Chapter 3

Reliability and validity of the assessment of daily activity performance (ADAP) in community-dwelling older women

Paul L. de Vreede[1] BSc, Monique M. Samson[1], MD, PhD, Nico L. van Meeteren[2,3], PT, PhD, Sijmen A. Duursma[1], MD, PhD, Harald J. Verhaar[1], MD, PhD

[1] Mobility Laboratory, Department of Geriatric Medicine, University Medical Center Utrecht, the Netherlands;
[2] Rudolf Magnus Institute of Neuroscience, Department of Neurology and Neurosurgery, University Medical Center Utrecht, the Netherlands;
[3] Department of Physiotherapy, Academy of Health Sciences, Utrecht, the Netherlands

Aging Clin Exp Res, accepted for publication.

ABSTRACT

Background and aims: The Assessment of Daily Activity Performance (ADAP) test was developed, patterned after the Continuous-scale Physical Functional Performance (CS-PFP) test, to provide a quantitative assessment of older adults' physical functional performance. The aim of this study was to determine the intra-examiner reliability and construct validity of the ADAP in a community-living older population, and to identify the importance of tester experience.

Methods: Forty-three community-dwelling, older women (mean age 75 yr ± 4.3) were randomised to the test-retest reliability study (n = 19) or the validation study (n = 24). Intra-examiner reliability of an experienced (tester 1) and an inexperienced tester (tester 2) was assessed by comparing test and retest scores of 19 participants. Construct validity was assessed by comparing the ADAP scores of 24 participants with self-perceived function using the SF-36 Health Survey, muscle function tests, and the Timed Up and Go test (TUG).

Results: Tester 1 had good consistency and reliability scores (mean difference between test and retest scores, -1.05 ± 1.99; 95% confidence interval (CI), -2.58 to .48; Cronbach's alpha (α) range, .83 to .98; intraclass correlation (ICC) range, .75 to .96; Limits of Agreement (LoA), -2.58 to 4.95). Tester 2 had lower reliability scores (mean difference between test and retest scores, -2.45 ± 4.36; 95% CI, -5.56 to .67; α range, .53 to .94; ICC range, .36 to .90; LoA, -6.09 to 10.99), with there being a systematic difference between test and retest scores for the ADAP domain lower-body strength (-3.81; 95% CI, -6.09 to -1.54). ADAP correlated with SF-36 Physical Functioning scale (r = .67), TUG test (r = -.91), and with isometric knee extensor strength (r = .80).

Conclusions: The ADAP test is a reliable and valid instrument. Our results suggest that testers should practise using the test, to improve reliability, before using it in clinical practice.

INTRODUCTION

In exercise studies, the most commonly used measures of physical function are self-report activities of daily living (ADL) questionnaires, such as the Katz and Barthel index (1-3), health-related quality of life questionnaires, such as the Short Form Health Survey (SF-36) (4-6), and selected intermediate outcome measures, such as muscle strength and gait speed.(4, 7-10) However, the extent to which these assessments are responsive to meaningful changes in the functioning of community-living, healthy individuals has been questioned (4, 11, 12). For example, ADL questionnaires usually fail to detect changes in healthy participants because of ceiling effects (11, 13, 14). Furthermore, although improved intermediate outcome measures, such as muscle strength or gait speed, have been equated with improved performance of daily activities (7, 9, 10, 15), an increase in muscle strength or walking speed does not necessarily mean that the performance of functional tasks is improved (4, 14). Thus, when evaluating interventions aimed at improving the ability of healthy individuals to perform everyday activities, it is essential to use measures of physical function that are not affected by ceiling effects.

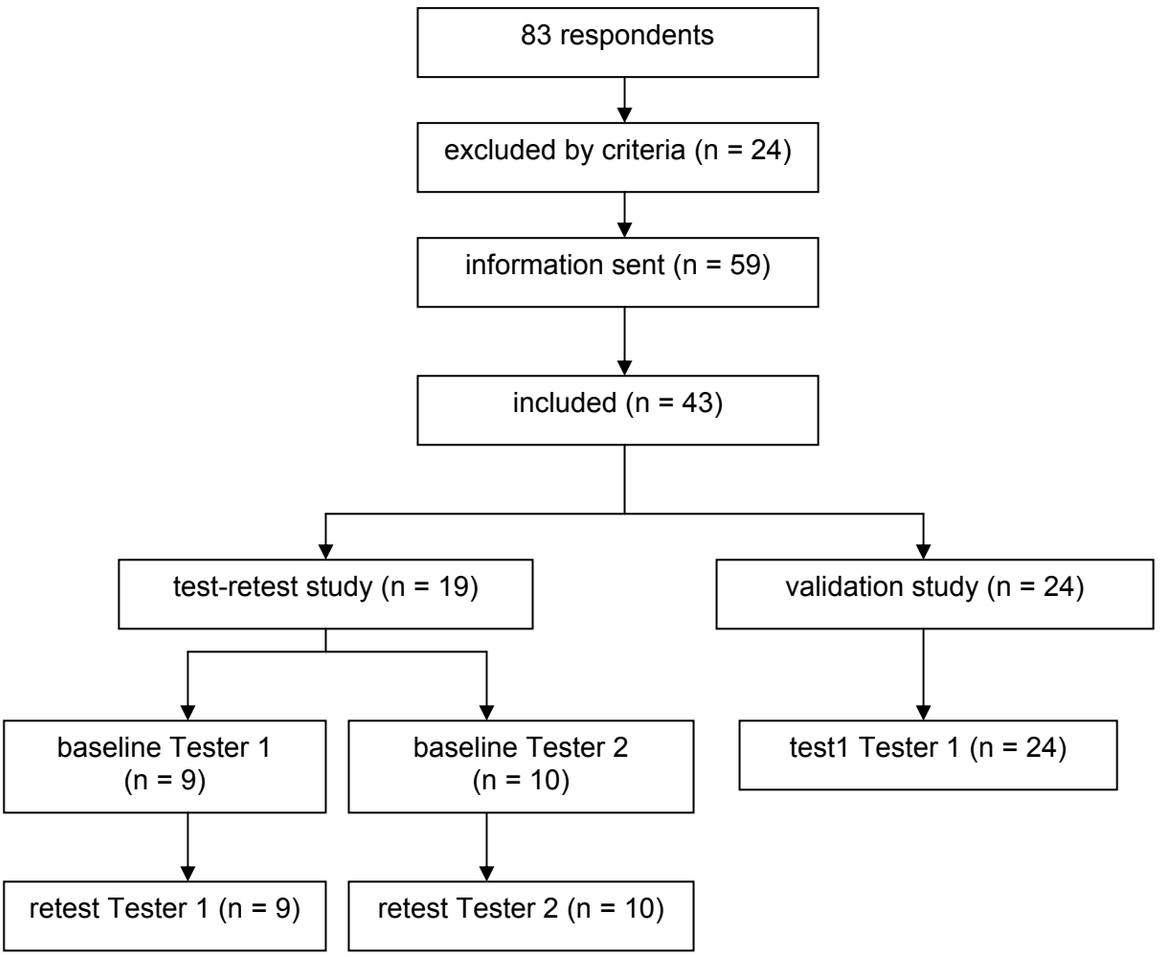
We developed the quantitative Assessment of Daily Activity Performance (ADAP) test (16, 17). This method was patterned after the Continuous-scale Physical Functional Performance (CS-PFP) test, as demonstrated by Cress et al to be reliable, valid, sensitive to change, and without ceiling or floor effects (11, 18). The CS-PFP test was modified to Dutch dimensions for bed size, height of the kitchen counter, and height of the washing machine. The vertical reach task was replaced by the functional reach test (19). These modifications of the CS-PFP test makes that the ADAP test differs on approximately 30% of the tasks performed during the test. Therefore, the ADAP test should be approached as a different test and validity and reliability need to be established. The aim of this study was to examine the reliability and construct validity of the ADAP test in a sample of community-living older people. Because a tester's experience may affect test results (20, 21), we compared the ADAP test results of an experienced tester from our mobility laboratory with those of an inexperienced tester.

METHODS

Participants and examiners

Eighty-three community-dwelling women older than 70 years were recruited from the Utrecht region through newspaper advertisements. Of the 83 respondents, 24 were excluded after telephone interviews. Exclusion criteria included: recent fractures; unstable cardiovascular, metabolic, musculoskeletal condition, or other chronic illnesses that might limit testing; severe airflow obstruction; recent depression or emotional distress; or any reason for a loss of mobility for more than 1 week in the previous 2 months. After reading about the study, 43 respondents participated in the present study. Figure 1 shows the flow of participants through the study. Nineteen participants were randomly assigned by computer to a test-retest trial to determine reliability, and 24 respondents were assigned to the validation study.

Figure 1. Study profile



The 19 participants of the test-retest trial were randomly assigned by computer to one of two testers (Tester 1 and Tester 2). Tester 1 was a 26-year-old female research assistant and Tester 2 was a 29-year-old male PhD-student. Before the start of this trial, Tester 1 had administered the ADAP 29 times and Tester 2 only 4 times. All measurements of the validation study were obtained by the experienced Tester 1 after the measurements of the reliability study.
This study was approved by the Medical Ethics Board of the University Medical Center Utrecht University Hospital in the Netherlands. Written informed consent was obtained from all participants after they had read the information brochure on the study.

Measurements

The tests were administered at the Mobility Laboratory. Participants of the test-retest study were tested on two separate occasions, 1 week apart at a similar time of day (early morning, late morning, early afternoon, or late afternoon) by the same examiner. At the beginning of each test session, participants were asked if during the week prior to the test something had occurred that might have influenced their performance on the ADAP test (e.g. illness, injury, or stressful situation). After the ADAP, the participants of the validation study completed the SF-36 Health Survey, followed by several muscle function tests, and the Timed Up and Go (TUG) test.

Assessment of Daily Activity Performance (ADAP)

The ADAP test was patterned after the Continuous-scale Physical Functional Performance© (CS-PFP) test, as demonstrated by Cress et al to be reliable, valid and sensitive to change in function (11, 18). Like the CS-PFP test, the ADAP includes 16 common tasks, such as transferring laundry and boarding a bus, and allows the participant to perform at maximal ability by maximizing the weight carried and working at the fastest speed possible or reaching the greatest distance (11, 16). The CS-PFP test was modified to Dutch dimensions for bed size (190 cm x 200 cm; height 60 cm), height of the kitchen counter (114 cm), and height of the washing machine (88.5 cm). Vertical reach was replaced by the functional reach test (19) because the combination of a forward standing reach and a sit-and-reach task (putting a Velcro-closed strap over the shoe) is a more familiar method in the literature to determine upper-body flexibility than the combination of the vertical reach

test and a sit-and-reach task as proposed by others (19, 22). Furthermore, Schenkman et al demonstrated a relationship between spinal flexibility and functional reach (23). Measurement protocols and participant instructions were standardised. The ADAP measures whole-body physical function, assessing upper and lower-body strength, upper-body flexibility, balance and coordination, and endurance. In general, scores on a specific task can contribute to one, two, or three domains. Tasks quantified by both weight carried and time are “carrying a weighted pan”, “pouring water from a jug into a cup”, “carrying weight up and down a bus platform”, and “carrying groceries”. Tasks quantified by time alone are “transferring laundry from a washer to a dryer”, “putting on and taking off a jacket”, “floor sweeping”, “vacuuming”, “making a bed”, “climbing stairs”, “getting down and up from the floor”, “opening a door”, “putting a hook-and-loop strap over a shoe”, and “picking up four scarves from the floor”. Tasks quantified by distance are “6-minute walk” and “functional reach”.
The scoring procedures of the ADAP test are provided in the Appendix. Each task was scaled 0 to 100 according to the formula: Observed score = (observed score – lower limit) / (upper limit – lower limit) x 100. If the observed score was less or equal to the lower limit, the score was 0. For an observed score greater than or equal to the upper limit, the score was 100. Unattempted tasks received a score of 0. Time was converted to speed (1/t) so that higher numbers reflect a better function for each of the units measured: weight, distance, and speed. Domain scores are calculated as the mean of task scores that contribute to the domain as presented in the Appendix. The ADAP total score is calculated as the mean of all task scores. The average time required to complete the test for community-living older women is 60 minutes. The main role of the tester in the ADAP consists of explaining the tasks to the participant and registering the time needed to complete a task and the weight carried during a task. We reported previously that the ADAP test can detect a change in daily task performance after a 12-week of exercise period in a relatively healthy group of older women (17). A description of the protocol to perform the ADAP can be obtained from the authors.

Self-Perceived Function

Self-perceived function of the 24 participants in the validation study was determined using the Dutch language version of the SF-36 Health Survey (24). The SF-36 is a 36-item questionnaire that measures physical and mental disability and well-being. It

includes eight multi-item scales that measure physical functioning (PF), role limitations due to physical health problems (RP), bodily pain (BP), general health perceptions (GH), vitality (VT), social functioning (SF), role limitations due to emotional problems (RE), and mental health (MH). Scales are scored from 0 (poorer health) to 100 (better health). The Dutch language version of the SF-36 has proven to be a practical, reliable, and valid instrument for use in general population surveys in the Netherlands (24).

Timed Up and Go (TUG)

In the Timed Up and Go, the time an individual needs to rise from a standard arm chair (seat 46 cm high), walk 3 meters, turn around, return to the chair, and sit down again is measured (16, 25, 26). The test was performed three times as quickly as possible. The quickest time, recorded in seconds (sec), was used for analysis.

Muscle function

Isometric knee extensor strength (IKES) was measured in both legs with a fixed strain gauge (AFG-Advanced Force Gauge, Mecmesin Inc, Santa Rosa, California, USA) (16, 27, 28). The highest score of five attempts was recorded in Newton (N).

Isometric elbow flexor strength (IEFS) was measured in both arms with a hand-held dynamometer (microFET, Hoggan Health Industries, Draper, Utah, USA) (16, 29, 30). The highest score of three attempts was recorded in Newton (N).

Handgrip strength (HGS) was measured with a mechanical handgrip dynamometer (Takei Kiki Kogyo 5101, Japan) (16, 29). The best score of five attempts was recorded in kilogram force (kgF).

Leg extension power (LEP) was measured in both legs with the Nottingham power rig (NUMAS, University of Nottingham Medical Faculty Workshops, Nottingham, UK) (16, 29, 31). The measurements were repeated until no further improvement was seen, up to a maximum of 10 pushes (16, 27). The highest recorded power output was recorded in Watt (W).

Peak values for the left and right legs, arms or hands of IKES, IEFS, HGS and LEP were averaged and used for analysis.

Statistics

All statistical analyses were performed with SPSS software (SPSS Inc. Spss reference guide. Chicago: SPSS Inc, 1990). Univariate analysis of variance was used to test for differences in baseline characteristics between groups.

Reliability and Internal Consistency

Often, the reliability of physical measures is established by calculating the Pearson correlation coefficient (11, 20), a method that is considered inadequate to determine reliability because of the incapacity to detect systematic differences (20, 32). To assess reliability, first, the coefficient of internal consistency was measured with Cronbach's alpha (α). An alpha of 0.6 or greater indicated that the items in the scale measured the same contribute. Second, test-retest reliability was measured with the intraclass correlation (ICC), calculated with a one-way random model, and with the mean difference and limits of agreement (20, 32, 33). The latter were calculated using Brand and Altman plots (32), in which the limits of agreement (D – 2s, D + 2s) were put into the standard mathematical expression as delta – 2SD and delta + 2SD, in which delta is the mean of the differences between two ratings for the same subject, and SD is the standard deviation of the differences. Because the measurement errors probably follow a Gaussian distribution, 95% of the differences will lie between these limits of agreement, more precisely, between delta – 1.96SD and delta + 1.96SD.

Levene's test for equality of variance was performed to compare the test-retest differences between Tester 1 and Tester 2.

Construct validity

We hypothesized that maximum muscle strength, muscle power, mobility, and self-perceived physical function would be positively associated with ADAP scores. The ADAP test results were compared with the results of IKES, IEFS, HGS, LEP, TUG, and SF-36 by calculating bivariate Pearson correlations between these tests and total and subscale scores of the ADAP.

RESULTS

Reliability

Characteristics of the participants are listed in Table 1. No differences were found between the participants examined by Tester 1, the participants examined by Tester 2 and the participants of the validation study for baseline scores for weight, height, age or physical functional performance. The nine participants randomised to Tester 1 had a mean age of 74.1 ± 3.4 years (range, 70 – 80 years) and the 10 participants randomised to Tester 2 had a mean age of 75.8 ± 3.9 years (range, 70 – 83 years). No participants reported incidents that might have influenced test performance.

Table 1. Baseline characteristics of participants

Characteristics	*Tester1 n = 9*	*Tester 2 n = 10*	*Validity n = 24*	*P-value*
Age, years	74.1 ± 3.4	75.8 ± 3.9	74.6 ± 4.8	.68
Weight, kilograms	73.8 ± 11.9	66.1 ± 7.4	65.5 ± 10.6	.12
Height, meters	1.62 ± 0.05	1.61 ± 0.04	1.59 ± 0.05	.46
ADAP test				
Total score	40.0 ± 7.4	47.7 ± 6.3	43.3 ± 14.3	.36
Upper-body strength	42.8 ± 9.7	48.4 ± 5.7	40.8 ± 13.3	.22
Upper-body flexibility	45.9 ± 9.0	47.9 ± 10.5	45.2 ± 16.9	.88
Lower-body strength	34.4 ± 8.7	40.8 ± 6.2	38.3 ± 15.5	.55
Balance and coordination	34.4 ± 8.7	44.7 ± 9.6	42.4 ± 16.0	.23
Endurance	37.8 ± 8.0	49.2 ± 8.4	44.4 ± 15.8	.18
SF-36				
Physical Component Summary (PCS)			72.2 ± 16.4	
Mental Component Summary (MCS)			80.7 ± 14.9	
Physical Functioning (PF)			75.4 ± 16.6	
Role-Physical (RP)			71.9 ± 36.4	
Bodily Pain (BP)			75.3 ± 18.8	
General Health (GH)			66.3 ± 16.9	
Vitality (VT)			69.8 ± 16.6	
Social Functioning (SF)			90.1 ± 13.3	

Role-Emotional (RE)	84.7 ± 34.0
Mental Health (MH)	78.0 ± 12.5
TUG, seconds	6.0 ± 1.9
IKES, N	244.1 ± 84.9
IEFS, N	144.8 ± 23.9
LEP, W	95.3 ± 38.5
HGS, kg Force	20.3 ± 5.7

Note: Values are means ± SD
Abbreviation: ADAP, assessment of daily activity performance; SF-36, Short Form 36 Health Survey; TUG, timed up and go; IKES, isometric knee extensor strength; IEFS, isometric elbow flexor strength; LEP, leg extensor power; HGS, handgrip strength; N, newtons; W, watts.

Cronbach's alpha, ICCs, and the parameters according to the Bland and Altman plot (mean difference, limits of agreement) are presented in Table 2. The values for Cronbach's alpha indicated a good internal consistency for Tester 1 (alpha range, .83 to .98) and for Tester 2 (alpha range, .80 to .94), with the exception of ADAP upper-body strength (alpha .53). The variance in the difference between test and retest scores differed between the testers for ADAP balance and coordination and endurance scores.

The total score of the ADAP test at baseline and for the retest are presented in Figure 2. The mean difference between test and retest scores did not differ significantly from zero for Tester 1, whereas it did for Tester 2 for ADAP lower-body strength (-3.81; 95% confidence interval [CI], -6.09 to -1.54). Tester 1 showed a high reliability for ADAP total and domain scores (ICC range, .75 to .96), whereas Tester 2 had lower ICC's for ADAP total and domain scores (ICC range, .36 to .76), except for upper-body flexibility (ICC .90). The ADAP upper-body strength measurements of Tester 2 were not reliable (ICC .36). A scatter plot of the difference between scores against the mean ADAP total score for Tester 1 and 2 is presented in Figure 3. The horizontal lines in these graphs represent the limits of agreement. There was a

greater difference between test and retest scores for Tester 2 (range, -6.54 to 6.48) than for Tester 1 (range, -2.10 to 3.59). The limits of agreement were also larger for Tester 2 (-6.09 to 10.99) than for Tester 1 (-2.58 to 4.95).

Figure 2. Scatterplot of the ADAP total score at baseline and retest.

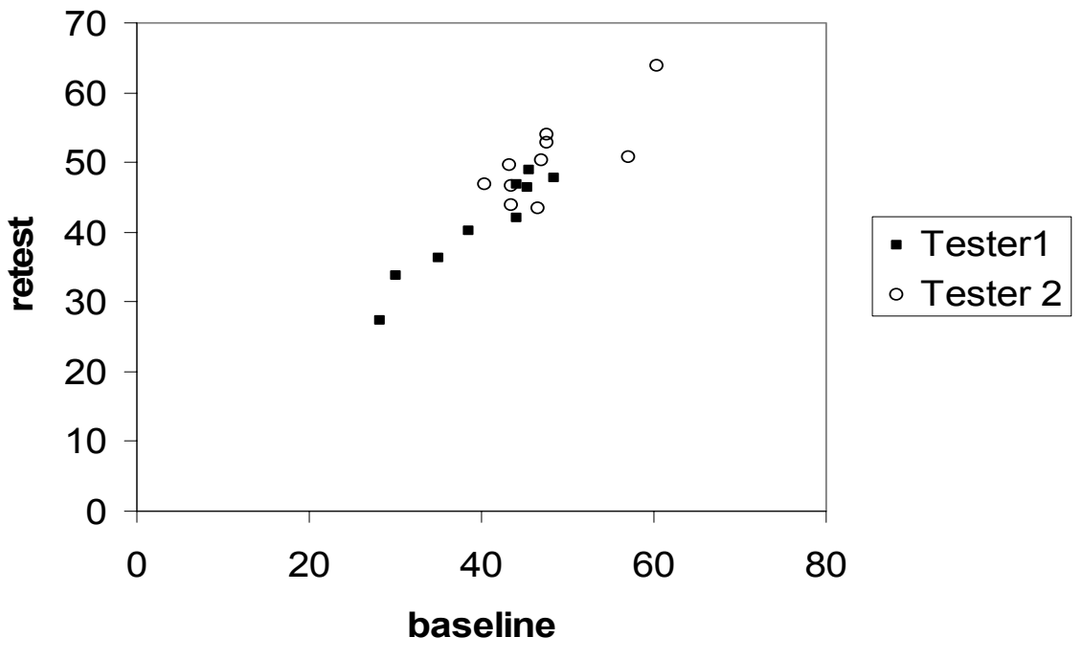
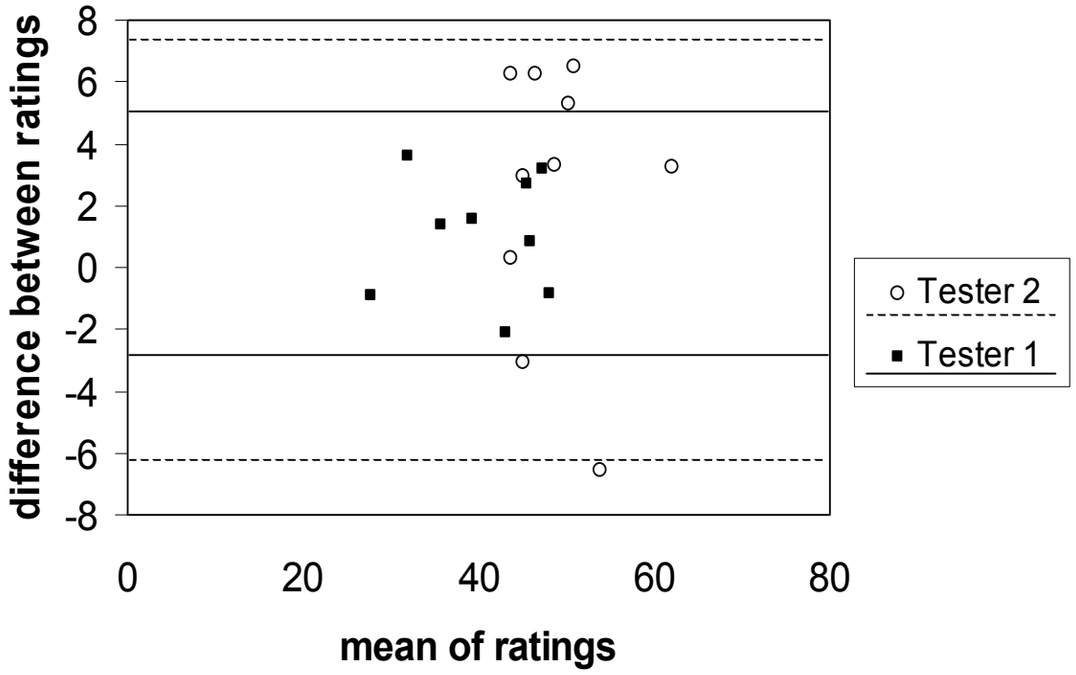


Table 2. Intra-tester consistency and reliability for the ADAP test in a test-retest design with an experienced tester (Tester 1) and an inexperienced Tester (tester 2).

	Tester 1 *n = 9*				*Tester 2* *n = 10*				*Tester 1 vs. Tester 2*
ADAP	Mean difference (95% CI)	Alpha	ICC	LoA	Mean difference (95% CI)	Alpha	ICC	LoA	P-value
Total score	-1.05 ± 1.99 (-2.58 to .48)	.98	.96	-2.85 to 4.95	-2.45 ± 4.36 (-5.56 to .67)	.86	.70	-6.09 to 10.99	.08
Upper-body strength	-1.42 ± 4.90 (-5.18 to 2.34)	.91	.84	-8.17 to 11.02	-1.70 ± 5.47 (-5.62 to 2.21)	.53	.36	-9.03 to 12.43	.83
Upper-body flexibility	1.66 ± 6.81 (-3.58 to 6.90)	.83	.75	-15.01 to 11.69	.15 ± 4.34 (-2.96 to 3.25)	.94	.90	-8.65 to 8.36	.07
Lower-body strength	-2.24 ± 3.65 (-5.05 to .57)	.96	.91	-4.92 to 9.40	-3.81 ± 3.19 (-6.09 to -1.54)	.94	.74	-2.43 to 10.06	.51
Balance and coordination	-1.43 ± 3.18 (-3.87 to 1.01)	.97	.93	-4.79 to 7.66	-3.03 ± 8.14 (-8.85 to 2.80)	.80	.65	-12.92 to 18.97	.03
Endurance	-1.44 ± 2.41 (-2.58 to .48)	.98	.95	-3.28 to 6.16	-2.27 ± 5.56 (-6.24 to 1.71)	.87	.76	-8.63 to 13.16	.03

Note: Abbreviations: ADAP, Assessment of Daily Activity Performance; CI, confidence interval; Alpha, Cronbach's alpha; ICC, intraclass correlation; LoA, limits of agreement.

Figure 3. Bland and Altman scatterplot of the intratester reliability of the ADAP total score.



Note: Difference between scores against the mean of ratings (sum scores). Horizontal lines show the limits of agreement for Tester 1 and Tester 2 (dotted lines).

Construct validity

The mean age of the participants was 74.6 ± 4.8 years (range 70 – 91 years) (Table 1). SF-36 scores and muscle strength results were high, indicating that the participants were in good physical and mental health. Bivariate correlations between ADAP scores, SF-36 scales, and strength and mobility tests are shown in Table 3. ADAP total and all domain scores correlated significantly with the physical component summary scale (PCS) and physical functioning scale (PF). Also, ADAP total and domain scores correlated with the scales Bodily Pain (BP) and General Health (GH). ADAP total and domain scores also were highly correlated with TUG test (range, r = -.77 to -.91), IKES (range, r = .64 to .80), LEP (range, r = .56 to .63), and HGS (range, r = .51 to .74) scores.

Table 3. Correlation coefficients between ADAP test, self-perceived function (SF-36), muscle function, and mobility measures.

SF-36	*GH*	*MH*	*PF*	*RP*	*VT*	*SF*	*RE*	*BP*	*PCS*	*MCS*
ADAP										
Total score	.45*	.23	.67**	.30	.38	.33	.39	.59**	.64**	.25
Upper-body strength	.50*	.30	.78**	.36	.51*	.34	.47*	.62**	.71**	.33
Upper-body flexibility	.35	.32	.56**	.43*	.42*	.23	.36	.58**	.62**	.27
Lower-body strength	.45*	.13	.69**	.26	.33	.35	.43*	.56**	.63**	.22
Balance & coordination	.43*	.11	.55**	.21	.25	.36	.32	.49*	.55*	.17
Endurance	.42*	.19	.60**	.24	.32	.32	.33	.55**	.59**	.20

	TUG	*IKES*	*IEFS*	*LEP*	*HGS*
ADAP					
Total score	-.91**	.80**	.54**	.63**	.62**
Upper-body strength	-.80**	.76**	.59**	.56**	.74**
Lower-body strength	-.84**	.77**	.55**	.53**	.63**
Upper-body flexibility	-.77**	.64**	.42*	.57**	.53**
Balance and coordination	-.85**	.76**	.50*	.59**	.51*
Endurance	-.91**	.77**	.50*	.62**	.56**

Note: Values are Pearson *r*; * $p<.05$; ** $p<.01$

Abbreviations: ADAP, Assessment of Daily Activity Performance; SF-36, Short Form 36 Health Survey; GH, General Health; MH, Mental Health; PF, Physical Functioning; RP, Role-Physical; VT, Vitality; SF, Social Functioning; RE, Role-Emotional; BP, Bodily Pain; PCS, Physical Component Summary; MCS, Mental Component Summary; TUG, timed up and go; IKES, isometric knee extensor strength; IEFS, isometric elbow flexor strength; LEP, leg extensor power; HGS, handgrip strength.

DISCUSSION

The results of this study show that the Assessment of Daily Activity Performance (ADAP) is a reliable and valid instrument for measuring physical function in community-dwelling older women.

While intraclass correlation coefficients (ICC) or Pearson product moment are often used to determine the reliability of an instrument (20), they are considered inappropriate because they do not detect systematic differences (20, 32). In the present study, we used the intraclass correlation coefficient (ICC), mean difference and Limits of Agreement, and Cronbach's alpha internal consistency analysis to determine the reliability of the ADAP, because of their complementary value (20, 33). Furthermore, according to Bland and Altman (32) the scatter plot of differences between test and retest scores plotted against the mean of the scores provides insight into the distribution of differences between two measurements, and the limits of agreement represent an estimate of the range of rating-pair differences with 95% of the differences between two ratings. Results showed that the internal consistency and intra-rater reliability of the test were higher when an experienced tester (Tester 1) administered the test. The limits of agreement were smaller for Tester 1 (-2.58 to 4.95), who administered the test 29 times before the study, than for Tester 2 (-6.09 to 10.99), who had administered the test only 4 times previously. The results obtained by the less experienced tester were less consistent and less reliable. There was also a statistically significant difference between test and retest scores for the ADAP domain lower-body strength. In the tests of the ADAP, participants are encouraged by the tester to exert maximum effort. These maximum capacity measurements probably were more consistent for the experienced tester, and thus a trained tester may be better able to stimulate participants. The main role of the tester in the ADAP consists of explaining the tasks to the participant and registering the time needed to complete a task and the weight carried during a task. The results of the present study suggest that before using the ADAP a tester first has to complete a learning phase to obtain reliable measurements.

Cress and colleagues (11) used the Pearson product moment to determine reliability of the CS-PFP test, on which the ADAP is based. Our data for the experienced tester (Tester 1) are consistent with their data. In a test-retest design, Cress et al found correlation coefficients ranging from .85 for upper-body flexibility to .97 for CS-PFP

total score. We found ICC values of .75 for upper-body flexibility and .96 for the ADAP total score.

We found that ADAP test scores correlated moderately with scores for the SF-36 physical health summary scale and physical functioning scale. ADAP scores were strongly correlated with knee extensor strength and TUG test scores, suggesting that the ADAP test is a good indicator of maximum physical performance. These findings are consistent with those of the validation study of the CS-PFP test of Cress and colleagues (11).

The ADAP test was patterned after the CS-PFP test because of its capacity to measure quantitatively, without ceiling effects, changes in performance that are expected in exercise interventions. The CS-PFP test is also sensitive to change in healthy, community-living older adults (18). In future research we intend to use the ADAP in descriptive and evaluation studies to determine the effect of a 12-week exercise programme on physical function in community-living older adults.

A limitation of the present reliability study is that only two testers were used to determine the reliability of the ADAP. Further, because the experienced and inexperienced examiners examined different samples of subjects, the difference in test-retest reliability between the two testers may not be necessarily caused by differences in the experience of the observers. More testers that examine the same sample of subjects should be used in future studies to evaluate the reliability of the ADAP and the influence of tester's experience. During recruitment, 16 potential participants withdrew after reading about the study. Often, the duration and physical load of the tests were mentioned as reasons for withdrawal, which suggests that the ADAP might be less suitable for testing fragile, older individuals. It would be of interest to examine the possibility to develop a short version of the ADAP test for testing fragile older people.

In conclusion, when administered by an experienced tester, ADAP is a reliable and valid instrument. Before the ADAP is used in research trials, it is recommended that testers gain experience in test administration and scoring. Further research is needed to evaluate the exact influence of tester experience and to determine how many test sessions are needed before a tester obtains reliable measurements.

ACKNOWLEDGEMENTS

This study was supported by Zorg Onderzoek Nederland-Medische Wetenschappen, a Dutch governmental organization supporting medical research.

REFERENCES

1. Singh NA, Clements KM, Fiatarone MA. A randomized controlled trial of progressive resistance training in depressed elders. J Gerontol A Biol Sci Med Sci 1997;52A:M27-35.
2. McMurdo ME, Johnstone R. A randomized controlled trial of a home exercise programme for elderly people with poor mobility. Age Ageing 1995;24:425-8.
3. Cohen ME, Marino RJ. The tools of disability outcomes research functional status measures. Arch Phys Med Rehabil 2000;81:S21-9.
4. Latham NK, Bennett DA, Stretton CM, Anderson CS. Systematic review of progressive resistance strength training in older adults. J Gerontol A Biol Sci Med Sci 2004;59A:48-61.
5. Chandler JM, Duncan PW, Kochersberger G, Studenski S. Is lower extremity strength gain associated with improvement in physical performance and disability in frail, community-dwelling elders? Arch Phys Med Rehabil 1998;79:24-30.
6. Buchner DM, Cress ME, de Lateur BJ, Esselman PC, Margherita AJ, Price R, et al. The effect of strength and endurance training on gait, balance, fall risk, and health service use in community-living older adults. J Gerontol A Biol Sci Med Sci 1997;52A:M218-24.
7. Ades PA, Ballor DA, Ashikaga T, Utton JL, Nair KS. Weight training improves walking endurance in healthy elderly persons. Ann Intern Med 1996;124:568-72.
8. Connelly DM, Vandervoort AA. Effects of detraining on knee extensor strength and functional mobility in a group of elderly women. J Orthop Sports Phys Ther 1997;26:340-6.
9. Rubenstein LZ, Josephson KR, Trueblood PR, Loy S, Harker JO, Pietruszka FM, et al. Effects of a group exercise program on strength, mobility, and falls among fall-prone elderly men. J Gerontol A Biol Sci Med Sci 2000;55A:M317-21.
10. Krebs DE, Jette AM, Assmann SF. Moderate exercise improves gait stability in disabled elders. Arch Phys med Rehabil 1998;79:1489-95.

11. Cress ME, Buchner DM, Questad KA, Esselman PC, deLateur BJ, Schwartz RS. Continuous-scale physical functional performance in healthy older adults: A validation study. Arch Phys Med Rehabil 1996;77:1243-50.
12. Fried LP, Ettinger WH, Lind B, Newman AB, Gardin J. Cardiovascular Health Study Group. Physical disability in older adults: a physiological approach. J Clin Epidemiol 1994;47:747-60.
13. Greig CA, Young A, Skelton DA, Pippet E, Butler FM, Mahmud SM. Exercise studies with elderly volunteers. Age Ageing 1994;23:185-9.
14. Keysor JJ, Jette AM. Have we oversold the benefit of late-life exercise? J Gerontol A Biol Sci Med Sci 2001;56A:M412-23.
15. Lord SR, Ward JA, Williams P, Strudwick M. The effect of a 12-month exercise trial on balance, strength, and falls in older women: A randomized controlled trial. J Am Geriatr Soc 1995;43:1198-1206.
16. De Vreede PL, Samson MM, van Meeteren NL, van der Bom JG, Duursma SA, Verhaar HJ. Functional Tasks Exercise versus Resistance Exercise to Improve Daily Function in Older Women: A Feasibility study. Arch Phys Med Rehabil 2004;85:1952-1961.
17. De Vreede PL, Samson MM, van Meeteren NL, Duursma SA, Verhaar HJ. Functional-Task exercise versus resistance strength exercise to improve daily function in older women: A randomized, controlled trial. J Am Geriatr Soc 2005;53:2-10.
18. Cress ME, Buchner DM, Questad KA, Esselman PC, deLateur BJ, Schwartz RS. Exercise: Effects on physical functional performance in independent older adults. J Gerontol A Biol Sci Med Sci 1999;54A:M242-8.
19. Duncan PW, Weiner DK, Chandler J et al. Functional reach: A new clinical measure of balance. Journal Gerontol A Biol Med Sci 1990;45:M192-M197.
20. Van Genderen FR, de Bie RA, Helders PJ, van Meeteren NL. Reliability research: towards a more clinically relevant approach. Phys Ther Rev 2003;8:169-76.
21. Ottenbacher KJ, Branch LG, Ray L, Gonzales VA, Peek MK, Hinman MR. The reliability of upper- and lower-extremity strength testing in a community survey of older adults. Arch Phys Med Rehabil 2002;83:1423-7.

22. Tanaka K, Kim HS, Yang JH, Shimamoto H, Kokudo S, Nishijima T. Index of assessing functional status in elderly Japanese men. Appl Human Sci. 1995;14:65-71.
23. Schenkman M, Morey M, Kuchibhatla M. Spinal flexibility and balance control among community-dwelling adults with and without Parkinson's disease. J Gerontol A Biol Sci Med Sci. 2000;55A:M441-5.
24. Aaronson NK, Muller M, Cohen PD, Essink-Bot M, Fekkes M, Sanderman R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. J Clin Epidemiol 1998;51:1055-68.
25. Podsiadlo D, Richardson S. The Timed 'Up & Go': A test of basic functional mobility for frail elderly persons. J Am Geriatric Soc 1991;39:142-8.
26. Skelton DA, McLaughlin AW. Training functional ability in old age. Physiotherapy. 1996;82:159-67.
27. Samson MM, Meeuwsen IB, Crowe A, Dessens JA, Duursma SA, Verhaar HJ. Relationships were observed between physical performance measures, age, height and body weight in healthy adults. Age Ageing 2000;29:235-42
28. Edwards RHT, Young A, Hosking GP, Jones DA. Human skeletal muscle function: description of tests and normal values. Clin Sci Mol Med 1977;55:283-90.
29. Skelton DA, Young A, Greig CA, Malbut KE. Effects of resistance training on strength, power, and selected functional abilities of women aged 75 and older. J Am Geriatr Soc 1995;43:1081-7.
30. Brown M, Sinacore DR, Ehsani AA, Binder EF, Holloszy JO, Kohrt WM. Low-intensity exercise as a modifier of physical frailty in older adults. Arch Phys Med Rehabil 2000;81:960-5.
31. Bassey EJ, Short AH. A new method for measuring power output in a single leg extension: feasibility, reliability and validity. Eur J Appl Physiol Occup Physiol 1990;60:385-90.
32. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;i:307-10.
33. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehabil 1998;12:187-99.

APPENDIX: The assessment of daily activity performance (ADAP) test

1. Carrying a weighted pan between kitchen counters
   Time score =
   (1/observed score – 1/8.33 sec) / (1/2.47 sec – 1/8.33 sec) x 100
   Weight score =
   (observed score – 1.4 kg) / (30.3 kg – 1.4 kg) x 100
2. Pouring water from a jug into a cup
   Time score =
   (1/observed score – 1/36.15 sec) / (1/6.8 sec – 1/36.15 sec) x 100
   Weight score =
   (observed score – 1.125 kg) / (4.5 kg – 1.125 kg) x 100
3. Carrying weight in a luggage bag up and down a 3-stair bus platform
   Time score =
   (1/observed score – 1/85.22 sec) / (1/11.75 sec – 1/85.22 sec) x 100
   Weight score =
   (observed score – 0.9 kg) / (30.6 kg – 0.9 kg) x 100
4. Carrying groceries through a door, up and down a 3-stair platform and lifting groceries on a counter.
   Time score =
   (1/observed score – 1/118.19 sec) / (1/33.15 sec – 1/118.19 sec) x 100
   Weight score =
   (observed score – 1.1 kg) / (27.69 kg – 1.1 kg) x 100
5. Transferring laundry from a washer to a dryer
   Time score =
   (1/observed score – 1/141.35 sec) / (1/21.31 sec – 1/141.35 sec) x 100
   Transferring laundry from a dryer to a counter
   Time score =
   (1/observed score – 1/113.06 sec) / (1/11.19 sec – 1/113.06 sec) x 100
6. Putting on and taking off a jacket
   Time score =
   (1/observed score – 1/39.76 sec) / (1/7.72 sec – 1/39.76 sec) x 100

7. Floor sweeping

   Time score =

   (1/observed score – 1/91.88 sec) / (1/18.78 sec – 1/91.88 sec) x 100

8. Vacuuming

   Time score =

   (1/observed score – 1/125.57 sec) / (1/19.34 sec – 1/125.57 sec) x 100

9. Making a bed

   Time score =

   (1/observed score – 1/151.41 sec) / (1/39.43 sec – 1/151.41 sec) x 100

10. Climbing stairs (13 steps)

    Time score =

    (1/(observed score/13) – 1/2.63 sec) / (1/0.32 sec – 1/2.63 sec) x 100

11. Getting down and up from the floor

    Time score =

    (1/observed score – 1/89.18 sec) / (1/3.53 sec – 1/89.18 sec) x 100

12. Opening a door

    Time score =

    (1/observed score – 1/11.94 sec) / (1/2.83 sec – 1/11.94 sec) x 100

13. Putting a hook-and-loop strap over a shoe

    Time score =

    (1/observed score – 1/17.15 sec) / (1/3.28 sec – 1/17.15 sec) x 100

14. Picking up four scarves from the floor

    Time score =

    (1/observed score – 1/36.09 sec) / (1/4.63 sec – 1/36.09 sec) x 100

15. 6-minute walk

    Distance score =

    (observed score m – 166 m) / (798 m – 166 m) x 100

16. Functional reach

    Distance score =

    ((observed score m / height m) – 0.033 m) / (0.294 m – 0.033 m) x 100

Allocation of task scores to ADAP domain scores

Tasks	***Upper-body strength***	***Upper-body flexibility***	***Lower-body strength***	***Balance & coordination***	***Endurance***
Weighted pan	Weight score			Time score	
Pouring water	Weight score			Time score	
Bus platform	Weight score		Weight score	Time score	
Groceries	Weight score		Weight score	Time score	
Laundry	Time scores		Time scores		
Jacket		Time score			
Floor sweeping			Time score	Time score	
Vacuuming			Time score	Time score	
Making a bed			Time score	Time score	
Climbing stairs			Time score		
Floor sit			Time score	Time score	
Opening a door	Time score				
Shoe strap		Time score			
Picking up scarves				Time score	
6-minute walk					Distance score
Functional reach		Distance score			
Total Time					Time score