Socializing with friends or family	98.2	99.1	n/a*	—
Attending religious services	74.9	91.0	.76	—
Attending club meetings or group activities	72.4	87.4	.66	—
Going out for enjoyment	84.1	92.8	.68	—
Providing care to another person	44.2	82.0	.63	—
Volunteering	51.2	78.4	.56	—
Participation restriction[†]				
Talking on the phone	1.5	97.3	n/a*	—
Using e-mail or texting	0.9	98.2	n/a*	—
Socializing with friends or family	6.8	91.0	.24	—
Attending religious services	11.4	90.1	.47	—
Attending club meetings or group activities	7.1	93.7	.60	—
Going out for enjoyment	8.6	91.0	.40	—
Providing care to another person	7.7	94.6	.54	—
Volunteering	12.0	85.6	.26	—

Notes: *Could not be calculated because no variation.
[†] Respondent's health or functioning limited how often or kept respondent from participating in an activity that is a valued some or a lot.

from the summary household activity measure improved both the percent agreement (86%) and kappa (.71).

Construct Validity

Table 6 shows the correlations between the NHATS summary measures and key demographic- and performance-based constructs. Three points are noteworthy. First, the NHATS summary measures of physical capacity demonstrated associations in the expected direction, and, compared with the traditional difficulty-based Nagi measures, the NHATS approach identified a marginally higher functioning group (a year younger with slightly better walking speed on average and less likely to be female or classified as frail). Second, the NHATS summary measures for activity limitations (self-care and mobility activities and household activities) demonstrated strong associations in the expected directions, and correlations were of similar magnitude to those for the NLTCS activities of daily living summary measure. Finally, NHATS' new measure of participation restrictions was positively correlated with indicators

Table 5. Frequency and Reliability of Derived Summary Measures From NHATS Validation Study

Summary measures	Percent			Kappa
	Percent (n = 326)	Alpha	Agree (n = 111)	
NHATS				
Physical capacity (less/more challenging)				
Unable to do any less challenging task	39.3	.84	89.2	.77
Able to do all more challenging tasks	16.6	.79	85.6	.63
Activity limitation				
Self-care and mobility				
Any help or difficulty	48.5	.82	82.0	.64
Any help, difficulty, or equipment	70.9	.81	86.5	.72
Household activity				
Any help or difficulty*	56.1	.84	81.1	.62
Any activity limitation	76.4	.90	84.7	.67
Participation restriction				
Any restriction in valued activity	27.9	.71	86.5	.64
Nagi				
Unable to do any less challenging task	45.1	.82	86.5	.72
No difficulty with all less challenging tasks	22.4	.85	83.8	.62
NLTCS				
Any problem with self-care or mobility activity	27.3	.86	93.7	.80

Notes: NLTCS = National Long Term Care Survey.
*Excluding light housekeeping increases % agreement to 86% and alpha to .71.

considered here, and its association with an indicator of living in a residential care facility was smaller than for all other disability measures (.07).

Confirmatory Factor Analysis

Eigenvalues and scree plots suggest that the NHATS self-care/mobility limitation measures and participation restriction items scale as expected into two domains (Table 7). All items loaded on a single domain and loadings ranged between .42 and .76, with most above .60, suggesting two strong factors.

DISCUSSION

NHATS has developed a new measurement protocol to fill gaps in long-standing self-reported measures of disability and functioning. By and large, the new items and derived summary measures demonstrate robustness over a short time period, with most kappas for recommended items to be retained in the .60-.80 range. The new measures also correlate strongly as expected with existing constructs such as frailty, memory tests, and performance-based mobility and form what appear to be two factors, one reflecting activity limitations and a second reflecting participation restrictions.

This study has several limitations. The sample was relatively small and purposeful by design. Moreover, the test-retest framework for assessing reliability is necessarily limited because there will always be some degree of true change in activity limitations over even a brief time period. This is less likely for more severe limitations in personal care activities and mobility but may in particular affect reports of participation. Some measurement properties,

Table 6. Construct Validity of Derived Summary Measures From NHATS Validation Study (n = 326)

Summary Measure	Correlations					M/Percentages					Mobility Score*	Word Recall Score	Frailty	N	
	Age	Female	Facility	Frailty	Word Recall Score	Mobility Score*	% Female	% Facility	% Frailty	% Frailty					
NHATS															
Physical capacity (less/more challenging)															
Unable to do any less challenging task	.27	.28	.29	.25	-.21	-.62	79.8	73.4	28.9	39.8	7.0	5.2	128		
Able to do all more challenging tasks	-.21	-.34	-.12	-.23	.16	.39	73.4	18.5	5.6	3.7	9.2	10.2	54		
Activity limitation															
Self-care and mobility															
Any help or difficulty	.21	.15	.24	.29	-.25	-.48	78.9	63.9	24.7	39.2	7.0	6.0	158		
Any help, difficulty, or equipment	.28	.21	.22	.20	-.19	-.41	78.5	63.2	20.8	31.6	7.5	6.7	231		
Household activities															
Any help or difficulty	.26	.13	.35	.28	-.21	-.49	79.1	62.7	27.8	37.9	7.2	6.1	169		
Any activity limitation	.26	.16	.22	.18	-.16	-.37	78.3	61.0	20.0	30.5	7.6	6.9	249		
Participation restriction															
Any restriction in valued activity	.14	.17	.07	.14	-.17	-.36	78.9	70.3	19.8	36.3	7.0	5.8	91		
Nagi															
Unable to do any less challenging task	.26	.25	.31	.22	-.24	-.55	79.4	70.1	27.9	36.7	7.0	5.7	147		
No difficulty with all less challenging tasks	-.19	-.22	-.17	-.20	.18	.39	74.4	35.6	4.1	9.6	9.1	9.8	73		
NLTCS															
Any problem with self-care or mobility activity	.28	.18	.32	.26	-.20	-.61	80.7	70.8	34.8	44.9	6.8	4.5	89		

Notes: NLTCS = National Long Term Care Survey.
*Measured with the short physical performance battery.

Table 7. Factor Loadings for NHATS Validation Study Measures
($n = 326$)*

	Activity Limitation	Participation Restriction
Doing laundry [†]	0.76	
Shopping for groceries [†]	0.66	
Making hot meals [†]	0.72	
Paying bills and banking [†]	0.59	
Keeping track of medication [†]	0.61	
Bathing [‡]	0.53	
Toileting [‡]	0.51	
Dressing [‡]	0.61	
Eating [‡]	0.44	
Getting around inside [‡]	0.73	
Getting outside [‡]	0.71	
Getting out of bed [‡]	0.47	
Socializing with friends or family [§]		0.60
Attending religious services [§]		0.52
Attending club meetings or group activities [§]		0.45
Going out for enjoyment [§]		0.49
Providing care to another person [§]		0.56
Volunteering [§]		0.42
Eigenvalues	8.4	1.2

Notes: *Confirmatory factor analysis with oblique (promax) rotation and correlated factor structure.

[†]Set = 1 if respondent has difficulty by him/herself or another person does the activity with or for the respondent because of a health- or functioning-related reason.

[‡]Set = 1 if gets help, has difficulty by oneself (when using equipment if used), or uses equipment.

[§]Set = 1 if health or functioning limited how often or kept respondent from doing the activity and the activity is valued (some or a lot).

such as responsiveness, also cannot be evaluated until longitudinal data are collected. In addition, we limited our investigation of summary measures here to dichotomous indicators, but future work should investigate the value of more continuous indices. Finally, analyses presented here were based on Cronbach's alpha and confirmatory factor analyses. Item Response Theory may be a promising next step to further investigate properties of individual items and underlying scale structures because the approach allows comparisons of individual ability with item (activity) difficulty (12).

Despite these limitations, this analysis points to some important improvements on the horizon in measuring the disablement process. Comparisons of NHATS' physical capacity measures with the more traditional Nagi items, for example, suggests equally good if not marginally better reliability and a slightly more discriminating measure of capacity at the higher end. When used in combination with measures of behavior change and assistive device use, such measures may offer researchers the ability to study earlier phases of disablement than previously possible on a national scale.

Our analysis also addressed the question of how the NHATS' measures compare with a subset of items used by its predecessor survey. Our analysis of the NLTCs self-care and mobility screener items, included in the NHATS validation study protocol, suggests on balance favorable findings.

We were able to create new measures of activity limitations and physical capacity that offer analysts more details about how activities are performed, while still reaching acceptable levels of reliability and validity.

By including new measures of participation and restrictions in valued activities, NHATS will allow researchers to investigate under what circumstances changes in physical capacity result in loss of engagement and conversely whether there are particular common factors that facilitate engagement despite such losses. Our analysis suggests that participation over the last month can be reliably captured, especially in the form of a summary measure of any participation restriction; investigators should be more cautious in analyzing changes in restrictions in particular activities, which are not as robust over a short time period.

Our analysis also has implications for the NHATS instruments currently in preparation for the first wave of national data collection. Based on the analyses reported here, we have made several refinements. First, the items reflecting higher and lower physical capacity will be administered as paired questions (more challenging, then if unable, less challenging). In one instance where the items were less consistent (using fingers to grasp small objects, opening a sealed jar), a change to specify "opening a sealed jar *with just your hands*" is expected to reduce the mismatch. A second substantive change is the decision to drop light housekeeping from the set of household activities included. Our analyses pointed to improved reliability of a summary measure of difficulty in one or more household activities without loss of validity. Finally, for participation and participation restrictions, we found that reliability improved appreciably when participation in the last month was framed as ever/never, in contrast to frequency (every day, most days, etc.); we have incorporated this approach. Items related to social interaction (phone, e-mail, and in person) have also been consolidated to shorten administration time.

NHATS is intended as a new platform for studying later life disability trends and dynamics. The final study protocol preserves the ability of researchers to examine more traditional measures of functioning while modernizing measures of disability and expanding their scope—in particular recognizing environmental aspects of activity limitations, broadening measures of physical capacity, and introducing new measures of late-life participation restrictions. Evidence for the validity and reliability of these new measures is a necessary foundation for their inclusion in a national study and use in future research to understand disability trends among older adults and how functional changes unfold in later life.

FUNDING

This research was supported by the National Institute on Aging (Cooperative Agreement 1U01AG032947-01).

ACKNOWLEDGMENTS

We thank Marlene Niefeld for her assistance in preparing the performance data. The views expressed are those of the authors alone and do not represent those of the funding agency or the authors' institutions.

REFERENCES

1. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged. The Index of ADL, a standardized measure of biological and psychosocial function. *JAMA*. 1963;185:914-919.
2. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*. 1969;9:179-186.
3. Nagi SZ. Some conceptual issues in disability and rehabilitation. In: Sussman MB, ed. *Sociology and Rehabilitation*; Washington, DC: American Sociological Association; 1965:100-113.
4. Agree EM. The influence of personal care and assistive technology on the measurement of disability. *Soc Sci Med*. 1999;48:427-443.
5. Simonsick EM, Newman AE, Nevitt MC, et al. Measuring higher level physical function in well-functioning older adults: expanding familiar approaches in the Health ABC Study. *J Gerontol A Biol Sci Med Sci*. 2001;56:M644-M649.
6. Freedman V. Adopting the ICF language for studying late-life disability: a field of dreams? *J Gerontol A Biol Sci Med Sci*. 2009;64:M1172-M1174.
7. Freedman VA, Agree EM, Cornman J. *Development of An Assistive Technology and Home Environment Assessment Instrument for National Surveys: Final Report. Part I. Recommended Modules and Instrument Development Process*. 2005. <http://aspe.hhs.gov/daltcp/reports/ATEAdev1.pdf>. Accessed March 7, 2011.
8. Bandeen-Roche K, Xue QL, Ferrucci L, et al. Phenotype of frailty: characterization in the women's health and aging studies. *J Gerontol A Biol Sci Med Sci*. 2006;61A:262-266.
9. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol Biol Sci Med Sci*. 2001;56A:M146-M156.
10. McClure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol*. 1987;126:161-169.
11. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol Biol Sci Med Sci*. 1994;49:M85-M94.
12. Jette AM, Haley SM, Kooyoomjian JT. Are the ICF activity and participation dimensions distinct? *J Rehabil Med*. 2003;35:145-149.