ELSEVIER

CrossMark

Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 8 (2017) 26-35

Alzheimer's & Dementia

Diagnostic Assessment & Prognosis

Detecting functional decline from normal aging to dementia: Development and validation of a short version of the Amsterdam IADL Questionnaire

Roos J. Jutten[a,*], Carel F. W. Peeters[b], Sophie M. J. Leijdesdorff[c], Pieter Jelle Visser[a,d], Andrea B. Maier[e,f], Caroline B. Terwee[b], Philip Scheltens[a], Sietske A. M. Sikkes[a,b]

[a]*Alzheimer Center, Department of Neurology, VU University Medical Center, Amsterdam Neuroscience, Amsterdam, The Netherlands*
[b]*Department of Epidemiology & Biostatistics, Amsterdam Public Health Research Institute, VU University Medical Center, Amsterdam, The Netherlands*
[c]*Alzheimer Center Rotterdam, Erasmus Medical Center, Rotterdam, The Netherlands*
[d]*Alzheimer Center, School for Mental Health and Neuroscience, University Medical Centre Maastricht, Maastricht, The Netherlands*
[e]*MOVE Research Institute Amsterdam, Department of Human Movement Sciences, VU University of Amsterdam, Amsterdam, The Netherlands*
[f]*Department of Medicine and Aged Care, Royal Melbourne Hospital, University of Melbourne, Melbourne, Australia*

Abstract

Introduction: Detecting functional decline from normal aging to dementia is relevant for diagnostic and prognostic purposes. Therefore, the Amsterdam IADL Questionnaire (A-IADL-Q) was developed: a 70-item proxy-based tool with good psychometric properties. We aimed to design a short version while preserving its psychometric quality.

Methods: Study partners of subjects (n = 1355), ranging from cognitively normal to dementia subjects, completed the original A-IADL-Q. We selected the short version items using a stepwise procedure combining missing data, Item Response Theory, and input from respondents and experts. We investigated internal consistency of the short version and concordance with the original version. To assess its construct validity, we additionally investigated concordance between the short version and the Mini–Mental State Examination (MMSE) and Disability Assessment for Dementia (DAD). Finally, we investigated differences in instrumental activities of daily living (IADL) scores between diagnostic groups across the dementia spectrum.

Results: We selected 30 items covering the entire spectrum of IADL functioning. Internal consistency (0.98) and concordance with the original version (0.97) were very high. Concordance with the MMSE (0.72) and DAD (0.87) scores was high. IADL impairment scores increased across the spectrum from normal cognition to dementia.

Discussion: The A-IADL-Q short version (A-IADL-Q-SV) consists of 30 items and has maintained the psychometric quality of the original A-IADL-Q. As such, the A-IADL-Q-SV is a concise measure of functional decline.

Keywords: Alzheimer's disease; Dementia; Instrumental activities of daily living; Item Response Theory; Functional decline; Mild cognitive impairment; Subjective cognitive decline

1. Introduction

Dementia is a syndrome characterized by progressive cognitive decline and significant interference in daily function [1]. The first observable problems in daily life often concern the instrumental activities of daily living (IADL).

*Corresponding author. Tel.: +31-20-4448527; Fax: +31 20-4448529.
E-mail address: r.jutten@vumc.nl

http://dx.doi.org/10.1016/j.dadm.2017.03.002

IADL can be defined as "complex activities for which multiple cognitive processes are necessary," such as cooking, managing finances, and driving [2]. Detecting functional decline along the continuum from normal aging to dementia is highly relevant for a number of reasons. First of all, subtle IADL problems may already be present in subjects with mild cognitive impairment (MCI) and predict progression to dementia [3–5]. This suggests that assessment of IADL can be used to select MCI subjects at an increased risk for dementia [6]. Once a diagnosis has been established, measuring IADL performance remains essential for the monitoring of clinical progression [7]. Finally, IADL assessment plays a pivotal role in clinical trials, particularly in the evaluation of symptomatic treatment in dementia caused by Alzheimer's disease (AD) [8–10].

IADL performance is often measured using proxy-based questionnaires [11]. Unfortunately, most of these questionnaires suffer from serious limitations. They focus on everyday activities that are outdated and less relevant for patients in the early stages of dementia [12]. Furthermore, psychometric properties such as reliability, validity, and responsiveness are often questionable or overlooked [13]. Recent studies have pointed out that improvements in IADL instruments are necessary, especially for detecting IADL problems in MCI and the early stages of dementia [14–17].

To overcome the aforementioned drawbacks of existing IADL scales, Sikkes et al. developed the Amsterdam IADL Questionnaire (A-IADL-Q). The A-IADL-Q is a 70-item proxy-based tool and was developed with input from clinicians, patients, and caregivers [18]. Previous studies have reported good psychometric properties with respect to reliability, validity, responsiveness, and diagnostic accuracy in dementia [19–21]. One disadvantage of the A-IADL-Q is its length, resulting in an administration time of 20 to 25 minutes. In addition, respondents often report that some items are redundant or unclear. To facilitate its administration and implementation on a wider scale, we aimed to design a short and more concise version of the A-IADL-Q.

The present article describes the development and validation of a short version of the A-IADL-Q (A-IADL-Q-SV). We aimed to select the most informative items, using a combined approach of quantitative and qualitative methods. We expected that the short version would maintain the good psychometric quality of the original A-IADL-Q. In addition, we expected that IADL scores based on the short version would differ between diagnostic groups across the spectrum from normal cognition (NC) to dementia.

2. Methods

2.1. Subjects

We selected 1355 subjects with different levels of cognitive functioning, ranging from NC to dementia. Their study partner, mainly a spouse, relative, or friend, completed the A-IADL-Q. We included subjects from neurologic memory clinics of the VU University Medical Center (VUmc) Alzheimer Center, Amsterdam, The Netherlands ($n = 1117$), and the Alzheimer Center Rotterdam, The Netherlands ($n = 32$), and from the geriatric memory clinic of the VUmc, Amsterdam, The Netherlands ($n = 102$). All these subjects underwent a dementia assessment, including clinical history, medical and neurologic examination, screening laboratory tests, a neuropsychological test battery, and brain imaging [22]. During this visit, study partners completed the A-IADL-Q on an iPad. Subjects' diagnoses were made in a multidisciplinary diagnostic meeting, containing at least a neurologist or geriatrician [3,22,23].

We included cognitively normal subjects ($n = 104$) from the Amsterdam site of the preclinAD cohort of the European Medical Information Framework for AD project. Inclusion criteria for this cohort were age $\geq$60, modified telephone interview for cognitive screening $>$22; Geriatric Depression Scale $<$11; Consortium to Establish a Registry for AD 10-word list delayed recall >-1.5 standard deviation of age adjusted normative data; and Clinical Dementia Rating score of 0 with a score on the memory subdomain of 0 [24–27]. During the baseline visit, study partners completed the A-IADL-Q on an iPad.

Data were collected between October 2012 and August 2016. All subjects gave written informed consent and all study partners gave oral informed consent. The Medical Ethical Committee of the VUmc approved the study.

2.2. The Amsterdam IADL Questionnaire

The original A-IADL-Q is a proxy-based scale with 70 items covering a broad range of cognitive IADL [18]. The items can be divided into eight subcategories: household, administration, work, computer use, leisure time, appliances, transport, and other activities. The A-IADL-Q is computerized and has an adaptive approach as the items are tailored to individual responses (see Fig. 1). This results in a minimum of 47 and a maximum of 70 items for each respondent. Before the start, it is emphasized that the questionnaire addresses day-to-day problems caused by cognitive problems, such as memory, attention, or planning problems. Difficulty in performance is rated on a five-point Likert scale, ranging from "no difficulty in performing this task" to "no longer able to perform this task." Scoring is based on Item Response Theory (IRT): a paradigm linking responses to a test battery to an underlying construct (or latent trait) [28]. For the A-IADL-Q, the construct underlying the items can be termed "IADL performance," that is, the latent trait reflects IADL impairment with higher estimated trait levels indicating more impairment.

Linking the probabilities of category-specific item responses to latent trait levels is based on an IRT model [28]. For the A-IADL-Q, the graded response model (GRM) is used: a polytomous IRT model appropriate for items with ordinal response categories [29]. In the GRM, each item is characterized by a discrimination parameter (α) and four extremity parameters (βs; the number of response categories minus 1). The discrimination (or slope) parameter indicates how well an item discriminates between individuals with

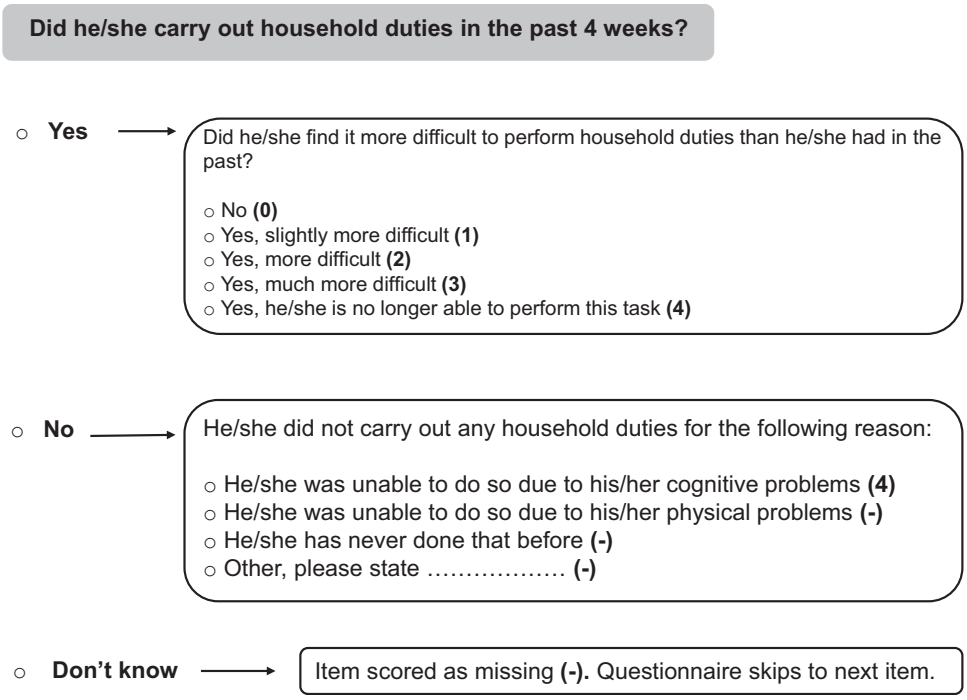

Fig. 1. Example item of the A-IADL-Q, including response options and scoring. Abbreviation: A-IADL-Q, Amsterdam IADL Questionnaire.

differing trait levels: higher discrimination parameters suggest higher ability to differentiate. The extremity (or category threshold) parameters represent the trait levels that mark the transition between response categories (in terms of cumulative probabilities for endorsement) [29]. An important advantage of IRT is that one's level of the latent trait can be estimated from any set of items for which the parameters are known. Therefore, IRT is able to handle missing data that may result from an adaptive approach. IRT is often preferred over classical scoring methods for scale development and refinement: it advances the development of more efficient scales by supporting item-reduction while preserving measurement precision [30,31].

The following basic assumptions underlie the IRT framework: (1) unidimensionality, implying that a single latent trait underlies the items; (2) local independence, which implies independence of item responses conditional on the latent trait; and (3) monotonicity, implying that the probability of endorsing (a category-specific response to) an item should increase as the trait level increases [32]. Previous work showed that the A-IADL-Q could be adequately described by a single latent factor and that the assumptions of local independence and monotonicity were met as well [19]. Because the present study contains a larger and more heterogenic sample, we have assessed these basic assumptions again.

2.3. Procedures

We divided the total sample into a training ($n = 677$) and validation set ($n = 678$), to use independent samples for the development and validation of the short version. We randomly split the Alzheimer Center Rotterdam, the VUmc geriatric, and the cognitively healthy cohorts. We conducted an alternative split procedure for the VUmc Alzheimer Center cohort ($n = 1117$), as a subsample ($n = 206$) of this cohort was used for the validation of the original A-IADL-Q. We therefore assigned this entire subsample to the current training set. From the remaining subjects ($n = 911$), we randomly assigned 35% to our training set and 65% to our validation set, to ensure that both sets had equal group sizes.

2.3.1. Development procedure

Item selection was performed in the training set, using a stepwise procedure that combined missing data, IRT, and content aspects. As shown in Fig. 1, a response is scored as missing when (1) the particular task has not been performed because of other reasons rather than cognitive problems, or (2) the study partner does not know whether the subject has performed that particular task in the past 4 weeks. Items with higher percentages of missing responses give us a less direct view of cognitive IADL, and are thus less applicable for our goal. We therefore eliminated items with more than 80% overall missing data. Items with more than 60% missing data in all diagnostic groups were candidates for elimination.

2.3.1.1. IRT analyses

We explored whether all items met the basic assumptions for IRT and eliminated items that did not meet these conditions. In the subsequent refitting rounds, we used IRT to identify items that contributed little unique information to the model, as reflected in either low item information values (an index representing the precision with which the trait is

measured) or overlapping item information curves (IICs; a mapping of the item information to the domain of the trait indicating how the information is distributed over the trait) with other items. After each elimination round, the GRM was refitted and an overall fit assessment was performed. This resulted in new item parameters and IICs that were used in the succeeding refitting round.

2.3.1.2. Content aspects

Comprehensibility was investigated in two ways: (1) by inspecting the comments that respondents provided in the "comment box" after completing the A-IADL-Q; and (2) performing thinking-out-loud interviews in a subsample of respondents (n = 17) while they were completing the A-IADL-Q. Items that were often commented as unclear or redundant, in either the comment box or interview, were candidates for removal. Furthermore, we investigated relevance and cultural applicability of all A-IADL-Q items with an online survey that we distributed among international experts. Between February 2016 and May 2016, we distributed the survey through contacts of the authors (R.J.J. and S.A.M.S.) via Qualtrics (www.qualtrics.com). All respondents (n = 33) were clinicians or researchers representing seven countries and had experience with the administration or cross-cultural validation of the A-IADL-Q. They were asked to rate the necessity of each original A-IADL-Q item for inclusion in the short version on a visual analogue scale ranging from 0 ("not necessary at all") to 100 ("very necessary").

2.3.2. Validation procedure

To confirm the quality of the final short version, we investigated missing data patterns, experts' ratings, adherence to IRT assumptions, and the overall fit of the short version items in the validation set. We subsequently investigated internal consistency of the short version and concordance between sum-scores derived from the short and original version. To assess construct validity [33], we investigated the relationship between the short version and measures of global cognition (Mini–Mental State Examination; MMSE [34]) and daily function (Disability Assessment for Dementia; DAD [35]), which were available for the VUmc Alzheimer Center cohort. Based on previous results [19], we expected moderate-to-high concordance between the short version and MMSE and DAD scores. To assess interpretability of the short version, we investigated differences in scores between six diagnostic groups that should represent different trait levels: (1) NC; (2) subjective cognitive decline (SCD); (3) MCI; (4) dementia caused by AD (AD dementia); (5) dementia other than AD (non-AD dementia); and (6) another neurologic or psychiatric disorder than dementia (Other).

2.4. Statistical analyses

Statistical analyses were performed using R and SPSS version 20.0 [36,37]. Statistical significance (for multiplicity corrections) was set at $P < .05$.

2.4.1. Development analyses

Item selection was partly based on IRT modeling. We used a GRM with a logit link function [29]. This model was fitted on the basis of approximate marginal maximum likelihood estimation [38]. The latent trait was assumed to follow a standard normal distribution. We assessed unidimensionality by performing an eigenvalue decomposition on the matrix of robust (Spearman) correlations between the items. A difference approximation to the second-order derivatives along the eigenvalue curve (scree plot) was calculated. This acceleration-approximation indicates points of abrupt change along the eigenvalue curve [39]. The number of eigenvalues before the point with the most abrupt change (the point with the maximum acceleration value) represents the number of latent dimensions that dominate the information content. Local independence was assessed by inspecting residual correlation matrices. We considered residual correlations >0.25 as indicative of problematic item pairs. We evaluated the monotonicity assumption using Mokken scale analysis [40]. Items that gave at least one significant violation of manifest monotonicity and had a crit value >30 were considered to violate latent monotonicity [41]. We assessed basic model fit by comparing nested models: we used a likelihood ratio test (LRT) to evaluate if the full GRM provided a better fit than a constrained GRM with equal slope parameters across items [28].

2.4.2. Validation analyses

We fitted a GRM on the final set of retained items. Estimation and assumption evaluation for this model were performed as described previously. This model was also compared with a constrained GRM as a means of basic model fit assessment. In addition, we evaluated global fitness of the final model with the comparative fit index (CFI) and root mean square error of approximation (RMSEA) [42]. Trait (or factor) scores were then based on empirical Bayes estimates: the mode of the posterior distribution of the trait given the retained items evaluated at the marginal maximum likelihood estimates. We calculated internal consistency of the retained items using a robust version of McDonald's omega [43]. We examined concordance between sum-scores derived from the short and original versions, and between short version sum-scores and MMSE and DAD scores, using Kendall's W [44]. To assess whether the short version scores differed between the diagnostic groups, we used a Kruskal-Wallis rank sum test on the trait scores followed by Dunn's pairwise test for multiple comparisons of mean rank sums (a nonparametric alternative to analysis of variance followed by post hoc tests) [45]. Multiple testing correction was based on the Bonferroni method.

3. Results

3.1. Sample and item characteristics

The study sample consisted of subjects with NC (n = 104), SCD (n = 219), MCI (n = 138), AD dementia

Table 1
Subject characteristics

	Total sample (n = 1355)	Training set (n = 677)	Validation set (n = 678)	P value
Age, M (SD)	65.7 (9.7)	66.1 (10.1)	65.3 (9.2)	.146*
Gender, female (%)	602 (44.4%)	301 (44.5%)	301 (44.4%)	.981†
Diagnosis				
NC	104 (7.7%)	52 (7.7%)	52 (7.6%)	
SCD	219 (16.2%)	116 (17.1%)	103 (15.2%)	
MCI	138 (10.2%)	84 (12.4%)	54 (8.0%)	
AD dementia	413 (30.5%)	209 (30.9%)	204 (30.1%)	
Non-AD dementia	235 (17.3%)	100 (14.8%)	135 (19.9%)	
Other	246 (18.2%)	116 (17.1%)	130 (19.2%)	

Abbreviations: AD, Alzheimer's disease; MCI, mild cognitive impairment; NC, normal cognition; SCD, subjective cognitive decline; SD, standard deviation.

*Tested using independent t test.

†Tested using Pearson's chi-square test.

(n = 413), non-AD dementia (n = 235), and 246 subjects with other diagnoses. Table 1 shows subject characteristics for the total sample and for the training and validation set separately. There were no age and gender differences between the two sets. The MCI group was slightly larger in the training set, whereas the non-AD dementia group was slightly larger in the validation set.

Missing responses on the item level in the training set ranged from 10.5% ("preparing sandwich meals") to 92.8% ("programming a video recorder"). Approximately half of the original version items (36/70) contained more than 50% missing data. Mean ratings from the 33 experts ranged from 23.9 ("programming a video recorder") to 86.9 ("paying when doing the shopping"), with an overall mean score of 62.3 (standard deviation = 14.9).

3.2. Development of the short version

Fig. 2 provides a flowchart of the item selection procedure. Our first step included the removal of two items that violated the assumption of monotonicity, together with seven items that contained more than 80% missing data. After the second round, we removed 11 items with missing responses more than 60% in all diagnostic groups and contributing little information to the model (item information <3.0). After the third round, we removed eight items that received low ratings of experts (mean rating <50) and had often been commented on as either unclear or redundant by respondents. We thereafter removed six items with overlapping IICs and overlapping content with other items within the same activity category (e.g., "cooking" versus "preparing hot meals"). Of these overlapping pairs, we removed the one containing higher missing data and lower content rating. After the fourth round, we removed four items that were often perceived as unclear and showed overlapping IICs with more specific

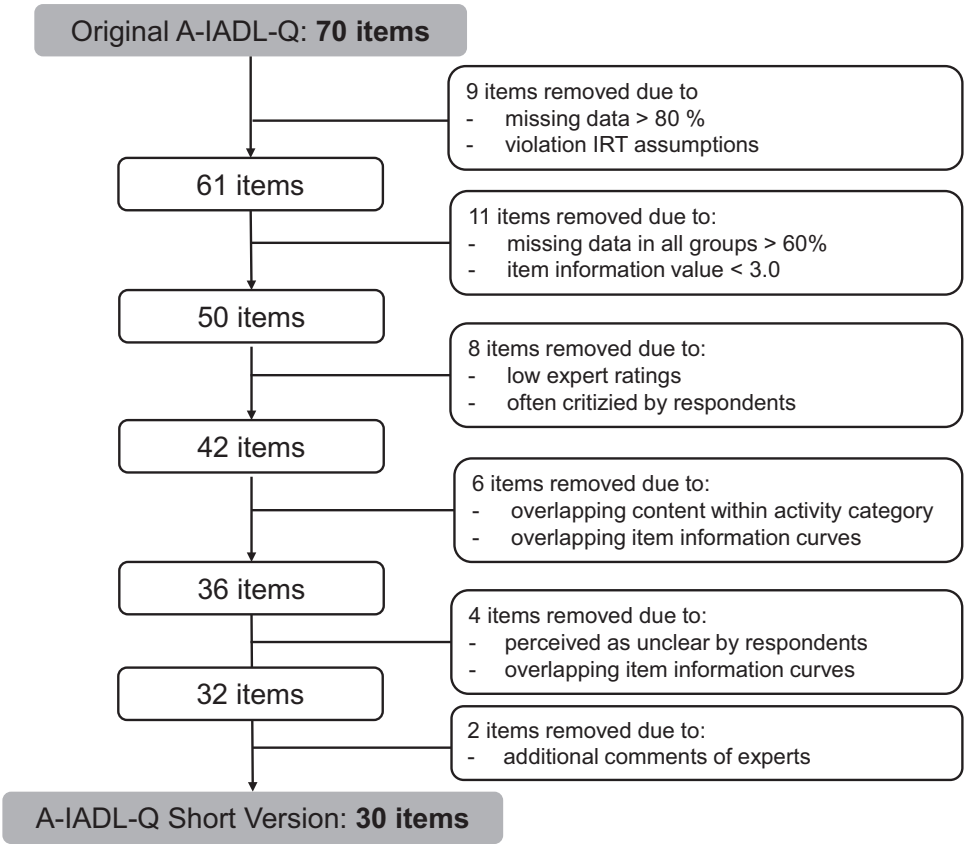

Fig. 2. Flowchart of the item selection procedure that led to the short version of the Amsterdam IADL Questionnaire (A-IADL-Q). Abbreviations: IADL, Instrumental Activities of Daily Living; IRT, Item Response Theory.

Table 2
Final selection of the short version items, including their missing data percentages, GRM parameters, item information values, and content ratings by experts

Item		% Missing	Item parameters					Item information	Expert rating
			α	β_1	β_2	β_3	β_4		
1	Carrying out household duties	12.1	1.830	−0.653	0.259	1.309	2.209	4.73	76
2	Doing the shopping	14.3	2.027	−0.597	0.481	1.206	1.732	4.93	79
4	Buying the correct articles	34.8	1.505	−0.209	0.795	1.249	1.282	2.57	73
6	Cooking	33.8	2.236	−0.613	0.426	0.990	1.388	5.25	76
9	Preparing sandwich meals	10.0	2.454	0.643	1.602	2.266	2.524	5.82	60
10	Making minor repairs to the house	53.1	2.330	−0.867	−0.013	0.510	0.897	5.28	60
11	Operating domestic appliances	8.3	2.120	−0.074	0.865	1.544	2.142	5.17	63
12	Operating the microwave oven	30.4	1.938	0.134	0.880	1.351	1.711	3.83	58
16	Operating the coffee maker	11.4	2.751	0.598	1.302	1.731	1.913	5.82	63
17	Operating the washing machine	42.0	3.662	0.584	1.254	1.533	1.610	7.71	58
19	Paying bills	34.4	2.659	−0.467	0.417	0.717	0.904	5.55	83
22	Using a mobile phone	17.0	2.126	−0.416	0.531	1.148	1.658	5	76
23	Managing the household budget	45.0	2.884	−0.773	0.138	0.448	0.738	6.56	79
25	Using electronic banking	46.9	3.632	−0.320	0.285	0.500	0.560	7.12	66
28	Using a pin code	11.8	2.030	0.190	1.082	1.533	1.984	4.34	77
29	Obtaining money from a cash machine	32.7	3.486	0.397	0.892	1.316	1.484	7.59	69
30	Paying using cash	15.6	2.665	0.415	1.288	1.712	2.257	6.59	72
31	Making appointments	17.4	1.945	−0.660	0.270	1.363	1.945	4.33	75
32	Filling in forms	24.8	2.516	−0.792	0.173	0.754	1.307	5.95	66
33	Working	47.1	1.579	−0.965	−0.237	0.379	0.742	2.95	70
35	Using a computer	22.6	2.229	−0.718	0.193	0.846	1.418	5.47	68
37	Emailing	44.0	3.100	−0.297	0.293	0.789	1.024	7.01	54
39	Printing documents	56.6	4.080	−0.059	0.653	0.814	0.857	8.3	70
46	Operating devices	16.4	3.374	−0.367	0.584	1.243	1.869	10.3	72
47	Operating the remote control	3.1	1.786	0.017	1.073	1.734	2.599	4.32	80
57	Playing card and board games	50.9	1.657	−0.462	0.556	1.194	1.662	3.49	62
59	Driving a car	25.7	1.592	−0.351	0.569	1.015	1.300	2.93	76
65	Using a sat-nav system	51.6	2.421	−0.394	0.384	0.777	0.823	4.62	61
66	Using public transport	47.9	3.123	−0.071	0.420	0.878	1.200	7.01	83
70	Being responsible for his/her own medication	27.7	1.478	−0.075	1.005	1.685	2.361	3.16	82

Abbreviations: GRM, Graded Response Model; α, discrimination parameter; β's, extremity parameters.

NOTE. Percentage missing, parameter estimates, and information characteristics are based on the validation set. Expert ratings were made per item on a visual analogue scale ranging from 0 to 100.

items (e.g., "looking for important things at home" versus "looking for his/her keys"). Finally, we removed two items because of disputable item characteristics and additional comments of experts. After this, we refitted the model with the remaining 30 items and concluded that further shortening was unnecessary. All 30 retained items in the training set were deemed to contribute substantially unique information to the latent trait. The full GRM improved fit on the constrained GRM (*LRT value* = 98.01, *df* = 29, $P < .001$).

3.3. Validation of the short version

The final selection of items can be found in column 1 of Table 2. This selection adheres to all assumptions underlying the IRT framework. The maximum acceleration factor on the consecutive eigenvalues of the robust correlation matrix occurs at the second eigenvalue (with a value of 1.26), implying that the first eigenvalue (with a value of 17.09) dominates the information content. Hence, a single latent dimension is sufficient. Moreover, no item pair sorted a residual correlation >0.25 and no item displayed significant violations of manifest monotonicity. Table 2 also presents information on missing percentages and the estimated GRM parameters based on the validation set. The last column shows the experts' ratings. As can be seen, all retained items contain less than 60% missing data and most items (26/30) had less than 50% missing data in the validation set. The extremity parameters were spread along the latent trait continuum (ranging from −4 to +4), which is also illustrated by the IICs presented in Fig. 3. For most items, item information values were more than 3 (on a total information of 163.68). Finally, all short version items received medium-to-high ratings from experts.

The full GRM provided better fit on the validation data than the constrained GRM (*LRT value* = 6644.47, *df* = 29, $P < .001$). The overall fit of the final model was considered good: CFI = 0.994, RMSEA = 0.032. Internal consistency of the short version was very high (robust McDonald's omega = 0.98). Concordance between the item sum-scores of the short version and the original version was also very high (Kendall's $W = 0.97$). Concordance with the MMSE (Kendall's $W = 0.72$) and DAD (Kendall's $W = 0.87$) was high.

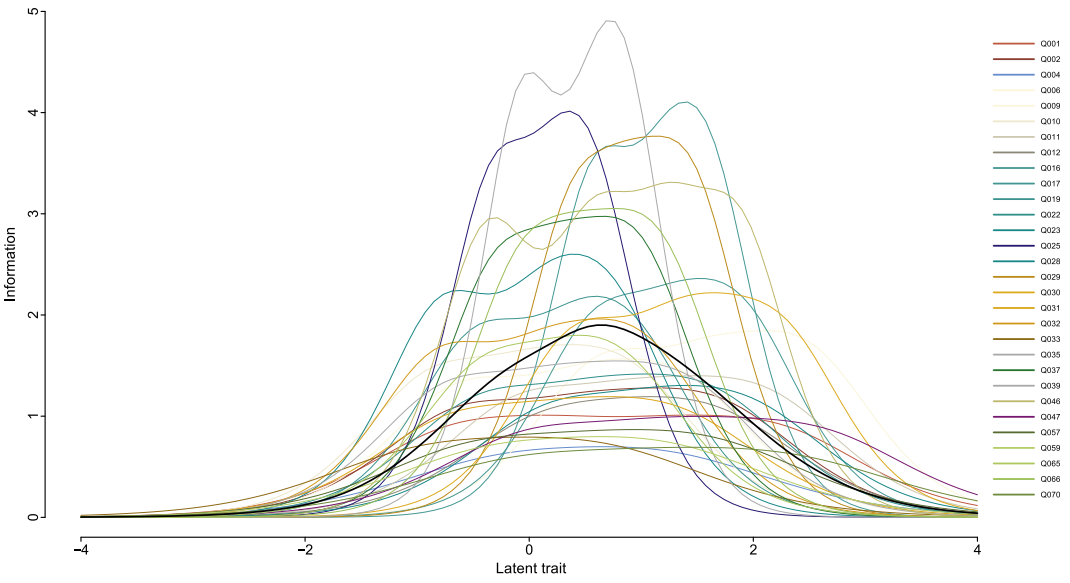

Fig. 3. Item information curves of the 30 Amsterdam IADL items that resulted in the short version. The bold black line represents the total test information curve. The latent trait ranges from −4 (good IADL functioning) to +4 (poor IADL functioning). Abbreviation: IADL, instrumental activities of daily living.

Table 3 presents the clinical characteristics of the different diagnostic groups within the validation set. Fig. 4 represents the trait score distributions for each diagnostic group. It can be seen that this score seems to increase from NC to dementia. The variances of the trait scores were not equal between diagnostic groups. Hence, a nonparametric test was used to assess diagnostic group differences between latent trait scores as derived from the final GRM. The Kruskal-Wallis rank sum test indicated that the mean trait score ranks of the diagnosis groups indeed differed ($\chi^2 = 187.01$, $df = 5$, $P < .001$). Pairwise comparisons (Dunn's test) with the Bonferroni correction indicated the following pairwise differences: (1) NC versus all other groups (all corrected P values $<.001$); (2) SCD versus AD dementia, non-AD dementia, and Other group (all corrected P values $<.001$); and (3) MCI versus AD dementia (corrected P value $= .002$).

4. Discussion

We designed a short version of the A-IADL-Q containing 30 items. We thereby reduced administration time by approximately 10 minutes. We showed that, although significantly shorter, the A-IADL-Q-SV has maintained the psychometric quality of the original A-IADL-Q. We demonstrated adequate measurement precision along the entire spectrum of IADL functioning. Short version scores were in high concordance with the MMSE and DAD, which supports the construct validity of the A-IADL-Q-SV. We also found that the A-IADL-Q-SV could differentiate between various diagnostic groups with respect to IADL impairment.

The present study expands on previous work on the A-IADL-Q, which already demonstrated good psychometric quality of the scale [18–21]. The A-IADL-Q-SV contains only the most informative items, and thereby possible "noise" caused by less informative or ambiguous items has been reduced. Because of its reduced length, the A-IADL-Q-SV may be perceived as a more user-friendly measure. The use of shorter tests is also encouraged from a psychometric point of view: a short form containing items of the same quality as the original form may yield less measurement error and thus be more reliable [28]. Longer tests are more likely to suffer from acquiescence bias and missing responses. Using the A-IADL-Q-SV may overcome these test-length related drawbacks.

Our findings suggest that the A-IADL-Q-SV can already detect IADL problems in subjects with SCD and MCI, which

Table 3
Clinical characteristics of different diagnostic groups in the validation set

	NC ($n = 52$)	SCD ($n = 103$)	MCI ($n = 54$)	AD dementia ($n = 204$)	Non-AD dementia ($n = 135$)	Other ($n = 130$)
Gender, female (%)	30 (57.7%)	43 (41.7%)	18 (33.3%)	111 (54.4%)	57 (42.2%)	42 (32.3%)
Age, M (SD)	70.5 (7.6)	62.8 (10.2)	69.8 (9.3)	66.9 (8.9)	66.5 (8.5)	60.8 (9.9)
MMSE score, M (SD)	NA	27.3 (2.2)	26.6 (2.1)	20.1 (4.8)	23.7 (4.4)	25.3 (4.3)
DAD score, M (SD)	NA	91.0 (12.6)	86.5 (14.3)	78.2 (20.6)	75.7 (25)	76.6 (23.7)

Abbreviations: AD, Alzheimer's disease; DAD, Disability Assessment for Dementia (lower scores reflect more dysfunction in activities of daily living); MCI, mild cognitive impairment; MMSE, Mini–Mental State Examination (higher scores reflect better cognitive functioning); NA, not available for this cohort; NC, normal cognition; SCD, subjective cognitive decline; SD, standard deviation.

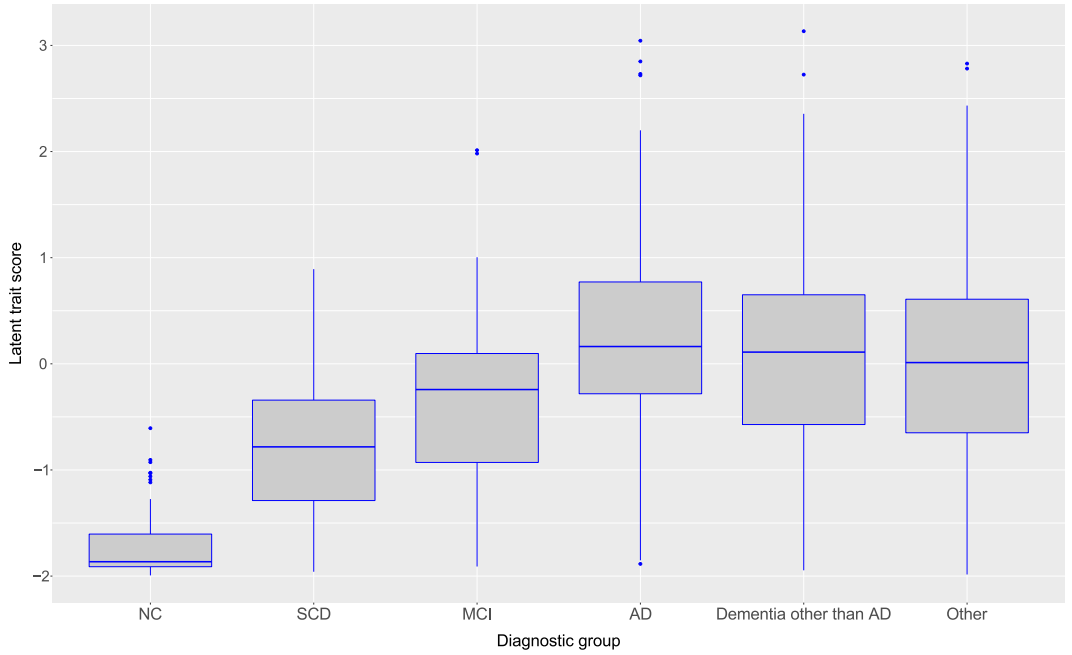

Fig. 4. Short version latent trait scores for each diagnostic group. Latent trait scores reflect IADL functioning, with higher scores indicating poorer IADL functioning. Post hoc analyses gave the following significant pairwise differences: (1) NC versus all other groups; (2) SCD versus AD dementia, non-AD dementia, and Other group; and (3) MCI versus AD dementia. Abbreviations: AD, Alzheimer's disease; IADL, instrumental activities of daily living; MCI, mild cognitive impairment; NC, normal cognition; SCD, subjective cognitive decline.

is in line with previous studies that report subtle functional impairment in these groups [46–48]. We found that IADL scores differed between subjects with NC and SCD. This is of particular interest because both groups are characterized by the absence of objective cognitive impairment, although SCD subjects may be at higher risk of developing dementia [49]. The A-IADL-Q-SV might thus be able to detect subtle functional decline that appears in preclinical stages of dementia, suggesting that it could be a promising measure for clinical trials in these earliest stages [10,50].

Strengths of this study include our large and heterogenic sample with subjects covering a broad range of the IADL spectrum along the continuum from normal aging to dementia. Another strength is the use of a validation set to replicate findings derived from the training set. After splitting the total sample, the training and validation set both contained more than 500 subjects, a number that is recommended for estimating accurate parameters based on the GRM [51]. Finally, combining statistical methods with input from respondents and experts is an important strength of this study, as it preserved both the psychometric quality and clinical relevance of the A-IADL-Q-SV.

There are some limitations that should be considered. Among them are our relatively small NC group, because of the fact that most subjects were recruited via memory clinics. Secondly, previous studies have shown that proxy-based IADL measures may be confounded by respondent characteristics such as caregiver burden and depression [52]. We did not take these characteristics into account in the present study. However, Sikkes et al. showed low correlations between the original A-IADL-Q, caregiver burden, and depression, indicating limited confounding by these variables [19].

Further research is needed to examine whether the A-IADL-Q-SV is sensitive to changes over time *within* subjects. We will investigate the A-IADL-Q-SV longitudinally in subjects with MCI and early dementia to determine whether it could be an effective measure for monitoring disease progression and evaluating disease-modifying therapies. Because the research field is shifting toward preclinical stages of dementia, it is also relevant to further investigate the A-IADL-Q-SV in subjects with SCD and the relation between IADL scores and dementia biomarkers in this group.

To conclude, we developed the A-IADL-Q-SV, which is a concise instrument to efficiently measure functional decline in the early stages of dementia. The A-IADL-Q-SV has retained the good qualities of the original A-IADL-Q; hence, we expect the A-IADL-Q-SV to be a promising outcome measure for daily function in dementia research and clinical practice.

Acknowledgments

The authors would like to thank all respondents and experts for their willingness to participate in this study. We also would like to thank Naomi Koster, Saskia de Vries, Judith Meurs, Iris Dalhuizen, and Tarik Binnenkade for their help with the data collection. The development of the Amsterdam IADL Questionnaire is supported by grants from Stichting VUmc Fonds and Innovatiefonds Zorgverzekeraars. The present study is supported by a grant from Memorabel (grant no. 733050205), which is the research program of the Dutch

Deltaplan for Dementia. This work has received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking (grant no. 115372). Part of this article has been presented at the 2016 Alzheimer's Association International Conference.

The Amsterdam IADL Questionnaire is free for use in all public health and not-for-profit agencies and can be obtained via https://www.alzheimercentrum.nl/professionals/amsterdam-iadl.

Conflicts of interest: R.J.J., C.F.W.P, S.M.J.L, P.J.V, A.B.M, and C.B.T. report no relevant conflicts of interest. P.S. has acquired grant support (for the institution; VUmc Alzheimer Center) from GE Healthcare, Danone Research, Piramal, and MERCK. In the past 2 years he has received consultancy/speaker fees (paid to the institution) from Lilly, GE Healthcare, Novartis, Sanofi, Nutricia, Probiodrug, Biogen, Roche, Avraham, and EIP Pharma. S.A.M.S. is supported by grants from JPND and Zon-MW, and has provided consultancy services in the past 2 years for Nutricia and Takeda. All funds were paid to her institution.

RESEARCH IN CONTEXT

1. Systematic review: The Amsterdam IADL Questionnaire (A-IADL-Q) was developed to detect problems in instrumental activities of daily living (IADL) in subjects with incipient dementia. Previous validation studies have demonstrated good psychometric properties. However, a shorter version is desired to facilitate administration and implementation on a wider scale. We therefore aimed to design and validate a short version of the A-IADL-Q (A-IADL-Q-SV).
2. Interpretation: We developed an A-IADL-Q-SV that consists of 30 items, resulting in an administration time of 10 to 15 minutes. The A-IADL-Q-SV has maintained the psychometric quality of the original version. We demonstrated adequate measurement precision along the entire spectrum of IADL functioning from normal aging to dementia.
3. Future directions: The A-IADL-Q-SV will be investigated longitudinally in subjects with early stages of dementia, to determine whether it could be an effective measure for the monitoring of disease progression and evaluation of symptomatic treatments.

References

[1] American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.

[2] Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969; 9:179–86.

[3] Albert MS, Dekosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:270–9.

[4] Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild cognitive impairment: a concept in evolution. J Intern Med 2014;275:214–28.

[5] Luck T, Luppa M, Wiese B, Maier W, van den Bussche H, Eisele M, et al. Prediction of incident dementia: impact of impairment in instrumental activities of daily living and mild cognitive impairment—results from the German study on ageing, cognition, and dementia in primary care patients. Am J Geriatr Psychiatry 2012;20:943–54.

[6] Tabert MH, Albert SM, Borukhova-Milov L, Camacho Y, Pelton G, Liu X, et al. Functional deficits in patients with mild cognitive impairment: prediction of AD. Neurology 2002;58:758–64.

[7] Rockwood K. The measuring, meaning and importance of activities of daily living (ADLs) as an outcome. Int Psychogeriatr 2007;19:467–82.

[8] Woodcock J, Sharfstein JM, Hamburg M. Regulatory action on rosiglitazone by the US Food and Drug Administration. N Engl J Med 2010; 363:1489–91.

[9] Vellas B, Andrieu S, Sampaio C, Coley N, Wilcock G. Endpoints for trials in Alzheimer's disease: a European task force consensus. Lancet Neurol 2008;7:436–50.

[10] Vellas B, Bateman R, Blennow K, Frisoni G, Johnson K, Katz R, et al. Endpoints for Pre-Dementia AD Trials: a report from the EU/US/CTAD Task Force. J Prev Alzheimers Dis 2015;2:128–35.

[11] Marshall GA, Amariglio RE, Sperling RA, Rentz DM. Activities of daily living: where do they fit in the diagnosis of Alzheimer's disease? Neurodegener Dis Manag 2012;2:483–91.

[12] Rosenberg L, Kottorp A, Winblad B, Nygård L. Perceived difficulty in everyday technology use among older adults with or without cognitive deficits. Scand J Occup Ther 2009;16:216–26.

[13] Sikkes SA, de Lange-de Klerk ES, Pijnenburg YA, Scheltens P, Uitdehaag BM. A systematic review of Instrumental Activities of Daily Living scales in dementia: room for improvement. J Neurol Neurosurg Psychiatry 2009;80:7–12.

[14] Law LL, Barnett F, Yau MK, Gray MA. Measures of everyday competence in older adults with cognitive impairment: a systematic review. Age Ageing 2012;41:9–16.

[15] Gold DA. An examination of instrumental activities of daily living assessment in older adults and mild cognitive impairment. J Clin Exp Neuropsychol 2012;34:11–34.

[16] Jekel K, Damian M, Wattmo C, Hausner L, Bullock R, Connelly PJ, et al. Mild cognitive impairment and deficits in instrumental activities of daily living: a systematic review. Alzheimers Res Ther 2015;7:17.

[17] Kaur N, Belchior P, Gelinas I, Bier N. Critical appraisal of questionnaires to assess functional impairment in individuals with mild cognitive impairment. Int Psychogeriatr 2016;28:1425–39.

[18] Sikkes SA, de Lange-de Klerk ES, Pijnenburg YA, Gillissen F, Romkes R, Knol DL, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. Alzheimers Dement 2012;8:536–43.

[19] Sikkes SA, Knol DL, Pijnenburg YA, de Lange-de Klerk ES, Uitdehaag BM, Scheltens P. Validation of the Amsterdam IADL Questionnaire(c), a new tool to measure instrumental activities of daily living in dementia. Neuroepidemiology 2013;41:35–41.

[20] Sikkes SA, Pijnenburg YA, Knol DL, de Lange-de Klerk ES, Scheltens P, Uitdehaag BM. Assessment of instrumental activities of daily living in dementia: diagnostic value of the Amsterdam Instrumental Activities of Daily Living Questionnaire. J Geriatr Psychiatry Neurol 2013;26:244–50.

[21] Koster N, Knol DL, Uitdehaag BM, Scheltens P, Sikkes SA. The sensitivity to change over time of the Amsterdam IADL Questionnaire©. Alzheimers Dement 2015;11:1231–40.
[22] van der Flier WM, Pijnenburg YA, Prins N, Lemstra AW, Bouwman FH, Teunissen CE, et al. Optimizing patient care and research: the Amsterdam Dementia Cohort. J Alzheimers Dis 2014; 41:313–27.
[23] Mckhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:263–9.
[24] Graff-Radford NR, Ferman TJ, Lucas JA, Johnson HK, Parfitt FC, Heckman MG, et al. A cost effective method of identifying and recruiting persons over 80 free of dementia or mild cognitive impairment. Alzheimer Dis Assoc Disord 2006;20:101–4.
[25] Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1983;17:37–49.
[26] Welsh KA, Butters N, Mohs RC, Beekly D, Edland S, Fillenbaum G, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part V. A normative study of the neuropsychological battery. Neurology 1994;44:609.
[27] Hughes CP, Berg L, Danziger WL, Coben LA, Martin R. A new clinical scale for the staging of dementia. Br J Psychiatry 1982; 140:566–72.
[28] Embretson SE, Reise SP. Item Response Theory for Psychologists. New York, NB: Psychology Press; 2013.
[29] Samejima F. Estimation of latent ability using a response pattern of graded scores. Psychometrika 1970;35:139.
[30] Reise SP, Waller NG. Item response theory and clinical measurement. Annu Rev Clin Psychol 2009;5:27–48.
[31] Chang CH, Reeve BB. Item response theory and its applications to patient-reported outcomes measurement. Eval Health Prof 2005; 28:264–82.
[32] Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. Qual Life Res 2007;16:5–18.
[33] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010; 63:737–45.
[34] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–98.
[35] Gelinas I, Gauthier L, McIntyre M, Gauthier S. Development of a functional measure for persons with Alzheimer's disease: the disability assessment for dementia. Am J Occup Ther 1999;53:471–81.
[36] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
[37] IBM. IBM SPSS Statistics for Windows, version 20.0. New York: IBM Corp; 2011.
[38] Rizopoulos D. ltm: an R package for latent variable modeling and item response theory analyses. J Stat Softw 2006;17:1–25.
[39] Raîche G, Walls TA, Magis D, Riopel M, Blais JG. Non-graphical solutions for Cattell's Scree Test. Methodology 2013;9:23–9.
[40] van der Ark LA. Mokken scale analysis in R. J Stat Softw 2007; 20:183–208.
[41] van Schuur W. Ordinal Item Response Theory: Mokken Scale Analysis, vol. 169. Los Angeles: Sage Publications; 2011.
[42] Maydeu-Olivares A. Goodness-of-fit assessment of item response theory models. Meas Interdiscip Res Perspect 2013;11:71–101.
[43] Zhang Z, Yuan KH. Robust coefficients alpha and omega and confidence intervals with outlying observations and missing data: methods and software. Educ Psychol Meas 2016;76:387–411.
[44] Kendall MG, Babington B. The problem of m rankings. Ann Math Stat 2016;10:275–87.
[45] Dunn OJ. Multiple comparisons using rank sums. Technometrics 1964;6:241–52.
[46] Farias ST, Mungas D, Reed BR, Harvey D, Cahn-Weiner D, Decarli C. MCI is associated with deficits in everyday functioning. Alzheimer Dis Assoc Disord 2006;20:217–23.
[47] Pedrosa H, De Sa A, Guerreiro M, Maroco J, Simoes MR, Galasko D, et al. Functional evaluation distinguishes MCI patients from healthy elderly people—the ADCS/MCI/ADL scale. J Nutr Health Aging 2010;14:703–9.
[48] Teng E, Becker BW, Woo E, Cummings JL, Lu PH. Subtle deficits in instrumental activities of daily living in subtypes of mild cognitive impairment. Dement Geriatr Cogn Disord 2010;30:189–97.
[49] Jessen F, Amariglio RE, Van Boxtel M, Breteler M, Ceccaldi M, Chételat G, et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimers Dement 2014;10:844–52.
[50] Snyder PJ, Kahle-Wrobleski K, Brannan S, Miller DS, Schindler RJ, Desanti S, et al. Assessing cognition and function in Alzheimer's disease clinical trials: do we have the right tools? Alzheimers Dement 2014;10:853–60.
[51] Tsutakawa RK, Johnson JC. The effect of uncertainty of item parameter estimation on ability estimates. Psychometrika 1990;55:371–90.
[52] Jorm A. Complaints of cognitive decline in the elderly: a comparison of reports by subjects and informants in a community survey. Psychol Med 1997;24:365–74.