For good measure

Dubbelman, Mark Anton

2022

document version
Publisher's PDF, also known as Version of record

FOR GOOD MEASURE:

Detecting and defining changes in everyday functioning in Alzheimer's disease and related disorders

Mark Anton Dubbelman

The studies described in this thesis were carried out at the Alzheimer Center Amsterdam, Amsterdam UMC, location VUmc, and embedded in the neurodegeneration research program of Amsterdam Neuroscience. The Alzheimer Center Amsterdam is supported by Alzheimer Nederland and Stichting VUmc fonds.

Cover design: Lisa den Dulk | steaze.nl
Layout by: Lisa den Dulk & Publiss | publiss.nl
Printing: Ridderprint | ridderprint.nl
ISBN: 978-94-6458-594-0

Printing of this thesis has been accomplished with gratefully acknowledged financial support from Alzheimer Nederland, Stichting Alzheimer & Neuropsychiatrie and Vrije Universiteit Amsterdam.

VRIJE UNIVERSITEIT

FOR GOOD MEASURE:

Detecting and defining changes in everyday functioning in Alzheimer's disease and related disorders

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor
aan de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op 15 november 2022 om 15.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door
Mark Anton Dubbelman
geboren te Utrecht

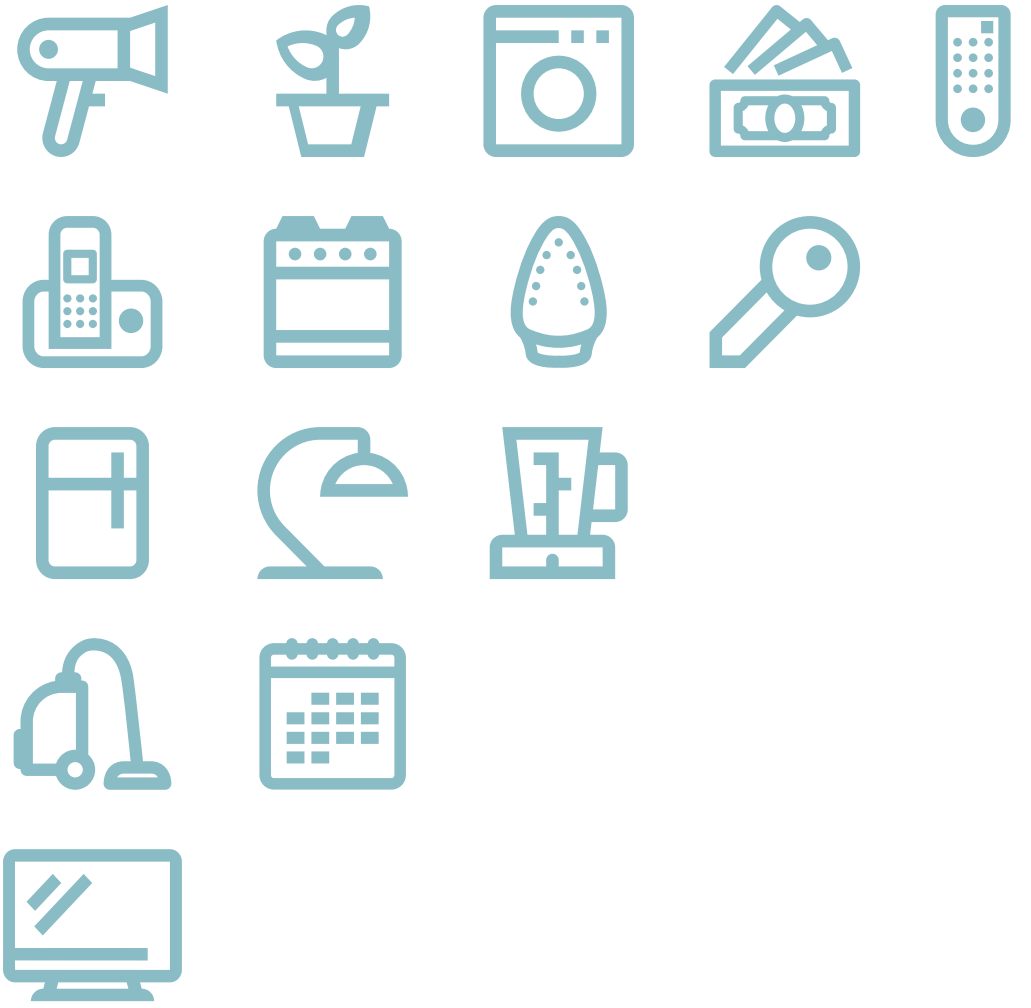
promotor:	prof.dr. P. Scheltens
copromotor:	dr. S.A.M. Sikkes
promotiecommissie:	prof.dr. Y.A.L. Pijnenburg prof.dr. B.M.J. Uitdehaag prof.dr. H.E. Hulst dr. L.B. Mokkink prof.dr. D.M. Rentz prof.dr. S.E. Tomaszewski Farias

“Valeu a pena?
Tudo vale a pena se a alma não é pequena”
— Fernando Pessoa, ‘Mar Português’ (1934)

Table of Contents

“(...) as long as the answer is right, who cares if the question is wrong?”

— Norton Juster, ‘The Phantom Tollbooth’ (1961)



CHAPTER 1

INTRODUCTION

1.1 General background

Neurodegenerative diseases are characterized by the progressive loss of structure and function of the brain, triggering the deterioration of cognitive functions. Eventually, cognitive impairments affect a person's ability to function independently in daily life to the point that they become dependent on others. This constitutes the dementia syndrome.[1] In 2021, approximately 55 million people worldwide had dementia.[2] With an incidence of nearly 10 million cases each year, the World Health Organization has identified dementia as a public health priority.[2] Increasing dementia-related healthcare costs place a burden on society as a whole. More importantly, because people who have dementia cannot live alone, dementia burdens not only the patients themselves, but also their caregivers and loved ones.

Before the onset of dementia, subtle cognitive symptoms emerge and increase over the years. Alzheimer's disease and related disorders can be categorized clinically into different stages, based on symptom severity. The 'preclinical stage' comprises the time when there are no evident clinical signs of the disease yet,[3] and is presumed to last about ten years.[4] An individual in the late preclinical stage may self-report to have impairments, but these complaints cannot be objectively measured using standard neuropsychological testing. This is referred to as 'subjective cognitive decline'.[5] As time passes, the disease moves into a 'prodromal stage', characterized initially by subtle cognitive impairment. This stage is also referred to as 'mild cognitive impairment'.[6] The point at which cognitive impairments begin to interfere significantly with everyday life is traditionally considered the start of dementia. The transition from one clinical stage to the next is predicated on changes in cognition.

Cognition in everyday life

Cognition covers a variety of brain functions, including attention, memory, language, and executive functioning.[7] These functions are also called cognitive domains. People employ the various cognitive domains in their everyday functioning. For example, when doing groceries, one uses their memory to remember what to buy and where the grocery store is. They use executive functions to plan a route to the supermarket and safely drive over. They require language and visual perception to read the product labels and use social cognition to greet the person at the register. Ergo, cognition is an integral part of daily life, and all cognitive domains are called upon—by various extents—for everyday functioning.

While commonly used pencil-and-paper tests have been shown to discriminate well between cognitively normal individuals and people with dementia, they often lack sensitivity to more subtle inter- and intra-individual differences in cognitive performance.[8,9] This renders them less valuable in early disease stages, both for finding group differences and for capturing early disease progression. Another important limitation of traditional tests, is that they have limited ecological validity—that is, performance on a test does not translate well to everyday life, nor can everyday cognitive functioning always be adequately captured by a neuropsychological test.[10]

Instrumental activities of daily living

An alternative avenue for measuring cognition is the assessment of performance of cognitively complex everyday activities. Activities such as preparing meals, managing finances, driving a car, or using a smartphone, require higher-order cognitive processing in multiple domains. These activities are formally referred to as 'instrumental activities of daily living' (IADL). IADL are distinguished from more simple everyday activities, such as getting dressed, eating, and using the bathroom, which are referred to as 'basic activities of daily living'. This terminology was introduced in the late 1960s by Lawton and Brody,[11] and has since been widely implemented.

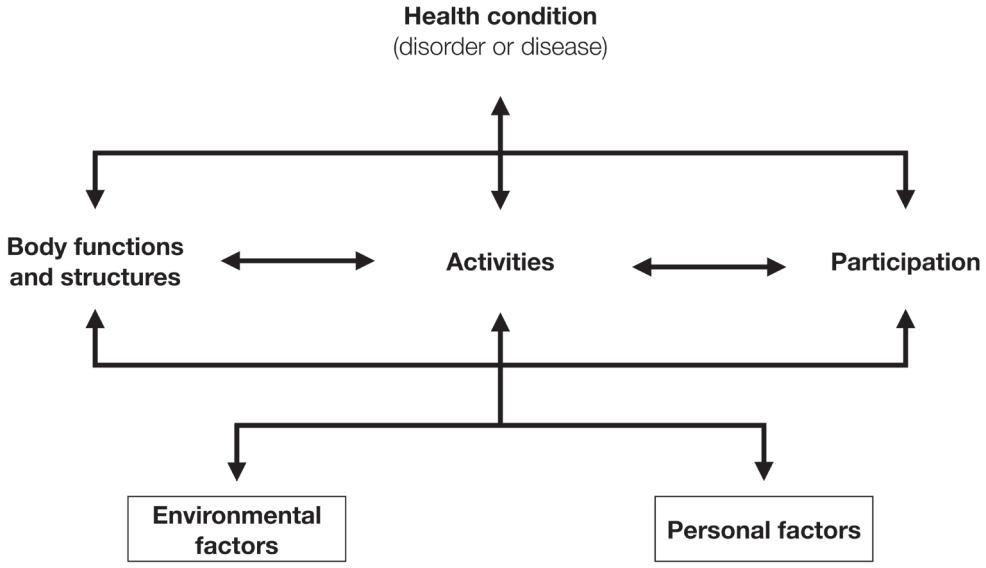


Figure 1 | International Classification of Functioning, Disability and Health

The performance of everyday activities is central in the biopsychosocial model of the International Classification of Functioning, Disability and Health (ICF, see Figure 1).[12] One of the aims of the ICF is to provide a scientific foundation for the study of health and health-related outcomes, as well as changes in health status and functioning. The model holds that an individual's level of functioning is an interaction between their health condition and environmental and personal factors.[12] Ultimately, the ICF supplies a framework within which to describe a person's level of functioning, considering the context of all characteristics and circumstances that influence the level of functioning.

In neurodegenerative diseases, a factor that has a substantial influence on a person's ability to carry out IADL independently is cognition.[13] As cognitive performance decreases, so does everyday functioning. IADL impairment is the defining feature of the dementia stage[1] and is a criterion for distinguishing mild cognitive impairment from dementia.[14] However, subtle problems in IADL may go unnoticed by a clinician, but still impact daily life to some extent. Research has indeed shown that some level of IADL impairment may be apparent before the onset of dementia.[15-18] This highlights the relevance of measuring IADL performance along the entire disease spectrum ranging from preclinical stages to dementia, using sensitive and specific tools.

The ability to perform activities of daily living is often systematically assessed in the context of Alzheimer's disease and related disorders. Most simply and directly, a self-reported questionnaire completed by an individual can provide an indication of their level of functioning. Alternatively, one can ask an informant, usually a partner, family member, friend, or other acquaintance, to rate a person's performance.[19] It is also possible to make use of performance-based instruments.[20,21] All these types of instruments can be used as outcome measures in clinical research. For the measurement of everyday functioning, informant-reported questionnaire are ubiquitous.[13]

Measurement matters

In recent years, the attention of the Alzheimer's disease scientific community has shifted to earlier disease stages. Researchers are increasingly searching for subtle, yet meaningful changes in cognition and functioning. New potentially disease-modifying drugs are being developed and investigated at a rapid rate,[22] yet many clinical trials rely on crude, insufficiently validated and outdated measures of

cognition and function. This is concerning because the clinical outcome measures in most cases determine whether trials are deemed successful. Many efforts to find disease-modifying treatments so far have failed to show a clinical benefit, including a drug now approved by the Food and Drug Administration for treatment of Alzheimer's disease in the United States.[23] This leaves one to wonder whether the outcome measures that are currently used to evaluate treatment effects are adequate for the purposes they intend to serve.

Outcome measure development comprises many steps to certify that the measure is consistent and adequately measures what it is designed to measure. The qualification for use of a new outcome measure in clinical trials by the United States Food and Drug Administration takes several years,[24] illustrating how elaborate of a process it is. Outcome measures are generally judged by two fundamental aspects of psychometric quality: reliability, i.e., the consistence of an instrument, and validity, i.e., the extent to which an instrument measures what it is intended to measure. There are various types of reliability and validity, including internal consistency, test–retest reliability and content, construct, and criterion validity.[25] For adequate assessment of everyday functioning, it is imperative that outcome measures have good psychometric qualities, yet the psychometric quality of many of these instruments is questionable or not supported by evidence.[26]

1.2 General aim and outline

In this thesis, we set out to improve our knowledge about the measurement of everyday functioning, focusing on early stages of Alzheimer's disease and related disorders. Our overall goal was to get a better understanding of when and in whom changes in everyday functioning start to occur, when they are noticeable, and how we should measure them. To achieve this goal, we formulated several aims. First, we aimed to assess the value of self-reported difficulties in everyday functioning among cognitively normal older adults. Second, in clinical populations, we aimed to investigate sources of bias in the measurement of everyday functioning. We also aimed to better understand the progression of everyday functional decline along the early disease continuum, as well as to determine what different patterns of change in everyday functioning exist among memory clinic patients. Finally, we aimed to determine the clinical meaningfulness of changes in everyday functioning.

We set out to answer the following questions:

1. What do various assessments of everyday functioning look like in cognitively normal individuals?
   1.1 To what extent can cognitively normal individuals reflect on their own level of everyday functioning?
   1.2 How is everyday functioning related to Alzheimer's disease biomarkers among cognitively normal individuals?
2. What is the influence of culture and language, age, education, and sex on the measurement of everyday functioning?
3. How does everyday functioning change over time and when are changes clinically meaningful?
   3.1 Does a mathematical scoring method improve detection of problems in everyday functioning in early disease stages?
   3.2 When do changes in everyday functioning start to occur along the Alzheimer's disease continuum?
   3.3 What groups of patients show similar changes in everyday functioning over time?
   3.4 What amount of change in everyday functioning is deemed clinically meaningful by stakeholders (caregivers and clinicians)?
   3.5 Among memory clinic patients, how often does clinically meaningful decline occur and after how much time?

Part I of this thesis addresses the first question. In **chapter 2**, we describe self-reported difficulties in everyday functioning, and compare these with study partner-reported difficulties to investigate the level of awareness of everyday functioning among cognitively normal individuals. In **chapter 3** we investigate the relationship between everyday functioning and cerebral tau in cognitively normal older adults and examine differences between participant and study partner-report in relation to tau burden. In **Part II**, we investigate the influence of various sources of bias on the measurement of everyday functioning. Chapters 4.1 and 4.2 address the potential influence of sources of bias on the measurement of everyday functioning. The source we investigate include culture/language, age, sex, and education.

The final overarching question is addressed in **Part III**. First, in **chapter 5**, we compare a standard scoring method with a mathematical scoring approach for the Amsterdam IADL Questionnaire, to investigate whether the scoring method

determines how well subtle changes in everyday functioning can be captured. In **chapter 6**, we aim to pinpoint the clinical stage in which a decline in everyday functioning can be observed, focusing specifically on early stages of Alzheimer's disease. In **chapter 7**, we try to identify groups of memory clinic patients that show similar patterns of changes in everyday functioning, as well as to find which baseline characteristics may help determine which patient belongs to what group.

In **chapter 8.1**, we determine cutoffs for mild, moderate, and severe problems in daily functioning, to facilitate interpretation of Amsterdam IADL Questionnaire scores. To determine these cutoffs, we make use of short summaries of fictional patients with different levels of functional impairment, which the participants need to compare with each other. Using similar summaries, in **chapter 8.2**, we answer the question of when changes in daily functioning are meaningful, by determining thresholds for clinically meaningful decline and improvement. Finally, we examine how often clinically meaningful decline in everyday functioning occurs by applying the thresholds to a cohort of memory clinic patients.

The main findings of this thesis are summarized and discussed in **chapter 9**.

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. 5th ed. American Psychiatric Association; 2013.
2. World Health Organization. Dementia. Updated September 2, 2021. Accessed November 19, 2021. https://www.who.int/news-room/fact-sheets/detail/dementia
3. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers & Dementia. May 2011;7(3):280-292. doi:10.1016/j.jalz.2011.03.003
4. Vermunt L, Sikkes SAM, van den Hout A, et al. Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease in relation to age, sex, and APOE genotype. Alzheimers Dement. Jul 2019;15(7):888-898. doi:10.1016/j.jalz.2019.04.001
5. Jessen F, Amariglio RE, Van Boxtel M, et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimers Dement. Nov 2014;10(6):844-52. doi:10.1016/j.jalz.2014.01.001
6. Albert MS, DeKosky ST, Dickson D, 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. May 2011;7(3):270-9. doi:10.1016/j.jalz.2011.03.008
7. Lezak MD. Neuropsychological assessment. Oxford University Press; 2004.
8. Rentz DM, Parra Rodriguez MA, Amariglio R, Stern Y, Sperling R, Ferris S. Promising developments in neuropsychological approaches for the detection of preclinical Alzheimer's disease: a selective review. Alzheimers Res Ther. 2013;5(6):58. doi:10.1186/alzrt222
9. Mortamais M, Ash JA, Harrison J, et al. Detecting cognitive changes in preclinical Alzheimer's disease: A review of its feasibility. Alzheimers Dement. Apr 2017;13(4):468-492. doi:10.1016/j.jalz.2016.06.2365
10. Tomaszewski Farias S, Harrell E, Neumann C, Houtz A. The relationship between neuropsychological performance and daily functioning in individuals with Alzheimer's disease: ecological validity of neuropsychological tests. Arch Clin Neuropsychol. 2003;18(6):655-72.
11. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. Autumn 1969;9(3):179-86.
12. World Health Organization. International Classification of Functioning, Disability and Health (ICF). Accessed 13 Apr, 2022. https://who.int/standards/classifications/international-classification-of-functioning-disability-and-health
13. Sikkes SAM, De Rotrou J. A qualitative review of instrumental activities of daily living in dementia: what's cooking? Neurodegener Dis Manag. 2014;4(5):393-400. doi:10.2217/nmt.14.24
14. Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild cognitive impairment: a concept in evolution. J Intern Med. Mar 2014;275(3):214-28. doi:10.1111/joim.12190
15. Peres K, Chrysostome V, Fabrigoule C, Orgogozo JM, Dartigues JF, Barberger-Gateau P. Restriction in complex activities of daily living in MCI: impact on outcome. Neurology. Aug 8 2006;67(3):461-6. doi:10.1212/01.wnl.0000228228.70065.f1
16. Luck T, Luppa M, Angermeyer MC, Villringer A, Konig HH, Riedel-Heller SG. Impact of impairment in instrumental activities of daily living and mild cognitive impairment on time to incident dementia: results of the Leipzig Longitudinal Study of the Aged. Psychol Med. May 2011;41(5):1087-97. doi:10.1017/S003329171000142X

17. Marshall GA, Rentz DM, Frey MT, et al. Executive function and instrumental activities of daily living in mild cognitive impairment and Alzheimer's disease. Alzheimers Dement. May 2011;7(3):300-8. doi:10.1016/j.jalz.2010.04.005

18. 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. Oct 1 2012;2(5):483-491. doi:10.2217/nmt.12.55

19. Desai AK, Grossberg GT, Sheth DN. Activities of daily living in patients with dementia: clinical relevance, methods of assessment and effects of treatment. Review. CNS Drugs. 2004;18(13):853-75. doi:10.2165/00023210-200418130-00003

20. Wales K, Clemson L, Lannin N, Cameron I. Functional Assessments Used by Occupational Therapists with Older Adults at Risk of Activity and Participation Limitations: A Systematic Review. Plos One. Feb 9 2016;11(2):e0147980. doi:10.1371/journal.pone.0147980

21. Wesson J, Clemson L, Brodaty H, Reppermund S. Estimating functional cognition in older adults using observational assessments of task performance in complex everyday activities: A systematic review and evaluation of measurement properties. Neurosci Biobehav Rev. Sep 2016;68:335-360. doi:10.1016/j.neubiorev.2016.05.024

22. Cummings J, Lee G, Zhong K, Fonseca J, Taghva K. Alzheimer's disease drug development pipeline: 2021. Alzheimers Dement (N Y). 2021;7(1):e12179. doi:10.1002/trc2.12179

23. Kuller LH, Lopez OL. ENGAGE and EMERGE: Truth and consequences? Alzheimers Dement. Apr 2021;17(4):692-695. doi:10.1002/alz.12286

24. Food and Drug Administration. Clinical Outcome Assessment (COA) Qualification Program. Updated Nov 9, 2021. Accessed Apr 23, 2022. https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/clinical-outcome-assessment-coa-qualification-program

25. Mokkink LB, Terwee CB, Patrick 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. Jul 2010;63(7):737-45. doi:10.1016/j.jclinepi.2010.02.006

26. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, Scheltens P, Uitdehaag BM. A systematic review of Instrumental Activities of Daily Living scales in dementia: room for improvement. J Neurol Neurosurg Psychiatry. Jan 2009;80(1):7-12. doi:10.1136/jnnp.2008.155838

"And the day came when the risk to remain tight in a bud was more painful than the risk it took to blossom"

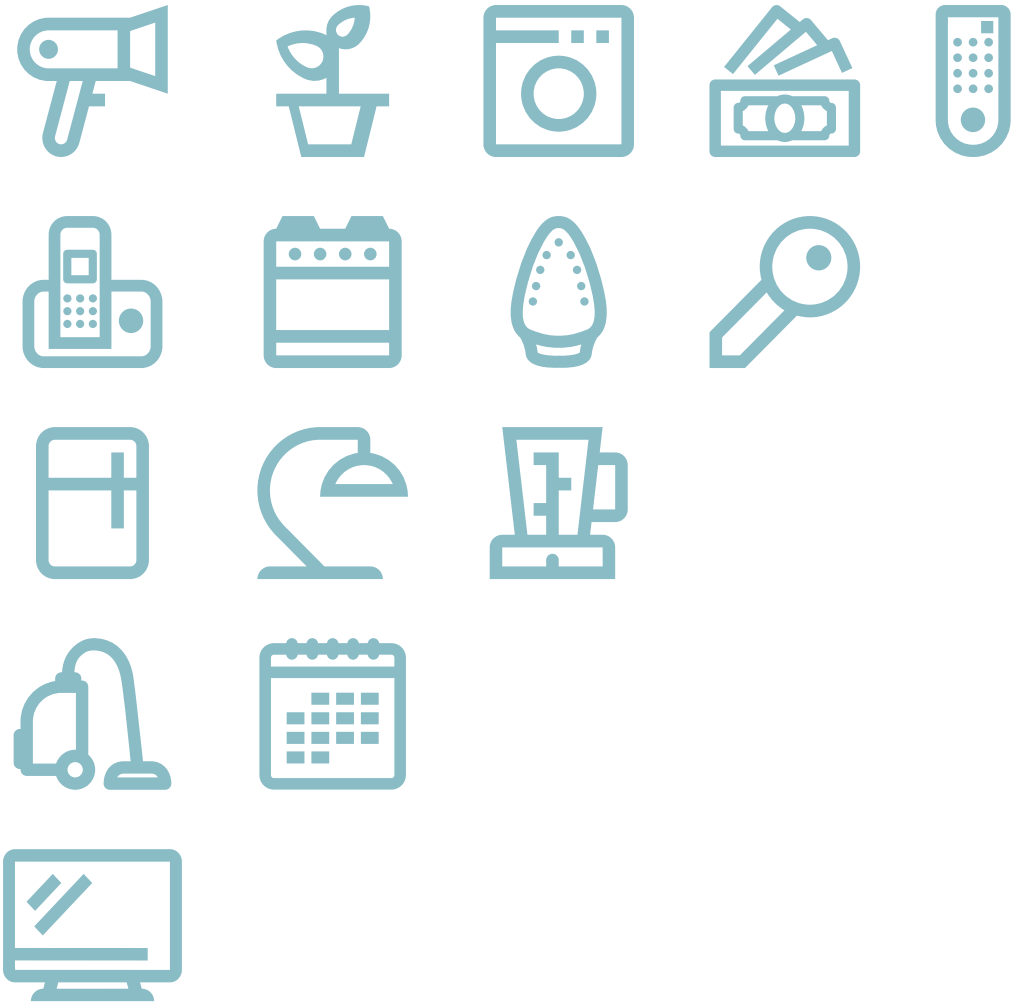
— Elizabeth Appell (1979)

PART I

EVERYDAY FUNCTIONING IN A COGNITIVELY NORMAL POPULATION

"Je n'ai fait celle-ci plus longue que parce que je n'ai pas eu le loisir de la faire plus courte"

— Blaise Pascal, 'Les Provinciales' (1657)



CHAPTER 2

EVERYDAY FUNCTIONING IN A COMMUNITY-BASED VOLUNTEER POPULATION: DIFFERENCES BETWEEN PARTICIPANT- AND STUDY PARTNER-REPORT

Merike Verrijp*, Mark A. Dubbelman*, Leonie N.C. Visser, Roos J. Jutten, Elke W. Nijhuis, Marissa D. Zwan, Hein P.J. van Hout, Philip Scheltens, Wiesje M. van der Flier, Sietske A.M. Sikkes

* contributed equally to this work

Published: January 5, 2022

In: Frontiers in Aging Neuroscience

Abstract

Introduction: Impaired awareness in dementia caused by Alzheimer's disease and related disorders made study partner-report the preferred method of measuring interference in "instrumental activities of daily living" (IADL). However, with a shifting focus toward earlier disease stages and prevention, the question arises whether self-report might be equally or even more appropriate. The aim of this study was to investigate how participant- and study partner-report IADL perform in a community-based volunteer population without dementia and which factors relate to differences between participant- and study partner-report.

Methods: Participants (*n* = 3,288; 18–97 years, 70.4% females) and their study partners (*n* = 1,213; 18–88 years, 45.8% females) were recruited from the Dutch Brain Research Registry. IADL were measured using the Amsterdam IADL Questionnaire. The concordance between participant- and study partner-reported IADL difficulties was examined using intraclass correlation coefficient (ICC). Multinomial logistic regressions were used to investigate which demographic, cognitive, and psychosocial factors related to participant and study partner differences, by looking at the over- and underreport of IADL difficulties by the participant, relative to their study partner.

Results: Most Amsterdam IADL Questionnaire scores represented no difficulties for both participants (87.9%) and study partners (89.4%). The concordance between participants and study partners was moderate (ICC = 0.55, 95% confidence interval [95%CI] = [0.51, 0.59]); 24.5% of participants (*n* = 297) overreported their IADL difficulties compared with study partners, and 17.8% (*n* = 216) underreported difficulties. The presence of depressive symptoms (odds ratio (OR) = 1.31, 95%CI = [1.12, 1.54]), as well as memory complaints (OR = 2.45, 95%CI = [1.80, 3.34]), increased the odds of participants overreporting their IADL difficulties. Higher IADL ratings decreased the odds of participant underreport (OR = 0.71, 95%CI = [0.67, 0.74]).

Conclusion: In this sample of community-based volunteers, most participants and study partners reported no major IADL difficulties. Differences between participants and study partners were, however, quite prevalent, with subjective factors indicative of increased report of IADL difficulties by the participant in particular. These findings suggest that self- and study partner-report measures may not be interchangeable, and that the level of awareness needs to be considered, even in cognitively healthy individuals.

2.1 Introduction

As the research field of Alzheimer's disease (AD) shifts its attention to earlier stages of the disease, clinically meaningful outcome measures that show early changes are becoming increasingly important.[1] One such outcome measure is the concept of "instrumental activities of daily living" (IADL), which refers to cognitively complex everyday activities.[2] Previous studies have shown that study partners report a decline in IADL in preclinical AD, even before cognitive problems can be detected by the standard cognitive testing.[3-6] Due to impairments in awareness in persons with dementia,[7] (I)ADL functioning has traditionally been assessed using study partner-report questionnaires.[8-15]

However, it has been suggested that study partner-report may be biased, by factors such as depression, anxiety, and caregiver burden.[16-18] With a shift toward studying cognitively normal or "at-risk" individuals, one might assume that participants are able to reliably reflect on their own level of functioning, as they are thought to have accurate or potentially heightened awareness of their functional and cognitive abilities, as reflected in the concept of subjective cognitive decline (SCD).[7,19] In such populations, participant-report may therefore be a more appropriate and direct assessment method.[16,17,20]

When investigating participant- and study partner-report, a few findings stand out. First, several studies have found that there is no perfect concordance between participants and study partners, even in cognitively normal populations.[13,21,22] Factors such as participant education, depression, and anxiety, as well as the nature of the relationship and the frequency and intensity of contact between participants and study partners, may affect how either party reports impairments, leading to discordance where one may report more or fewer impairments than the other. Second, studies investigating the interplay of these factors in cognitively normal populations are scarce. Furthermore, findings are difficult to compare between studies, due to differences in IADL measurements and in the definition and operationalization of concordance and discordance.

The Amsterdam IADL Questionnaire (A-IADL-Q) was developed as a study partner-rated questionnaire and has been extensively validated in memory clinic and community-based international aging populations.[23-32] It is not yet known how the participant-report version of the A-IADL-Q performs and how it relates to study partner-report. The aim of this study was to investigate how the participant- and

study partner-reported versions of the A-IADL-Q perform in a community-based population, without dementia, and what factors relate to differences between participant- and study partner-reported IADL functioning.

2.2 Materials and Methods

Participant selection and study design

Participants were selected through the Dutch Brain Research Registry (Hersenonderzoek.nl), which is an online platform for people interested in cognition and brain-related research.[33] All eligible registrants were invited by email to participate in the study. The only inclusion criterion was participants being 18 years or older. Those who self-reported to have received a dementia-related diagnosis (i.e., dementia or mild cognitive impairment [MCI]) were excluded.

Data collection started in August 2018 and ended in December 2018. The study was approved by the medical ethical committee of the VU University Medical Center. The participants provided consent via Hersenonderzoek.nl. Since study partners were not recruited through Hersenonderzoek.nl, they provided consent prior to completing the online IADL questionnaire.

Measures

Amsterdam Instrumental Activities of Daily Living Questionnaire

The main outcome measure was the A-IADL-Q. The A-IADL-Q was developed as a study partner-report instrument aimed at measuring problems in cognitively complex everyday functioning.[23] For the current study, we adapted the study partner-report version to a participant-report version. Both versions consist of the same 30 items, covering a broad range of cognitive IADL. Each item assesses difficulty performing an activity due to cognitive problems, such as problems with memory, attention, or executive functioning. Item responses were rated on a five-point Likert scale, ranging from “no difficulty in performing this activity” (0) to “no longer able to perform this activity” (4). The total score is calculated using item response theory (IRT), assuming a single underlying construct,[34] that is, IADL functioning, ranging from disability to ability. Total scores range from 20 to 70 and were reversed so that higher scores reflect better IADL functioning. A cutoff value for dementia was previously placed at 51.4,[25] while scores above 60 were considered to indicate no IADL difficulties.[32] The study partner-report version of the A-IADL-Q

has undergone extensive validation, showing a good content and construct validity, high internal consistency, high test-retest reliability, good responsiveness to change and ability to measure IADL across cultures and languages.[24-27,30] The study partner version of the A-IADL-Q also includes questions about the type of relation to the participant and cohabitation. Study partners were classified as spouses, children, siblings, or "other". Study partners in the "other" category included friends, coworkers, or other family members.

Other measures

Cognitive functioning was assessed using the Cognitive Online Self-Test Amsterdam (COST-A), an online cognitive self-test developed and validated by Van Mierlo et al.[35] The COST-A included 10 tasks, namely: orientation, digit-sequence learning, immediate word recall, two trail-making tasks (i.e., connecting numbered dots and alternately connecting lettered and numbered dots), delayed word recall, delayed word recognition, immediate recall of word pairs, recognition of word pairs, and semantic comprehension. Performance on each of the tasks was standardized and averaged into a Z-score to represent overall cognitive functioning, where higher scores indicate better cognition. Visser et al.[36] provide a more detailed description of the COST-A.

In addition, a single yes/no question ("Do you have memory complaints?") assessed subjective memory complaints. Depressive symptoms were assessed with the five-item short form of the Geriatric Depression Scale (GDS5),[37] with higher scores indicating more depressive symptoms. The education level was classified as low-medium (up to high school) and high education (college degree).

Defining awareness of IADL functioning

In line with other studies, we defined concordance based on the discrepancy between participant- and study partner-report.[7] Based on a previously determined clinically meaningful difference over time of 2.4 points,[38] we categorized concordance into three groups, namely, (1) concordance between dyads, (2) discordance between dyads with the participant "over reporting" difficulties (i.e., scoring ≥2.4 points lower than their study partner), and (3) discordance between dyads with the participant "underreporting" difficulties (i.e., scoring ≥2.4 points higher than their study partner).

Statistical analysis

Demographic differences between study partners and participants were tested using independent *t*-tests or chi-square tests. The frequency of IADL difficulties among cognitively normal participants and their study partners was determined. Then, in separate linear regression analyses, A-IADL-Q scores of both raters were associated with age, education, objective cognitive functioning, subjective cognitive functioning, and depressive symptoms.

The intraclass correlation coefficient (ICC) was computed to examine the absolute agreement between participant and study partner ratings. According to the criteria suggested by Koo et al.,[39] an ICC < 0.5 shows poor agreement, an ICC of 0.5–0.75 shows moderate, and an ICC > 0.75 shows good agreement.

Using multinomial logistic regression models, we investigated which factors related to concordance and discordance between dyads. The variables included the following parameters of participants: education level, sex, age, COST-A scores, memory complaints, GDS5 total score, study partner-reported IADL functioning, the type of relationship, cohabitation (yes/no), and the absolute age difference between dyads. For this analysis, COST-A scores were dichotomized into normal (higher than -1.5) and low (lower than or equal to -1.5) cognitive functioning. All analyses were performed using R version 4.0.3 software.[40]

2.3 Results

Of the 11,060 eligible registrants, 4,817 individuals (44%) were interested in participation and received study instructions. After receiving instructions, 3,288 (68%) individuals completed the participant-reported A-IADL-Q. On average, participants were 61.0 ± 12.1 years old and the majority of them were women (*N* = 2,315, 70.4%). Approximately half the participants experienced memory complaints. Table 1 displays all participant and study partner characteristics. Participant and study partner characteristics stratified by age groups are shown in the Supplementary Material.

Table 1 | Participant and study partner characteristics.

	Participants (*n* = 3,288)	Dyads (*n* = 1,213)	
		Participants	Study partners
Age, mean (SD)	61.0 (12.1)	62.5 (11.1)	58.8 (14.2)
Range	18–97	18–93	18–88
Female, n (%)	2,315 (70.4)	828 (68.3)	556 (45.8)
High level of education, n (%)	2,323 (70.7)	854 (70.4)	—
A-IADL-Q score, mean (SD)	65.9 (4.8)	65.9 (4.7)	66.1 (4.6)
Range	40.9–70.0	40.9–70.0	42.7–70.0
Memory complaints present,[a] n (%)	1,429 (47.5)	586 (49.9)	—
COST-A, abnormal performance,[b] n (%)	225 (7.6)	86 (7.5)	—
GDS5,[a] median (IQR)	0 (0–1)	0 (0–1)	—
Type of relationship, n (%)	—		
Spouse		956 (78.8)	
Child		155 (12.8)	
Sibling		32 (2.6)	
Other		70 (5.8)	
Duration of relationship, n (%)	—		
<5 years		33 (2.7)	
5–10 years		58 (4.8)	
>10 years		1,119 (92.5)	
Living together, n (%)	—	960 (79.3)	

"—" denotes that the data were not available. [a] Data were available for 3,011 participants, of whom 1,175 were part of a dyad. [b] Data were available for 2,945 participants, of whom 1,149 were part of a dyad. Abbreviations: A-IADL-Q, Amsterdam Instrumental Activities of Daily Living Questionnaire; COST-A, Cognitive Online Self-Test Amsterdam; GDS5, 5-item Geriatric Depression Scale; IQR, interquartile range; SD, standard deviation.

For 1,213 participants (36.9% of complete sample), the A-IADL-Q was also completed by a study partner (participant and study partner pairs will be referred to as "dyads"). Participants who were part of a dyad were 62.5 ± 11.1 years old and the majority of them were women (*N* = 828, 68.3%). They were older ($p < .001$) and more often male ($p = .046$) than participants who were not part of a dyad. Within dyads, the participants were older ($p < .001$) and more likely to be female ($p < .001$) than study partners.

IADL difficulties in a cognitively normal population

Figure 1 shows the distribution of participant- and study partner-reported A-IADL-Q scores. Among dyads, the participant reported A-IADL-Q scores (65.9 ± 4.8) did not

differ from the study partner-reported A-IADL-Q scores (66.1 ± 4.6; p = .186). Virtually all participants (3,232/3,288, 98.3%) and study partners (1,195/1,213, 98.5%) reported A-IADL-Q scores above the previously established cutoff for dementia (total score of 51.4). Moreover, the vast majority of both participant-reported (87.9%) and study partner-reported (89.4%) total scores were higher than 60, indicating no difficulties.

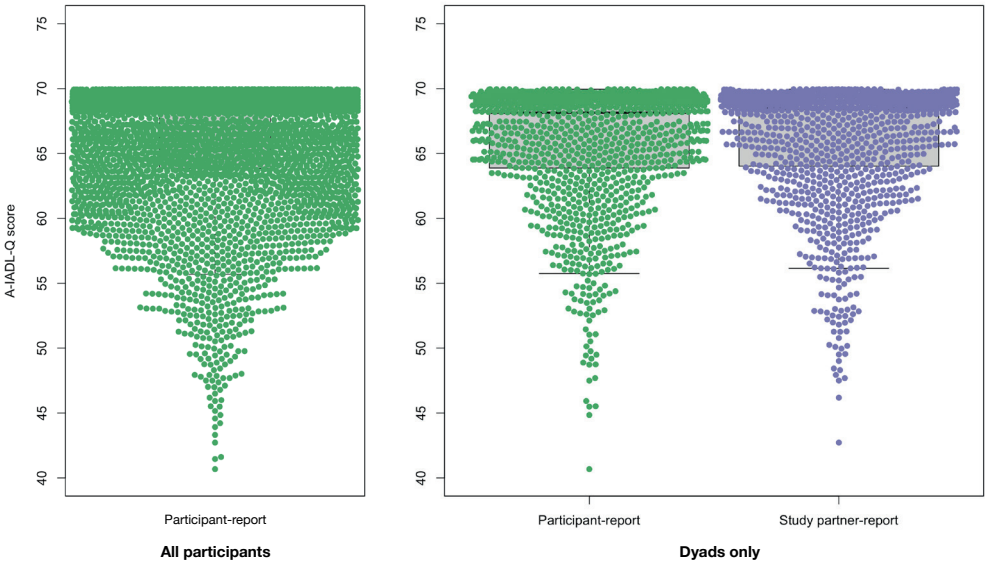


Figure 1 | A-IADL-Q total score distribution among all participants (left panel, n = 3,288) and among dyads (right panel, n = 1,213; participants are shown in green; study partners are shown in purple).

Then, we examined IADL difficulties at an item level. Half of all participants (n = 1,750/3,288, 53.2%) and study partners (n = 722/1,213, 59.5%) reported no difficulties in any activity. Those who reported difficulties mostly did so in only one activity (35.2% of participants and 35.8% of study partners). Figure 2 shows the percentage of participants and study partners who reported difficulties for each IADL activity. Most frequently reported IADL difficulties for both participants and study partners were working (26.9 and 19.9%, respectively), household duties (22.2 and 16.5%, respectively), and making minor repairs at home (16.4 and 12.7%, respectively).

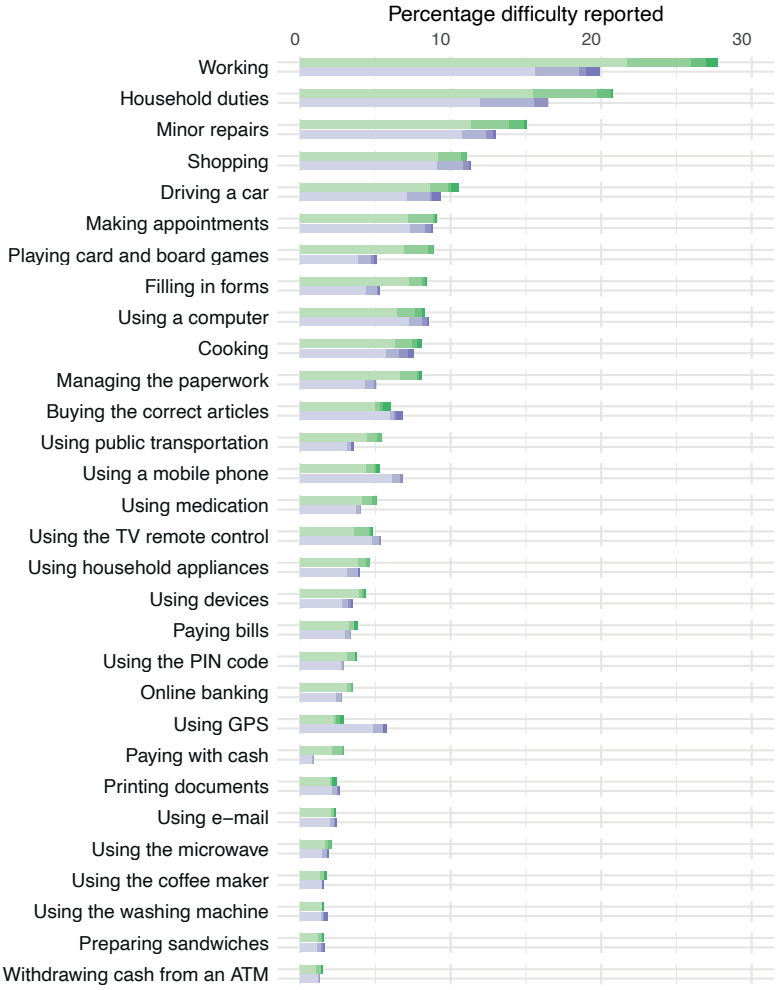


Figure 2 | Stacked bar chart showing the percentage of participants (denoted in shades of green) and study partners (denoted in shades of purple) who reported difficulties (*n* = 1,213). The dark shades represent difficulty with the activities: "no longer able to perform this activity" (4), "much more difficulty" (3), "more difficulty" (2), and "slightly more difficulty" (1). The lightest shade represents "no difficulty in performing this activity" (0). Displaying data from dyads only.

Table 2 shows the associations between age, education level, cognitive complaints, COST-A, GDS5, and participant- and study partner-reported IADL performance. Higher age was associated with lower A-IADL-Q scores, and high education was associated with better A-IADL-Q scores, but associations were weak. For example, with every 10 years increase in age, A-IADL-Q participant- and study partner-

reported scores decreased with 1.2 and 1.8 points, respectively. Both participant- and study partner-reported A-IADL-Q scores were more highly associated with COST-A scores, memory complaints, and GDS5. Higher COST-A scores, indicating better cognitive functioning, were associated with better IADL functioning, whereas a higher GDS5, indicating more depressive symptoms, and presence of memory complaints were associated with worse IADL functioning. Associations with age, education, and COST-A scores were comparable for participant- and study partner-report, whereas associations with GDS5 and memory complaints were more strongly associated with participant-reported IADL scores.

Table 2 | Linear regressions to investigate associations with participant- and study partner-reported IADL performance.

Measure	Participant-report	Study partner-report
Age	-0.12 [-0.16, -0.09]	-0.18 [-0.26, -0.14]
High education	0.09 [0.06, 0.13]	0.07 [0.02, 0.13]
Memory complaints present	-0.33 [-0.36, -0.29]	-0.24 [-0.30, -0.19]
COST-A	0.23 [0.19, 0.26]	0.25 [0.20, 0.31]
GDS5	-0.33 [-0.36, -0.29]	-0.21 [-0.30, -0.17]

Associations are shown as standardized beta [95% confidence interval]. Some measures were not available for the entire sample. Memory complaints were available for n = 3,011 participants and n = 1,175 participants who were part of a dyad. COST-A scores were available for n = 2,945 participants and n = 1,149 participants who were part of a dyad. GDS5 scores were available for n = 3,017 participants and n = 1,177 participants who were part of a dyad.

Concordance and discordance between dyads

There was a moderate agreement between participant- and study partner-reported IADL functioning (ICC = 0.55, 95%CI = [0.51, 0.59], p < .001; see Supplementary Material). Of all 1,213 dyads, 700 (57.7%) were in concordance. Two hundred sixteen participants (17.8%) underreported difficulties, compared with their study partners, and 297 participants (24.5%) overreported IADL difficulties. Compared with concordant dyads, participants with memory complaints (odds ratio [OR] = 2.44, 95%CI = [1.80, 3.32], p < .001) and with a higher GDS5 (OR = 1.31, 95%CI = [1.12, 1.53], p = .001) were more likely to overreport IADL difficulties (see Table 3). Participant underreport was less likely when there were fewer IADL difficulties (OR = 0.71, 95%CI = [0.67, 0.74], p < .001). Thus, concordance was more likely when the participant did not experience memory complaints, when they had lower GDS5 scores, and when IADL performance was higher. Education, age, gender, and COST-A scores of participants were not related to concordance between dyads.

Table 3 | Multivariable multinomial logistic regression models comparing study partners reporting more IADL difficulties than participants (*n* = 216) and participants reporting more IADL difficulties than study partners (*n* = 297), set against agreement between participants and study partners (*n* = 700).

	Study partner > participant (*n* = 216)		Participant > study partner (*n* = 297)	
	OR [95%CI]	**p**	**OR [95%CI]**	**p**
COST-A ≤ -1.5 SD	0.47 [0.21, 1.07]	0.070	1.36 [0.78, 2.39]	0.283
A-IADL-Q (study partner report)	0.71 [0.67, 0.74]	< 0.001	1.04 [0.99, 1.09]	0.148
Memory complaints presents	0.76 [0.50, 1.15]	0.194	2.44 [1.80, 3.32]	< 0.001
High education	0.92 [0.60, 1.40]	0.689	1.30 [0.93, 1.80]	0.121
Absolute age difference between dyads in years	1.00 [0.97, 1.04]	0.924	1.01 [0.98, 1.04]	0.924
Age in years (participant)	1.01 [0.99, 1.03]	0.467	1.01 [0.99, 1.02]	0.272
Female sex (participant)	0.74 [0.53, 1.02]	0.159	1.08 [0.78, 1.49]	0.661
GDS5[a]	0.58 [0.50, 0.68]	< 0.001	1.31 [1.12, 1.53]	< 0.001
Type of relationship, study partner is a[b]				
Child	2.19 [0.63, 7.60]	0.216	0.83 [0.30, 2.27]	0.716
Sibling	0.75 [0.13, 4.35]	0.744	0.57 [0.18, 1.85]	0.350
Other	0.81 [0.22, 2.98]	0.755	0.63 [0.24, 1.68]	0.355
Dyads live together	1.58 [0.70, 3.57]	0.277	1.04 [0.57, 1.90]	0.898

[a] More depressive symptoms; [b] Using spouse as a reference category.

Abbreviations: A-IADL-Q, Amsterdam Instrumental Activities of Daily Living Questionnaire; COST-A, Cognitive Online Self Test Amsterdam; GDS, Geriatric Depression Scale; OR, odds ratio; 95%CI, 95% confidence interval.

2.4 Discussion

In this study, we showed that the majority of IADL scores fell within the range of normal IADL functioning in this community-based population, but that discordance among dyads was quite prevalent. A small proportion reported subtle IADL difficulties, which was associated with older age, lower education, worse cognitive performance, presence of self-reported memory complaints, and more depressive symptoms of participants, for both participant- and study partner-report. A moderate agreement between participant- and study partner-reported IADL was found with discordance between dyads being more likely when the participant reported memory complaints and had depressive symptoms and lower IADL performance.

While the large majority of participant- and study partner-reported IADL functioning fell within the range of normal IADL functioning, approximately a tenth of both participants and study partners scored below the previously established cutoff for normal IADL functioning.[32] This prevalence of impaired IADL is comparable with other population-based studies.[41-44] For example, Scheel-Hincke and colleagues[42] reported a prevalence of impaired IADL of 12 to 20% in Western Europe, with impaired IADL defined as presence of any difficulties. Another population-based study by Pudaric and colleagues[43] reported a prevalence of impaired IADL (inability to carry out shopping, cooking, or housework) of 6 to 11%. Despite this comparable prevalence of abnormal IADL functioning, it is important to note that approximately half of our population reported more subtle difficulties. If we applied the definition of Scheel-Hincke et al.,[42] the prevalence of impaired IADL in our study would be approximately 50%, which is substantially higher than the prevalence they reported. There are two potential explanations for this difference: first, we included more activities, and second, and more importantly, we included more cognitively complex activities than other studies. This is illustrated by the fact that most problems were reported in working, household duties, and making repairs, which are especially cognitively complex.[27] These activities were not included in other IADL scales. For example, a population-based study that assessed five IADL items reported most problems for shopping.[45] In our population, problems with shopping were fourth most prevalent. We found a higher proportion of difficulties for more complex activities, supporting the notion that including more complex activities enabled detection of more fine-grained difficulties in IADL functioning.

With regard to potential sources of bias in the report of IADL functioning, we found low associations between both study partner and participant-reported IADL functioning and age and education. This finding is supported by previous validation studies for the study partner version of the A-IADL-Q.[24,27,30] Participant- and study partner-report were similarly associated with objective cognitive performance, but participant-reported IADL functioning was more strongly related to depressive symptoms, as well as subjective cognitive performance (i.e., presence of self-reported memory complaints). Consistent with recent literature suggesting that study partners are better able to assess the functioning of participants than the participant themselves,[46] our findings might imply that study partner-report is less biased than participant-report by participant-related subjective factors.

Our findings demonstrated only a moderate concordance between dyads. While the distributions of study partner- and participant-reported IADL scores were largely

similar, we found a moderate ICC and a high proportion of discordance (either over- or underreport). Other studies have also shown discordance in cognitively normal participants and, specifically, participant overreport.[13,21,41,47] For example, a study by Okonkwo and colleagues[21] showed slight discordance between participant- and study partner-report of specific finance-related IADL. The proportion of discordance that we found in our study is substantially higher, which is probably due to differences in IADL measures, definitions of concordance, and population differences. As opposed to Okonkwo and colleagues,[21] who calculated concordance based on an individual item, we determined concordance based on a more global measure of IADL with a wider range of activities. We calculated concordance based on a clinically meaningful difference in total scores. Another potential explanation may be that, even though we used a population-based sample, we did not screen for cognitive impairment. As such, it is possible that there were participants who had subtle cognitive impairment but did not meet criteria for MCI or dementia. Thus, while the proportion of discordance is difficult to compare with other studies, the fact that other studies also reported discordance suggests that participant and study partner-report might not be interchangeable.

The potential limited interchangeability is further supported by our results, which indicate that concordance is influenced by self-reported memory complaints and depressive symptoms. Participants with memory complaints reported more difficulties, compared with their study partners. Participant overreport of memory complaints has previously been described as a heightened awareness,[7] which is thought to characterize early stages of AD and related disorders.[7,48,49] Following this theory, a subgroup of our study sample may have a heightened functional awareness. While no other studies have investigated the effect of subjective cognitive functioning on the concordance of functional impairment, several studies[13,21,41,47,50-52] related objective cognitive functioning to concordance. These studies show that patients with poorer global cognition are more likely to underreport IADL difficulties. We did not find a significant association between concordance and objective cognition within our healthy volunteer population. This could be due to the fact that our population is presumably cognitively healthy, and lowered awareness may not occur until cognitive problems start to develop.[7,53] Although not significant, in this population, lower cognitive performance seems to be related to reduced odds for participant underreport. This might suggest that the subtle cognitive problems of these individuals do not interfere with their disease insight, but rather, that they increase their awareness. Furthermore,

participants with depressive symptoms were more likely to overreport, and less likely to underreport, IADL difficulties. This was also reported in studies in MCI and dementia that showed a greater chance of discordance when participants had depressive symptoms.[21,54] This is in line with the idea that negative self-perception in patients with depressive symptoms causes exaggeration of deficits,[55] as has also been shown by Okonkwo and colleagues,[21] who reported that underestimation of financial abilities was related to higher depressive symptoms. Thus, memory complaints and depressive symptoms both need to be taken into consideration when using participant-reported IADL measures.

The findings discussed earlier may have important implications for study design decisions and should be considered carefully when considering the use of a participant-reported IADL instrument. Although a concordance of 60% might seem low, the majority of both participant- and study partner-reported difficulties fell within the category of "no difficulties". This crude overlap indicates that participant-report IADL can be useful in cognitively normal populations in cross-sectional studies. However, when a deterioration of cognitive functioning and subsequently everyday functioning is to be expected, study partner-report might provide a more reliable indication of change in IADL functioning. The combination of participant- and study partner-report can be used to establish awareness, which is informative since it has been shown to predict future disease progression[56,57] and greater discordance seems to be related to a greater risk of Alzheimer pathology.[7,52] The combination of participant- and study partner-report might also be valuable as they seem to reflect different perspectives. This is reflected in the current study as participant-report seems to be more influenced by subjective factors than study partner-report. The different perspectives were also implied in an article by Amariglio and colleagues[58] who showed that distinct IADL items were related to amyloid pathology for participants and study partners. Thus, participant self-report can be used in cognitively normal populations but should ideally be supplemented by study partner-report, not only when considering the cognitive decline of participants in longitudinal studies, but also to gain multiple perspectives and insight into the awareness of participants.

Some limitations should be considered when interpreting our findings. For the lack of an objective IADL measure, we cannot ascertain whether participants indeed overreport their difficulties or whether participants actually have IADL difficulties that the study partner does not yet notice. In contrast, a heightened participant awareness may also reflect lowered study partner awareness. This caveat

notwithstanding, the absence of an association between participant overreport and objective cognitive functioning could indicate that participant overreport is more strongly influenced by subjective than objective factors. It should also be noted that objective cognition and IADL performance cannot be completely separated, as IADL performance is dependent on cognition. This may introduce some level of circularity into the analyses. However, the association between our objective cognitive measure and the A-IADL-Q scores was only moderate. Furthermore, as the study partner-report is generally considered a gold standard in dementia research and clinical practice,[59] we used it as such in the current study. Another limitation is the selective nature of the volunteer registry, which consists mostly of highly educated and highly motivated individuals. This may limit generalizability to the general population. We did not include factors such as caregiver burden, personality traits, or more detailed information on the amount of contact between the participant and the study partner. Future studies should consider assessing these factors to obtain more detailed insight into the accuracy of assessments and possible biases. Furthermore, follow-up studies are needed to determine the pivot point until which the participant is still able to reliably evaluate their own level of daily functioning.

An important strength of this study is the large sample of cognitively healthy volunteers, representing a large range of ages, from early adulthood to late life. We included detailed information about the level of IADL difficulties from both self- and study partner-report in a cognitively healthy population, providing valuable new insights into the occurrence of more subtle IADL difficulties. While the clinically meaningful cutoff was determined for decline and not for differences between respondents, a strength of this clinically meaningful cutoff to distinguish concordance from discordance is that we believed that discordance actually represented an important, non-negligible difference in IADL report.

Conclusion

Our findings show a moderate concordance between participants and study partners in reporting IADL difficulties, with subjective factors influencing the level of concordance. These findings suggest caution in using self- and study partner-report measures interchangeably, even in cognitively healthy community-based samples. Our results suggest that participant-report might be more related to subjective factors and that study partner-report is less associated with these factors, possibly reflecting differing perspectives.

References

1. Edgar CJ, Vradenburg G, Hassenstab J. The 2018 Revised FDA Guidance for Early Alzheimer's Disease: Establishing the Meaningfulness of Treatment Effects. J Prev Alz Dis. 2019;6doi:10.14283/jpad.2019.30
2. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. Autumn 1969;9(3):179-86.
3. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers & Dementia. May 2011;7(3):280-292. doi:10.1016/j.jalz.2011.03.003
4. 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. Oct 1 2012;2(5):483-491. doi:10.2217/nmt.12.55
5. Marshall GA, Aghjayan SL, Dekhtyar M, et al. Activities of daily living measured by the Harvard Automated Phone Task track with cognitive decline over time in non-demented elderly. J Prev Alzheimers Dis. 2017;4(2):81-86. doi:10.14283/jpad.2017.10
6. Zoller AS, Gaal IM, Royer CA, et al. SIST-M-IR activities of daily living items that best discriminate clinically normal elderly from those with mild cognitive impairment. Curr Alzheimer Res. 2014;11(8):785-791. doi:Doi 10.2174/1567205011081409101222 20
7. Hanseeuw BJ, Scott MR, Sikkes SAM, et al. Evolution of anosognosia in alzheimer's disease and its relationship to amyloid. Ann Neurol. Feb 2020;87(2):267-280. doi:10.1002/ana.25649
8. Desai AK, Grossberg GT, Sheth DN. Activities of daily living in patients with dementia: clinical relevance, methods of assessment and effects of treatment. Review. CNS Drugs. 2004;18(13):853-75. doi:10.2165/00023210-200418130-00003
9. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, Scheltens P, Uitdehaag BM. A systematic review of Instrumental Activities of Daily Living scales in dementia: room for improvement. J Neurol Neurosurg Psychiatry. Jan 2009;80(1):7-12. doi:10.1136/jnnp.2008.155838
10. Loewenstein DA, Arguelles S, Bravo M, et al. Caregivers' judgments of the functional abilities of the Alzheimer's disease patient: a comparison of proxy reports and objective measures. J Gerontol B Psychol Sci Soc Sci. Mar 2001;56(2)(2):P78-84. doi:10.1093/geronb/56.2.p78
11. Howorth P, Saper J. The dimensions of insight in people with dementia. Aging Ment Health. Mar 2003;7(2):113-22. doi:10.1080/1360786031000072286
12. Wadley VG, Harrell LE, Marson DC. Self- and informant report of financial abilities in patients with Alzheimer's disease: reliable and valid? J Am Geriatr Soc. Nov 2003;51(11):1621-6. doi:10.1046/j.1532-5415.2003.51514.x
13. Farias ST, Mungas D, Jagust W. Degree of discrepancy between self and other-reported everyday functioning by cognitive status: dementia, mild cognitive impairment, and healthy elders. Int J Geriatr Psychiatry. Sep 2005;20(9):827-34. doi:10.1002/gps.1367
14. Graham DP, Kunik ME, Doody R, Snow AL. Self-reported awareness of performance in dementia. Brain Res Cogn Brain Res. Sep 2005;25(1):144-52. doi:10.1016/j.cogbrainres.2005.05.001
15. Hackett K, Mis R, Drabick DAG, Giovannetti T. Informant Reporting in Mild Cognitive Impairment: Sources of Discrepancy on the Functional Activities Questionnaire. J Int Neuropsychol Soc. May 2020;26(5):503-514. doi:10.1017/S1355617719001449

16. Zanetti O, Geroldi C, Frisoni GB, Bianchetti A, Trabucchi M. Contrasting results between caregiver's report and direct assessment of activities of daily living in patients affected by mild and very mild dementia: the contribution of the caregiver's personal characteristics. J Am Geriatr Soc. Feb 1999;47(2):196-202. doi:10.1111/j.1532-5415.1999.tb04578.x

17. Arguelles S, Loewenstein DA, Eisdorfer C, Arguelles T. Caregivers' judgments of the functional abilities of the Alzheimer's disease patient: impact of caregivers' depression and perceived burden. J Geriatr Psychiatry Neurol. Summer 2001;14(2):91-8. doi:10.1177/089198870101400209

18. Ready RE, Ott BR, Grace J. Validity of informant reports about AD and MCI patients' memory. Alzheimer Dis Assoc Dis. Jan-Mar 2004;18(1):11-6. doi:10.1097/00002093-200401000-00003

19. Steward KA, Kennedy R, Erus G, Nasrallah IM, Wadley VG. Poor awareness of IADL deficits is associated with reduced regional brain volume in older adults with cognitive impairment. Neuropsychologia. Jun 2019;129:372-378. doi:10.1016/j.neuropsychologia.2019.04.023

20. DeBettignies BH, Mahurin RK, Pirozzolo FJ. Insight for impairment in independent living skills in Alzheimer's disease and multi-infarct dementia. J Clin Exp Neuropsychol. Mar 1990;12(2):355-63. doi:10.1080/01688639008400980

21. Okonkwo OC, Wadley VG, Griffith HR, et al. Awareness of deficits in financial abilities in patients with mild cognitive impairment: going beyond self-informant discrepancy. Am J Geriatr Psychiatry. Aug 2008;16(8):650-9. doi:10.1097/JGP.0b013e31817e8a9d

22. Marshall GA, Sikkes SAM, Amariglio RE, et al. Instrumental activities of daily living, amyloid, and cognition in cognitively normal older adults screening for the A4 Study. Alzheimers Dement (Amst). 2020;12(1):e12118. doi:10.1002/dad2.12118

23. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. Alzheimers Dement. Nov 2012;8(6):536-43. doi:10.1016/j.jalz.2011.08.006

24. Sikkes SAM, Knol DL, Pijnenburg YAL, 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(1):35-41. doi:10.1159/000346277

25. Sikkes SAM, Pijnenburg YAL, 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. Dec 2013;26(4):244-50. doi:10.1177/0891988713509139

26. Koster N, Knol DL, Uitdehaag BM, Scheltens P, Sikkes SAM. The sensitivity to change over time of the Amsterdam IADL Questionnaire((c)). Alzheimers Dement. Oct 2015;11(10):1231-40. doi:10.1016/j.jalz.2014.10.006

27. Jutten RJ, Peeters CFW, Leijdesdorff SMJ, et al. Detecting functional decline from normal aging to dementia: Development and validation of a short version of the Amsterdam IADL Questionnaire. Alzheimers Dement (Amst). 2017;8:26-35. doi:10.1016/j.dadm.2017.03.002

28. Facal D, Carabias MAR, Pereiro AX, et al. Assessing everyday activities across the dementia spectrum with the Amsterdam IADL Questionnaire. Curr Alzheimer Res. 2018;15(13):1261-1266. doi:10.2174/1567205015666180925113411

29. Villeneuve SC, Houot M, Cacciamani F, et al. Latent class analysis identifies functional decline with Amsterdam IADL in preclinical Alzheimer's disease. Alzheimers Dement (N Y). 2019;5:553-562. doi:10.1016/j.trci.2019.08.009

30. Dubbelman MA, Verrijp M, Facal D, et al. The influence of diversity on the measurement of functional impairment: An international validation of the Amsterdam IADL Questionnaire in eight countries. Alzheimers Dement (Amst). 2020;12(1):e12021. doi:10.1002/dad2.12021
31. Bruderer-Hofstetter M, Dubbelman MA, Meichtry A, et al. Cross-cultural adaptation and validation of the Amsterdam Instrumental Activities of Daily Living questionnaire short version German for Switzerland. Health Qual Life Outcomes. Oct 2 2020;18(1):323. doi:10.1186/s12955-020-01576-w
32. Dubbelman MA, Verrijp M, Visser LNC, et al. A mixed-methods approach to establish clinically meaningful categories of impairment in instrumental activities of daily living. Alzheimer's & Dementia. 2020;16(S9)doi:10.1002/alz.045693
33. Zwan MD, Van der Flier WM, Cleutjens S, et al. Dutch Brain Research Registry for online study participant recruitment: Design and first results. Alzheimer's & Dementia. 2020;16(S10)doi:10.1002/alz.044738
34. Reise SP, Waller NG. Item response theory and clinical measurement. Annu Rev Clin Psychol. 2009;5:27-48. doi:10.1146/annurev.clinpsy.032408.153553
35. Van Mierlo LD, Wouters H, Sikkes SAM, et al. Screening for mild cognitive impairment and dementia with automated, anonymous online and telephone cognitive self-tests. J Alzheimers Dis. 2017;56(1):249-259. doi:10.3233/JAD-160566
36. Visser LNC, Dubbelman MA, Verrijp M, et al. The Cognitive Online Self-Test Amsterdam (COST-A): Establishing norm scores in a community-dwelling population. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. 2021;13(1) doi:10.1002/dad2.12234
37. Hoyl MT, Alessi CA, Harker JO, et al. Development and testing of a five-item version of the Geriatric Depression Scale. J Am Geriatr Soc. Jul 1999;47(7):873-8. doi:10.1111/j.1532-5415.1999.tb03848.x
38. Abstract: Symposia, Conferences, Oral communications: 13th Clinical Trials on Alzheimer's Disease (CTAD) November 4-7, 2020. J Prev Alzheimers Dis. 2020;7(S1):S2-S54. doi:10.14283/jpad.2020.54
39. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. Jun 2016;15(2):155-63. doi:10.1016/j.jcm.2016.02.012
40. R: A language and environment for statistical computing. Version 4.0.3. 2020.
41. Ostbye T, Tyas S, McDowell I, Koval J. Reported activities of daily living: Agreement between elderly subjects with and without dementia and their caregivers. Age Ageing. Mar 1997;26(2):99-106. doi:DOI 10.1093/ageing/26.2.99
42. Scheel-Hincke LL, Moller S, Lindahl-Jacobsen R, Jeune B, Ahrenfeldt LJ. Cross-national comparison of sex differences in ADL and IADL in Europe: findings from SHARE. Eur J Ageing. Mar 2020;17(1):69-79. doi:10.1007/s10433-019-00524-y
43. Pudaric S, Sundquist J, Johansson S-E. Country of birth, instrumental activities of daily living, self-rated health and mortality: a Swedish population-based survey of people aged 55–74. Social Science & Medicine. 2003;56(12):2493-2503. doi:10.1016/s0277-9536(02)00284-8
44. Crimmins EM, Kim JK, Sole-Auro A. Gender differences in health: results from SHARE, ELSA and HRS. Eur J Public Health. Feb 2011;21(1):81-91. doi:10.1093/eurpub/ckq022
45. Chan KS, Kasper JD, Brandt J, Pezzin LE. Measurement equivalence in ADL and IADL difficulty across international surveys of aging: findings from the HRS, SHARE, and ELSA. J Gerontol B Psychol Sci Soc Sci. Jan 2012;67(1):121-32. doi:10.1093/geronb/gbr133

46. Howland M, Allan KC, Carlton CE, Tatsuoka C, Smyth KA, Sajatovic M. Patient-rated versus proxy-rated cognitive and functional measures in older adults. Patient Relat Outcome Meas. 2017;8:33-42. doi:10.2147/PROM.S126919

47. Pol MC, Buurman BM, de Vos R, de Rooij SE. Patient and proxy rating agreements on activities of daily living and the instrumental activities of daily living of acutely hospitalized older adults. J Am Geriatr Soc. Aug 2011;59(8):1554-6. doi:10.1111/j.1532-5415.2011.03514.x

48. Jessen F, Amariglio RE, Van Boxtel M, et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimers Dement. Nov 2014;10(6):844-52. doi:10.1016/j.jalz.2014.01.001

49. Slot RER, Sikkes SAM, Berkhof J, et al. Subjective cognitive decline and rates of incident Alzheimer's disease and non-Alzheimer's disease dementia. Alzheimers Dement. Mar 2019;15(3):465-476. doi:10.1016/j.jalz.2018.10.003

50. Weinberger M, Samsa GP, Schmader K, Greenberg SM, Carr DB, Wildman DS. Comparing proxy and patients' perceptions of patients' functional status: results from an outpatient geriatric clinic. J Am Geriatr Soc. Jun 1992;40(6):585-8. doi:10.1111/j.1532-5415.1992.tb02107.x

51. Albert SM, Michaels K, Padilla M, et al. Functional significance of mild cognitive impairment in elderly patients without a dementia diagnosis. Am J Geriatr Psychiatry. Summer 1999;7(3):213-20. doi:10.1097/00019442-199908000-00005

52. Tabert MH, Albert SM, Borukhova-Milov L, et al. Functional deficits in patients with mild cognitive impairment: prediction of AD. Neurology. Mar 12 2002;58(5):758-64. doi:10.1212/wnl.58.5.758

53. Starkstein SE, Jorge R, Mizrahi R, Robinson RG. A diagnostic formulation for anosognosia in Alzheimer's disease. J Neurol Neurosurg Psychiatry. Jun 2006;77(6):719-25. doi:10.1136/jnnp.2005.085373

54. Magaziner J, Bassett SS, Hebel JR, Gruber-Baldini A. Use of proxies to measure health and functional status in epidemiologic studies of community-dwelling women aged 65 years and older. Am J Epidemiol. Feb 1 1996;143(3):283-92. doi:10.1093/oxfordjournals.aje.a008740

55. Lahr D, Beblo T, Hartje W. Cognitive performance and subjective complaints before and after remission of major depression. Cogn Neuropsychiatry. Jan 2007;12(1):25-45. doi:10.1080/13546800600714791

56. Nosheny RL, Jin C, Neuhaus J, et al. Study partner-reported decline identifies cognitive decline and dementia risk. Ann Clin Transl Neurol. Dec 2019;6(12):2448-2459. doi:10.1002/acn3.50938

57. Nosheny RL, Camacho MR, Jin C, et al. Validation of online functional measures in cognitively impaired older adults. Alzheimers Dement. Oct 2020;16(10):1426-1437. doi:10.1002/alz.12138

58. Amariglio RE, Sikkes SAM, Marshall GA, et al. Item-Level Investigation of Participant and Study Partner Report on the Cognitive Function Index from the A4 Study Screening Data. J Prev Alzheimers Dis. 2021;8(3):257-262. doi:10.14283/jpad.2021.8

59. Sikkes SAM, De Rotrou J. A qualitative review of instrumental activities of daily living in dementia: what's cooking? Neurodegener Dis Manag. 2014;4(5):393-400. doi:10.2217/nmt.14.24

2.5 Supplementary Material

Supplemental Table 1 displays the characteristics of participants in the entire sample, as well as characteristics of participants from dyads and study partners, stratified by age group. The age groups are based on the participant's age, with the following groups: participants younger than 40 years of age, participants aged 40 through 49 years, participants aged 50 through 59 years, participants aged 60 through 69 years and participants aged 70 years and older.

Supplemental Table 1 | Participant and study partner characteristics, by age group.

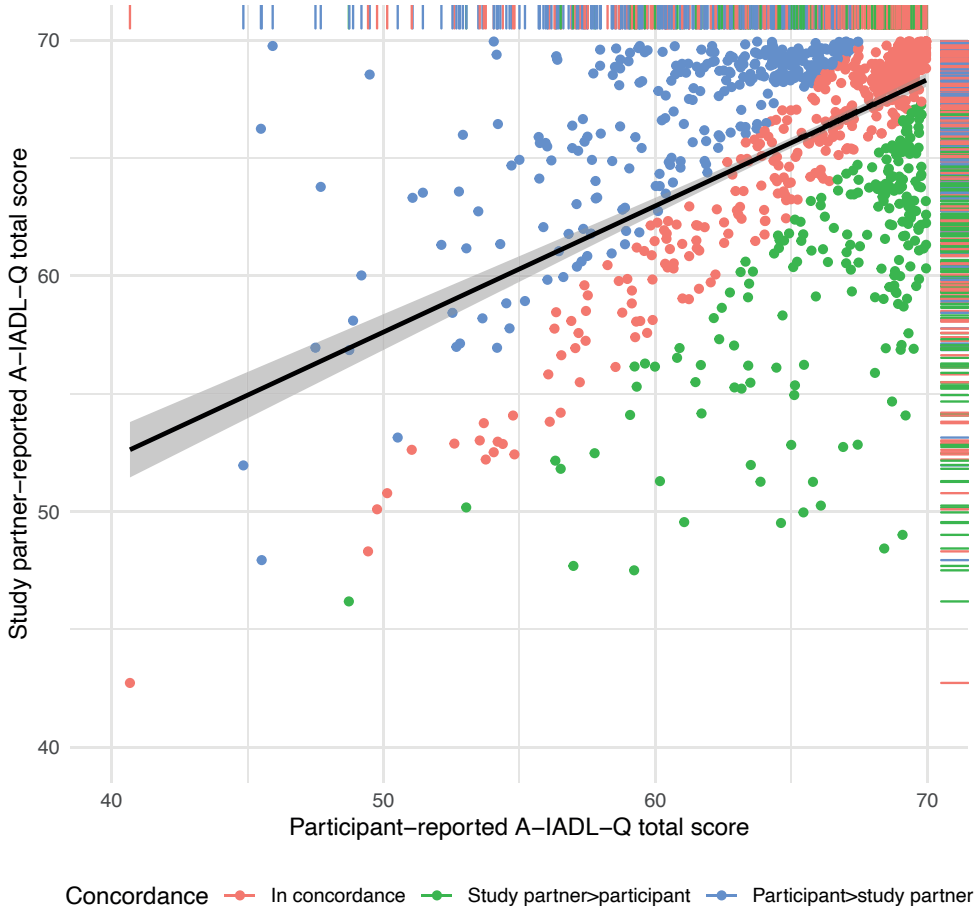
	Participants	**Dyads**	
		Participants	**Study partners**
Age groups, n (%)a	3,288 (100.0)	1,213 (100.0)	58.8 (14.2)
Under 40	189 (5.7)	46 (3.8)	38.3 (13.0)
40–49	238 (7.2)	63 (5.2)	46.6 (12.3)
50–59	833 (26.9)	295 (24.3)	51.4 (13.3)
60–69	1,224 (37.2)	504 (41.5)	61.5 (10.8)
Over 70	754 (22.9)	305 (25.1)	67.0 (12.6)
Female, n (%)	2,315 (70.4)	828 (68.3)	556 (45.8)
Under 40	149 (78.8)	37 (80.4)	18 (39.1)
40–49	179 (75.2)	49 (77.8)	24 (38.1)
50–59	696 (78.8)	217 (73.6)	118 (40.0)
60–69	871 (71.2)	362 (71.8)	204 (40.5)
Over 70	420 (55.7)	163 (53.4)	192 (63.0)
High level of education, n (%)	2,323 (70.7)	854 (70.4)	—
Under 40	169 (89.4)	43 (93.5)	
40–49	180 (75.6)	46 (73.0)	
50–59	647 (73.2)	209 (70.8)	
60–69	814 (66.5)	343 (68.1)	
Over 70	513 (68.0)	213 (69.8)	
A-IADL-Q score, mean (SD)	65.9 (4.8)	65.9 (4.7)	66.1 (4.6)
Under 40	67.1 (4.2)	67.3 (5.1)	68.1 (3.3)
40–49	66.1 (4.8)	66.0 (4.5)	67.1 (3.6)
50–59	66.3 (4.6)	66.2 (4.9)	66.6 (4.7)
60–69	65.8 (4.8)	66.0 (4.5)	66.2 (4.4)
Over 70	65.0 (4.9)	65.2 (4.7)	65.1 (4.9)
Memory complaints, n (%)	1,429/3,011 (47.5)	586/1,175 (49.9)	—
Under 40	36/175 (20.6)	4/46 (8.7)	
40–49	98/225 (43.6)	27/61 (44.3)	
50–59	355/818 (43.4)	143/285 (50.2)	
60–69	574/1,137 (50.5)	249/493 (50.5)	
Over 70	366/656 (55.8)	163/290 (56.2)	
Abnormal performance (≤-1.5) on COST-A, n (%)	218/2,945 (7.4)	83/1,149 (7.2)	—
Under 40	0/173 (0.0)	0/45 (0.0)	
40–49	7/223 (2.1)	1/60 (1.70)	
50–59	26/805 (3.2)	14/283 (4.9)	
60–69	77/1,103 (7.0)	23/484 (4.8)	
Over 70	108/631 (17.1)	45/277 (16.2)	

	Participants	Dyads	
		Participants	Study partners
GDS5, median (IQR)	0 (0–1)	0 (0–1)	—
Under 40	0 (0–1)	0 (0–1)	
40–49	0 (0–1)	0 (0–1)	
50–59	0 (0–1)	0 (0–1)	
60–69	0 (0–1)	0 (0–1)	
Over 70	0 (0–1)	0 (0–1)	
Dyads are spouses, n (%)	—	956 (78.8)	
Under 40		30 (65.2)	
40–49		47 (74.6)	
50–59		233 (79.0)	
60–69		424 (84.1)	
Over 70		222 (72.8)	
Duration relationship >10 years, n (%)	—	1,119 (92.5)	
Under 40		28 (60.9)	
40–49		56 (88.9)	
50–59		265 (89.8)	
60–69		479 (95.0)	
Over 70		291 (95.4)	
Living together, n (%)	—	960 (79.3)	
Under 40		29 (63.0)	
40–49		51 (81.0)	
50–59		250 (84.7)	
60–69		418 (82.9)	
Over 70		212 (69.5)	

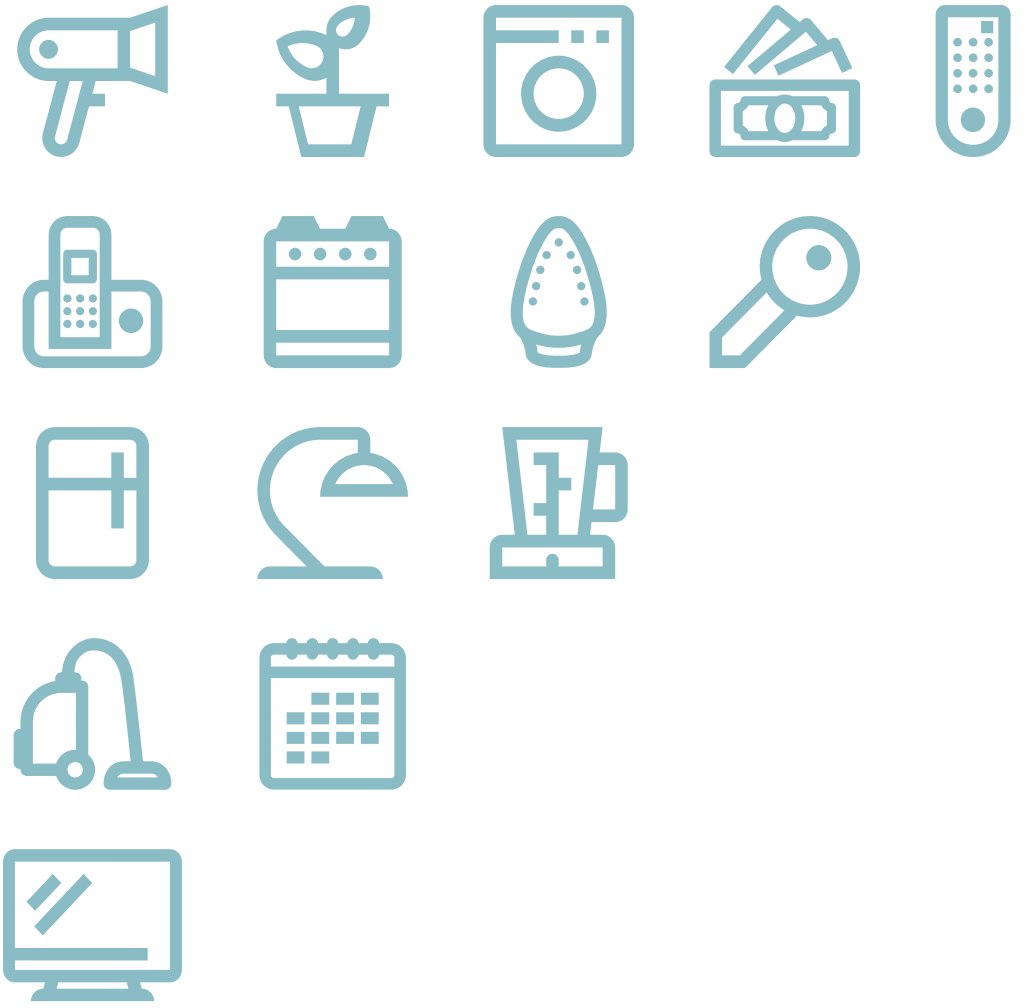
[a] For study partners, the table displays the mean age (SD).

Abbreviations: A-IADL-Q, Amsterdam Instrumental Activities of Daily Living Questionnaire; COST-A, Cognitive Online Self-Test Amsterdam; GDS5, 5-item Geriatric Depression Scale; IQR, interquartile range; SD, standard deviation.

Relationship between participant and study partner-reported IADL scores



Supplemental Figure 1 | Scatterplot showing the relationship (black line) between participant reported (horizontal axis) and study partner reported IADL functioning (vertical axis).
Each dot represents an individual; dots are colored based on a difference in Amsterdam IADL Questionnaire scores of ≥2.4 points: dyads in concordance are red, dyads where the study partner reported better Amsterdam IADL Questionnaire scores than the participant are green, dyads where the participant reported better Amsterdam IADL Questionnaire scores than the study partner are blue.



CHAPTER 3

EVERYDAY FUNCTIONING AND ENTORHINAL AND INFERIOR TEMPORAL TAU BURDEN IN COGNITIVELY NORMAL OLDER ADULTS

Mark A. Dubbelman, Justin Sanchez, Aaron P. Schultz, Dorene M. Rentz, Rebecca E. Amariglio, Sietske A.M. Sikkes, Reisa A. Sperling, Keith A. Johnson, Gad A. Marshall, on behalf of the A4 Study team

Published: June 16, 2022

In: Journal of Prevention of Alzheimer's Disease

Abstract

Background: Performance of cognitively complex "instrumental activities of daily living" (IADL) has previously been related to amyloid deposition in preclinical Alzheimer's disease.

Objectives: We aimed to investigate the relationship between IADL performance and cerebral tau accumulation in cognitively normal older adults.

Design: Cross-sectional.

Setting: Data was collected in the Anti-Amyloid Treatment in Asymptomatic Alzheimer's (A4) and Longitudinal Evaluation of Amyloid Risk and Neurodegeneration (LEARN) studies.

Participants: Participants (n = 447, age 71.9 ± 4.9 years, 57.5% female) who underwent tau positron emission tomography were selected from the A4 and LEARN studies.

Measurements: IADL performance was measured using the self- and study partner-reported versions of the Alzheimer's Cooperative Study Activities of Daily Living – Prevention Instrument (ADCS ADL-PI). We also investigated discordance between participants and their study partners. Cross-sectional associations between entorhinal and inferior temporal tau (independent variables) and ADCS ADL-PI total scores, item-level scores, and discordance (dependent variables) were investigated in linear and logistic regressions. Analyses were adjusted for age, sex and education and a tau by amyloid interaction was also included.

Results: Participants and their study partners reported high levels of IADL performance. Entorhinal and inferior temporal tau were related to study partner but not to self-reported total ADCS ADL-PI scores. The association was not retained after adjustment for global cerebral amyloid burden. At the item level, greater entorhinal tau was associated with study partner-reported difficulties remembering important dates (odds ratio (OR) = 1.24, 95% confidence interval (95%CI) = [1.06, 1.45, p = .008) and difficulties remembering the details of TV programs and movies (OR = 1.32, 95%CI = [1.08, 1.61], p = .007). Greater inferior temporal tau was associated with self-reported difficulties managing to find personal belongings (OR = 1.23, 95%CI = [1.04, 1.46], p = .018) and study partner-reported difficulties remembering the details of TV programs and movies (OR = 1.39, 95%CI = [1.11, 1.75], p = .005). Discordance between participant and study partner-report was more likely with

greater entorhinal (OR = 1.18, 95%CI = [1.05, 1.33], p = .005) and inferior temporal tau burden (OR = 1.29, 95%CI = [1.10, 1.51], p = .002).

Discussion: We found a cross-sectional relationship between study partner-reported everyday functioning and tau in cognitively normal older adults. Participants were more likely to self-report difficulties differently from their study partners when tau burden was higher. This may hint at an altered early-disease awareness of functional changes and underscores the importance of self-report of IADL functioning in addition to collateral report by a study partner.

3.1 Introduction

In prodromal stages of Alzheimer's disease (AD), difficulties with everyday activities that require higher-order cognitive functioning have been shown to increase over time.[1-5] These activities, referred to as 'instrumental activities of daily living' (IADL),[6] reflect cognition in daily life and are related to autonomy and quality of life. As such, IADL comprise an inherently clinically meaningful outcome. IADL performance in the prodromal stage has been related to both amyloid[7,8] and tau[9] accumulation, which form the two key components of the biological definition of AD.[10]

Before AD enters the prodromal stage, there exists a period of amyloid and tau accumulation in the absence of apparent clinical signs: the preclinical stage. Studies have shown that in this stage, higher amyloid burden seems to be associated with poorer IADL performance.[8,11,12] However, to our knowledge, only one study investigated the relationship between IADL performance and tau in the preclinical stage of AD; that study did not show a direct association between regional tau deposition and IADL in cognitively normal older adults. However, in individuals with elevated amyloid, greater regional tau burden was associated with greater IADL difficulties.[13]

IADL performance is usually rated by a study partner, but self-report in the preclinical stage may also be of value. There is evidence arguing for the utility of both self- and study partner-report, as both have been shown to be related to subjective memory concerns,[14] objective cognitive performance,[15,16] and future cognitive decline.[17,18] As such, both sources may provide valuable information and the combination of the two may be greater than the sum of its parts.

A previous investigation of IADL performance in cognitively normal older adults who participated in the Anti-Amyloid Treatment in Asymptomatic Alzheimer's Disease (A4)[19,20] and Longitudinal Evaluation of Amyloid Risk and Neurodegeneration (LEARN) Studies in relation to amyloid showed that worse IADL performance was associated with a higher amyloid burden.[11] In the present study, we aimed to investigate the association between IADL and cerebral tau. IADL performance was measured using the Alzheimer's Disease Cooperative Study Activities of Daily Living Prevention Instrument (ADCS ADL-PI) and tau burden was measured using flortaucipir positron emission tomography (PET). We hypothesized that difficulties with IADL performance would be associated with greater tau burden when assessed independently of amyloid, as well as in conjunction with amyloid.

3.2 Methods

Participants

We included participants from the A4 Study and its companion study, LEARN. The A4 Study cohort is described in more detail elsewhere.[20] Participants in A4 were selected for high cortical amyloid burden, while participants in LEARN had low amyloid. In brief, all participants were cognitively normal older adults between the ages of 65 and 85, who had a study partner who could provide collateral information about the participant's IADL performance. In addition to the A4 and LEARN inclusion criteria, to be included in the present study, participants also had to have undergone flortaucipir (tau) PET. We only used data from the baseline visit.

Assessments

IADL performance

Both participants and their study partners completed the Alzheimer's Disease Cooperative Study Activities of Daily Living Prevention Instrument (ADCS ADL-PI)[21] to assess IADL performance. The ADCS ADL-PI was previously found to have adequate reliability.[21] Each item was scored on a 4-point scale with response options of the participant "did not do the activity" (0), the participant did the activity "with a lot of difficulty" (1), "with a little difficulty" (2) or "as well as usual, with no difficulty" (3). Total (sum) scores for the 15-item version ranged from 0 to 45, with higher scores indicating better IADL performance. For the present study, item responses were dichotomized as the participant did the activity "with no difficulty" (0) or "with a little/a lot of difficulty" (1) because very few participants endorsed "a lot of difficulty". When the participant did not do the activity, or when the study partner did not know whether the participant did the activity, the item was considered missing. Missing item scores were prorated based on the individual's mean response on the non-missing items. This was done for a total of 152 items across all self-reported ADCS ADL-PI (2.3%) and 286 items across all study partner-reported ADCS ADL-PI (4.3%). Additionally, total scores were dichotomized as "without difficulty" (total score = 45, coded 0) or "with difficulty" (total score < 45, coded 1) because of a highly skewed distribution.

Based on a previous investigation in the larger A4/LEARN screening sample of 4,486 participants,[11] we focused on four items from the ADCS ADL-PI that showed a relatively high endorsement in the difficulty range: (a) remembering important

dates and times, (b) managing to find personal belongings at home, (c) following TV programs or movies and remembering the details, and (d) talking about and remembering current events.

Finally, we investigated discordance between participant and study partner by subtracting the study partner total score from the participant total score. The resulting discordance score potentially ranges from -45 (self-report < study partner-report) to +45 (self-report > study partner-report). We divided the discordance score into two groups: (1) in concordance (i.e., participant and study partner report the same score) and (2) in discordance (i.e., participant and study partner report different scores). We further distinguished between participants who self-reported less IADL difficulty (i.e., participant self-reported score is higher than the study partner-reported score, "participant underreport") and participants who self-reported more IADL difficulty (i.e., participant self-reported score is lower than the study partner-reported score, "participant overreport").

PET imaging

Tau burden was visualized in vivo with ^{18}F-flortaucipir PET using the A4 PET scanning protocol. Our analyses focused on two regions of interest: the entorhinal and inferior temporal cortices. These regions were selected based on prior findings with other IADL instruments.[13,22] For our primary analyses, we used non-partial volume corrected (non-PVC) standard uptake value ratios (SUVr) with cerebellar gray matter as reference region. To facilitate interpretation of analyses, SUVr were multiplied by 10 so that a one-unit increase in tau burden represents a 0.1 SUVr increase.

A global composite of cortical amyloid was obtained from ^{18}F-florbetapir PET SUVr, which used the whole cerebellum as the reference region, as previously described.[20] Global cortical amyloid was used as a continuous measure. Additionally, we dichotomized amyloid using a threshold SUVr of ≥1.15 to distinguish elevated from non-elevated amyloid.

Statistical analyses

R version 4.1.2 was used for all analyses.[23] Group differences were tested using t-tests or chi-squared tests, as appropriate. The Tukey Honest Significant Difference (HSD) correction for multiple comparisons was applied as necessary. Linear regressions

were fitted with the ADCS ADL-PI total scores for participant and study partner as the dependent variables, and continuous tau in entorhinal and inferior temporal cortices as the main independent variables. Subsequently, we employed logistic regressions to analyze the associations between tau and the dichotomized total and item ADCS ADL-PI scores. We additionally explored the interaction between tau and amyloid.

We analyzed agreement between participants and study partners in ADCS ADL-PI total scores using the intraclass correlation coefficient (ICC), focusing on differences in the raters' mean ratings. An ICC < 0.5 has been suggested to show poor agreement, an ICC between 0.5 and 0.75 shows moderate and an ICC > 0.75 shows good agreement.[24]

In logistic regressions, we first analyzed the relationship between cerebral tau burden and rater discordance, with the concordant group as the reference group. Then, in multinomial logistic regressions we further analyzed the discordance, distinguishing between participant under and overreport, again using the concordant group as the reference.

For all linear and logistic regressions, we report betas or odds ratios (OR), as appropriate, 95% confidence intervals (95%CIs) and *p*-values. All models were adjusted for age, sex, and education.

3.3 Results

A total of 447 participants (71.9 ± 4.9 years old; 58% female) were included in the present study, *n* = 392 from A4 and *n* = 55 from LEARN, based on availability of tau PET. Approximately one third of participants (*n* = 151, 34%) self-reported any IADL difficulties (ADCS ADL-PI self-reported total score median 45, range 37–45). Somewhat less than a third of study partners (*n* = 127, 28%) reported any IADL difficulties (ADCS ADL-PI study partner-reported total score median 45, range 35–45). Most participants (*n* = 346; 77.4%) had elevated global cortical amyloid. Those who self-reported difficulties were more likely to be male (*p* = .004), more likely to be of Asian descent (*p* = .020) and had a greater global composite amyloid SUVr (*p* = .009), compared to participants who self-reported to have no difficulties. Table 1 shows the characteristics of the sample.

Table 1 | Sample characteristics.

Characteristic	Whole group	No self-reported difficulty	Self-reported difficulty	P
N (%)	447 (100.0)	296 (66.2)	151 (33.8)	
Age in years	71.87 ± 4.9	71.63 ± 4.7	72.33 ± 5.2	0.155
Female sex, n (%)	257 (57.5)	185 (62.5)	72 (47.7)	0.004
Education in years	16.22 ± 2.8	16.07 ± 2.8	16.52 ± 2.9	0.117
Race, n (%)				
White	409 (91.5)	278 (94.9)	131 (88.5)	0.020
African American	11 (2.5)	7 (2.4)	4 (2.7)	
Asian	20 (4.5)	7 (2.4)	13 (8.8)	
Native American	1 (0.2)	1 (0.3)	0 (0.0)	
Not reported	6 (1.3)	4 (1.4)	4 (2.7)	
Ethnicity, n (%)				
Hispanic	8 (1.8)	4 (1.4)	4 (2.6)	0.554
Non-Hispanic	431 (96.4)	286 (96.6)	145 (96.0)	
Not reported	8 (1.8)	6 (2.0)	2 (1.3)	
ADCS ADL-PI* self-report, M (IQR)	45 (44–45)	45 (45–45)	44 (42–44)	< 0.001
ADCS ADL-PI* study partner-report, M (IQR)	45 (44–45)	45 (45–45)	45 (44–45)	< 0.001
Amyloid pet SUVr, global composite	1.28 ± 0.20	1.26 ± 0.18	1.31 ± 0.21	0.009
Tau pet SUVr, entorhinal cortex	1.18 ± 0.16	1.18 ± 0.16	1.18 ± 0.18	0.587
Tau pet SUVr, inferior temporal cortex	1.24 ± 0.13	1.23 ± 0.12	1.25 ± 0.15	0.139

All data are shown as mean ± standard deviation, unless noted otherwise.

* Shown here are the total scores based on the first 15 items of the ADCS ADL-PI.

Abbreviations: ADCS ADL-PI, Alzheimer's Disease Cooperative Study Activities of Daily Living Prevention Instrument; IQR, interquartile range; M, median; PET, positron emission tomography; SUVr, standard uptake value ratio.

Total scores

In the entire sample, study partner-reported ADCS ADL-PI total scores were significantly associated with both entorhinal tau (B = -1.13, 95%CI = [-1.89, -0.37], p = .004) and inferior temporal tau (B = -1.72, 95%CI = [-2.68, -0.76], p < .001). Self-reported ADCS ADL-PI total scores were not related to tau in either region (both p > .05, Table 2). The associations are visualized in Supplementary Figure 1. The relationship between entorhinal and inferior temporal tau and study partner-reported DACS ADL-PI total scores was not retained when correcting for global cerebral amyloid burden. There was no interaction between tau and amyloid, nor was amyloid associated with ADCS ADL-PI total scores in this sample when adjusting for tau. Table 2 shows the associations between tau and ADCS ADL-PI total scores, both with and without covarying for continuous amyloid. Results with dichotomized

amyloid were largely similar (see Supplemental Table 1). Results from the models with dichotomized ADCS ADL-PI total scores showed no associations with either entorhinal or inferior temporal tau (see Supplemental Table 2).

Table 2 | Coefficients and 95% confidence intervals from models with continuous ADCS ADL-PI scores.

	Self-report		Study partner-report	
	B [95%CI]	**P**	**B [95%CI]**	**P**
Entorhinal cortex				
Model 1: tau	-0.40 [-1.15, 0.36]	0.300	-1.13 [-1.89, -0.37]	0.004
Model 2: tau	0.24 [-0.27, 0.75]	0.349	0.17 [-0.34, 0.68]	0.515
Model 2: tau–amyloid interaction	-0.20 [-0.56, 0.17]	0.295	-0.19 [-0.56, 0.18]	0.310
Inferior temporal tau				
Model 1: tau	-0.59 [-1.55, 0.36]	0.222	-1.72 [-2.68, -0.76]	< 0.001
Model 2: tau	0.46 [-0.20, 1.11]	0.172	-0.19 [-0.85, 0.47]	0.568
Model 2: tau–amyloid interaction	-0.36 [-0.83, 0.10]	0.126	0.03 [-0.44, 0.50]	0.892

Model 1 includes only tau, model 2 includes tau, amyloid, and a tau–amyloid interaction. Both models are adjusted for age, sex, and education.
Abbreviations: B, estimate; 95%CI, 95% confidence interval.

Greater entorhinal tau burden was associated with greater odds for study partner-reported difficulties remembering important dates and times (OR = 1.24, 95CI = [1.06, 1.45], p = .008), as well as greater odds for study partner-reported difficulties following TV programs or movies and remembering the details (OR = 1.32, 95%CI = [1.08, 1.61], p = .007). Entorhinal tau was not associated with self-reported difficulties. A greater inferior temporal tau burden was associated with greater odds for study partner-reported difficulties following TV programs or movies and remembering the details (OR = 1.39, 95%CI = [1.11, 1.75], p = .005). Greater inferior temporal tau burden was associated with greater odds for self-reported difficulties managing to find personal belongings (OR = 1.23, 95%CI = [1.04, 1.46], p = .018). All odds ratios are shown in Table 3. The distributions of tau for no and some level of difficulty performing each of the four activities are shown in Figure 1, for both entorhinal tau (panel A) and inferior temporal tau (panel B) and both participant self-report (gold) and study partner-report (blue).

Item-level analysis

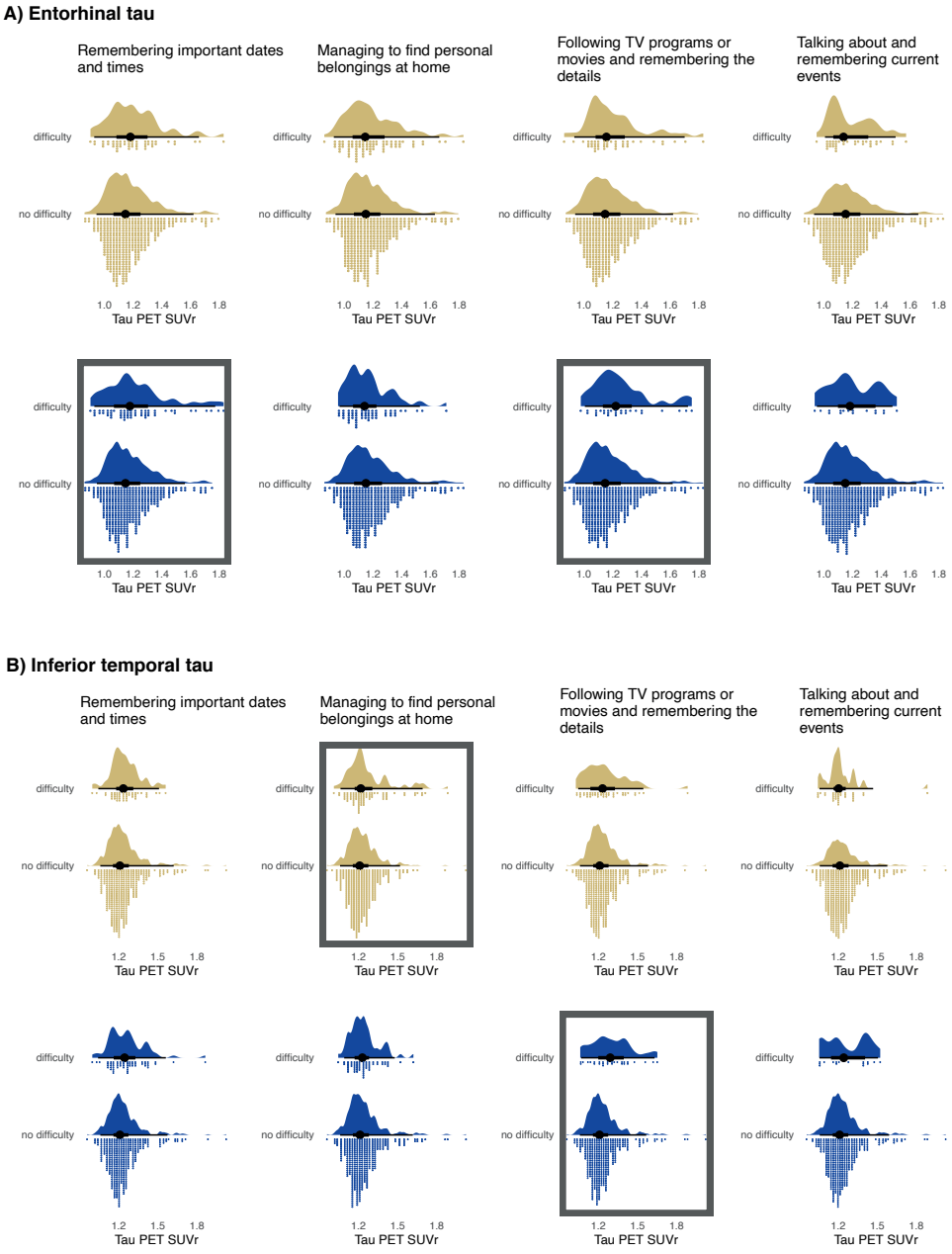


Figure 1 | Distributions and individual data points of entorhinal tau burden (panel A) and inferior temporal tau burden (panel B) set against the four ADCS ADL-PI items, as self-reported by the participant (gold) and the study partner (blue). Borders are placed around significantly different distributions.

Table 3 | Coefficients and 95% confidence intervals from models with continuous ADCS ADL-PI scores.

Activity	Entorhinal tau		Inferior temporal tau	
	OR [95%CI]	P	OR [95%CI]	P
Remembering important dates and times				
Self-report	1.12 [0.94, 1.32]	0.198	1.09 [0.88, 1.35]	0.431
Study partner-report	1.24 [1.06, 1.45]	0.008	1.21 [1.00, 1.46]	0.055
Managing to find personal belongings at home				
Self-report	1.02 [0.88, 1.18]	0.829	1.23 [1.04, 1.46]	0.018
Study partner-report	0.95 [0.80, 1.13]	0.562	1.03 [0.83, 1.28]	0.784
Following tv programs or movies and remembering the details				
Self-report	1.13 [0.95, 1.35]	0.181	1.13 [0.90, 1.41]	0.298
Study partner-report	1.32 [1.08, 1.61]	0.007	1.39 [1.11, 1.75]	0.005
Talking about and remembering current events				
Self-report	1.04 [0.85, 1.28]	0.687	0.94 [0.70, 1.25]	0.651
Study partner-report	1.17 [0.87, 1.56]	0.298	1.24 [0.90, 1.70]	0.188

All models are adjusted for age, sex, and education.
Abbreviations: OR, odds ratio; 95%CI, 95% confidence interval.

Participant and study partner discordance

There was a low intraclass correlation coefficient of 0.21 (95%CI = [0.12, 0.30]) for the concordance between participants and study partners in the total ADCS ADL-PI scores. Two hundred and fifty-four participants (56.8%) had the same score as their study partners, while 112 participants (25.1%) had lower scores (indicating more IADL difficulty) and 81 (18.1%) had higher scores than their study partners (indicating less IADL difficulty). Participants who self-reported less IADL difficulty than their study partner were more likely to be male ($\chi^2(1) = 12.52$, $p < .001$) and had a higher inferior temporal tau burden than participants who reported the same amount of difficulty as their study partner (Tukey HSD-adjusted $p = .019$).

Participants and their study partners were more likely to be in discordance when the participant had a greater entorhinal tau (OR = 1.18, 95%CI = [1.05, 1.33], $p = .005$) or inferior temporal tau burden (OR = 1.29, 95%CI = [1.10, 1.51], $p = .002$). This was not found for amyloid alone (OR = 1.71, 95%CI = [0.62, 4.70], $p = .296$). The discordance between dyads was significant in both the direction where the participant reported more IADL difficulty than their study partner and where the participant reported less IADL difficulty than their study partner. An overview of all model results can be

found in the Supplementary Material. Figure 2 shows the distributions of entorhinal and inferior temporal tau for the three groups.

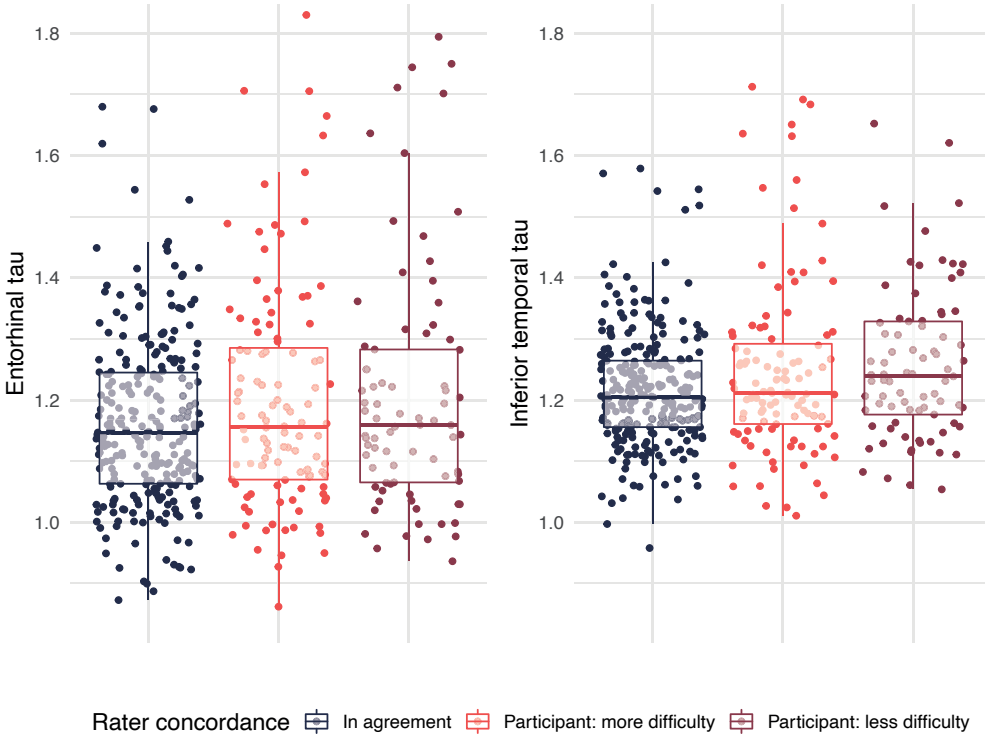


Figure 2 | Entorhinal (left) and inferior temporal tau distributions (right), stratified by rater difference (no difference [blue], participant self-reported more IADL difficulties than their study partner [red] or participant self-reported less difficulties than their study partner [brown]).

3.4 Discussion

In this study, we investigated the cross-sectional relationship between cerebral tau burden and the performance of higher-order cognitive everyday activities. We observed a relationship with overall IADL performance as reported by the study partners but not the participants themselves. We also found that difficulty performing specific activities may be increased when there is a greater regional cerebral tau burden. Furthermore, we observed that participants reported difficulties differently from their study partners when tau burden was greater.

First, as previously reported in a larger sample of cognitively normal older adults from the same cohort,[11] relatively few participants reported having difficulties carrying out the various tasks included in the ADCS ADL-PI. While a third of our sample reported at least some difficulty with one or more of the activities, the range in total scores was restricted and the ADCS ADL-PI showed a ceiling effect. Impairments in everyday functioning are uncommon in cognitively normal individuals, but more subtle difficulties carrying out activities do occur and have been reported in other cohorts as well.[14] Because changes in daily functioning may be considered an early marker of cognitive decline, it is relevant to investigate these subtle difficulties even in early stages of AD. As ceiling effects compromise analysis of effects, outcome measures that have a broader range in total scores among individuals who have no to very mild impairments are needed.

We observed a relationship between global IADL functioning as reported by a study partner and cerebral tau in the entorhinal and inferior temporal lobe. This relationship was not found with participant self-reported IADL functioning, nor with dichotomized self or study partner-reported total scores. A previous study showed that tau in the medial temporal and medial frontal regions was associated with IADL functioning in cognitively impaired individuals with elevated amyloid,[9] and it has been previously established that tau in various brain regions is related to clinical outcomes such as cognition.[25] Most individuals in our sample had elevated amyloid, yet we found inconsistent associations between tau burden and global IADL functioning, which were not retained after adjusting for amyloid. We also did not observe an interaction between tau and amyloid in the associations with IADL functioning. One possible explanation for these null findings is that difficulties with daily functioning in this sample may have been too subtle to detect, or that the instrument used, the ADCS ADL-PI, is not sensitive enough.

When looking at individual activities, we found more consistent associations. Both inferior temporal and entorhinal tau were related to a slightly increased odds of study partner-reported difficulties following a TV show or movie and remembering the details. Entorhinal tau was also related to study partner-reported difficulties remembering important dates and times, while inferior temporal tau was associated with self-reported difficulties managing to find personal belongings at home. These same items showed an association with amyloid in the larger A4/LEARN screening sample.[11] Unfortunately, we were unable to investigate the

interplay of tau and amyloid in the association with the performance of these activities due to infrequent endorsement of difficulty.

At the group level, participants and study partners reported overall similar levels of daily functioning, yet there was a low level of agreement between participants and their study partners. We divided our sample into participant–study partner dyads who were in concordance (reported the same level of daily functioning) and dyads who were not in concordance (reported different levels of daily functioning). We then further distinguished between dyads where the participant self-reported more difficulty and dyads where the participant self-reported less difficulty in daily functioning than their study partner. Slightly more than half the dyads were in concordance, while approximately a quarter of participants self-reported more difficulty and approximately a fifth of participants self-reported less difficulty than their study partners. Another study previously found a similar proportion of cognitively normal participants who reported more difficulties than their study partners.[14]

Employing the above concordance groups, we found an association between inferior temporal tau and discordance between participants and their study partners. Participants with a greater tau burden in both regions of interest were more likely to report IADL difficulties differently from their study partner. Interestingly, it seemed that both under and overreport by the participant were more likely with greater tau burden. The finding of participant overreport may hint at an early stage increase in awareness of functional impairments among those with more AD pathology: those with a greater tau burden report having more overall difficulties than study partners, albeit only slightly more difficulties. Conversely, the finding of participant underreport might reflect a decrease in awareness. Our findings add to a growing body of literature indicating that participants and study partners report differently on the participants' cognitive and functional performance, even when participants are cognitively normal.[14,26-29] As both self-reported and study partner-reported IADL functioning has been related to objective cognitive performance and future cognitive decline,[16,18] including both assessments in early disease stages may have added value and be beneficial for identifying those at risk for disease progression.

This study had a few limitations. The restricted range on the outcome measure, the ADCS ADL-PI, potentially reduced our power to detect associations, and the

ceiling effect might have inflated correlation coefficients. While we checked the assumptions for linear regression and found that they were sufficiently met, we additionally dichotomized the ADCS ADL-PI total scores. We could not replicate the findings from the linear models in the dichotomized models, suggesting that our results should be interpreted with caution. Finding a way to optimally assess the first, subtle change in daily functioning is an important challenge for future work. Further, our sample was predominantly non-Hispanic White and was highly educated, limiting the generalizability of our findings to individuals who do not match this profile. Recruitment of participants who represent the entirety of the population at-risk for AD should be a priority in the future. Future studies will also investigate the longitudinal association between AD biomarkers and daily functioning in the earliest stages of the disease. On the other hand, an important strength of our study was the large sample of cognitively normal older adults who underwent tau PET scans, combined with assessment of daily functioning as provided by the participant and a study partner.

In conclusion, we found evidence of a cross-sectional relationship between cerebral tau and study-partner reported difficulties with overall IADL, as well as both participant self and study partner-reported difficulties performing specific cognitively complex activities among cognitively normal older adults. Moreover, participants were more likely to self-report their difficulties differently from their study partners when tau burden was greater. These findings may hint at an altered awareness of functional changes among those with underlying AD pathology but who are still cognitively normal and underscore the importance of using assessments of IADL from multiple sources.

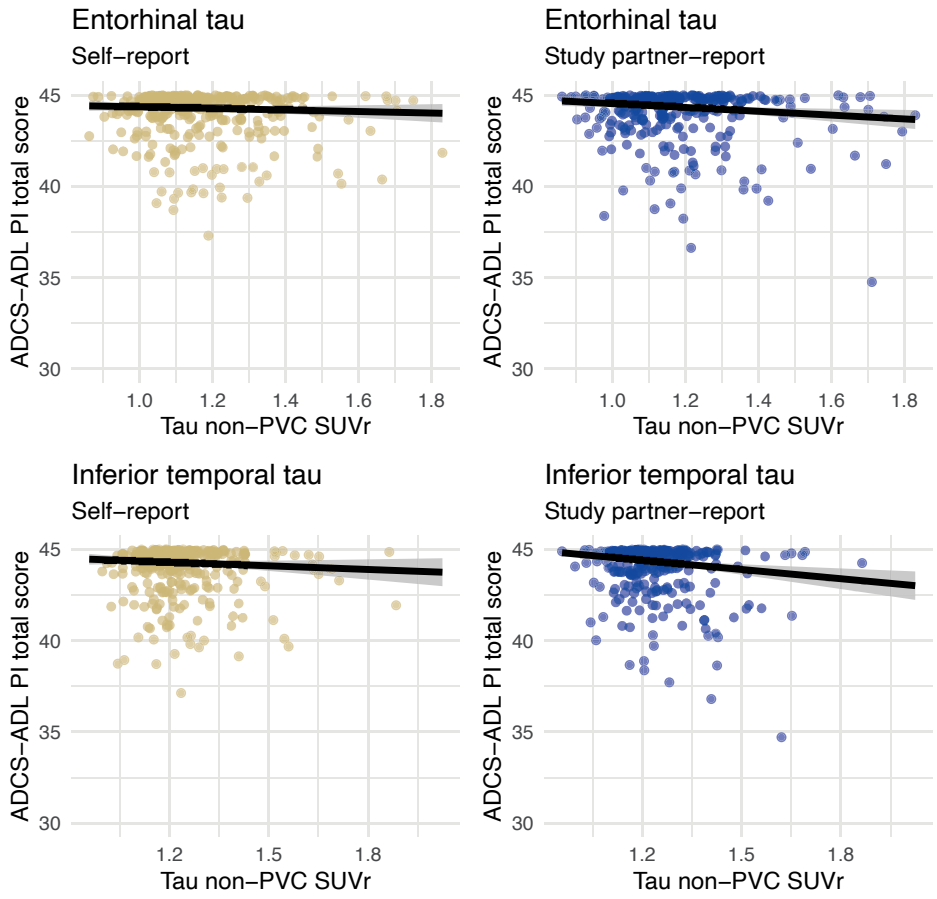
References

1. 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. Oct 1 2012;2(5):483-491. doi:10.2217/nmt.12.55
2. Marshall GA, Rentz DM, Frey MT, et al. Executive function and instrumental activities of daily living in mild cognitive impairment and Alzheimer's disease. Alzheimers Dement. May 2011;7(3):300-8. doi:10.1016/j.jalz.2010.04.005
3. Marshall GA, Aghjayan SL, Dekhtyar M, et al. Activities of daily living measured by the Harvard Automated Phone Task track with cognitive decline over time in non-demented elderly. J Prev Alzheimers Dis. 2017;4(2):81-86. doi:10.14283/jpad.2017.10
4. Roehr S, Riedel-Heller SG, Kaduszkiewicz H, et al. Is function in instrumental activities of daily living a useful feature in predicting Alzheimer's disease dementia in subjective cognitive decline? Int J Geriatr Psychiatry. Jan 2019;34(1):193-203. doi:10.1002/gps.5010
5. Dubbelman MA, Jutten RJ, Tomaszewski Farias SE, et al. Decline in cognitively complex everyday activities accelerates along the Alzheimer's disease continuum. Alzheimers Res Ther. 2020;doi:10.1186/s13195-020-00706-2
6. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. Autumn 1969;9(3):179-86.
7. Marshall GA, Olson LE, Frey MT, et al. Instrumental activities of daily living impairment is associated with increased amyloid burden. Dement Geriatr Cogn Disord. 2011;31(6):443-50. doi:10.1159/000329543
8. Lilamand M, Cesari M, Cantet C, et al. Relationship Between Brain Amyloid Deposition and Instrumental Activities of Daily Living in Older Adults: A Longitudinal Study from the Multidomain Alzheimer Prevention Trial. J Am Geriatr Soc. Oct 2018;66(10):1940-1947. doi:10.1111/jgs.15497
9. Marshall GA, Gatchel JR, Donovan NJ, et al. Regional Tau Correlates of Instrumental Activities of Daily Living and Apathy in Mild Cognitive Impairment and Alzheimer's Disease Dementia. J Alzheimers Dis. 2019;67(2):757-768. doi:10.3233/JAD-170578
10. Jack CR, Jr., Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement. Apr 2018;14(4):535-562. doi:10.1016/j.jalz.2018.02.018
11. Marshall GA, Sikkes SAM, Amariglio RE, et al. Instrumental activities of daily living, amyloid, and cognition in cognitively normal older adults screening for the A4 Study. Alzheimers Dement (Amst). 2020;12(1):e12118. doi:10.1002/dad2.12118
12. Tolbert S, Liu Y, Hellegers C, et al. Financial Management Skills in Aging, MCI and Dementia: Cross Sectional Relationship to 18F-Florbetapir PET Cortical beta-amyloid Deposition. J Prev Alzheimers Dis. 2019;6(4):274-282. doi:10.14283/jpad.2019.26
13. Gonzalez C, Tommasi NS, Briggs D, et al. Financial Capacity and Regional Cerebral Tau in Cognitively Normal Older Adults, Mild Cognitive Impairment, and Alzheimer's Disease Dementia. J Alzheimers Dis. 2021;79(3):1133-1142. doi:10.3233/JAD-201122
14. Verrijp M, Dubbelman MA, Visser LNC, et al. Everyday Functioning in a Community-Based Volunteer Population: Differences Between Participant- and Study Partner-Report. Front Aging Neurosci. 2021;13:761932. doi:10.3389/fnagi.2021.761932
15. Nosheny RL, Camacho MR, Insel PS, et al. Online study partner-reported cognitive decline in the Brain Health Registry. Alzheimers Dement (N Y). 2018;4:565-574. doi:10.1016/j.trci.2018.09.008
16. Tomaszewski Farias S, Giovannetti T, Payne BR, et al. Self-perceived Difficulties in Everyday Function Precede Cognitive Decline among Older Adults in the ACTIVE Study. J Int Neuropsychol Soc. Jan 2018;24(1):104-112. doi:10.1017/S1355617717000546

17. Farias ST, Mungas D, Jagust W. Degree of discrepancy between self and other-reported everyday functioning by cognitive status: dementia, mild cognitive impairment, and healthy elders. Int J Geriatr Psychiatry. Sep 2005;20(9):827-34. doi:10.1002/gps.1367
18. Nosheny RL, Jin C, Neuhaus J, et al. Study partner-reported decline identifies cognitive decline and dementia risk. Ann Clin Transl Neurol. Dec 2019;6(12):2448-2459. doi:10.1002/acn3.50938
19. Sperling RA, Rentz DM, Johnson KA, et al. The A4 study: stopping AD before symptoms begin? Sci Transl Med. Mar 19 2014;6(228):228fs13. doi:10.1126/scitranslmed.3007941
20. Sperling RA, Donohue MC, Raman R, et al. Association of Factors With Elevated Amyloid Burden in Clinically Normal Older Individuals. JAMA Neurol. Jun 1 2020;77(6):735-745. doi:10.1001/jamaneurol.2020.0387
21. Galasko D, Bennett DA, Sano M, et al. ADCS Prevention Instrument Project: assessment of instrumental activities of daily living for community-dwelling elderly individuals in dementia prevention clinical trials. Alzheimer Dis Assoc Dis. Oct-Dec 2006;20(4 Suppl 3):S152-69. doi:10.1097/01.wad.0000213873.25053.2b
22. Halawa OA, Gatchel JR, Amariglio RE, et al. Inferior and medial temporal tau and cortical amyloid are associated with daily functional impairment in Alzheimer's disease. Alzheimers Res Ther. Jan 31 2019;11(1):14. doi:10.1186/s13195-019-0471-6
23. R: A language and environment for statistical computing. Version 4.1.3. 2022. https://www.R-project.org/
24. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. Jun 2016;15(2):155-63. doi:10.1016/j.jcm.2016.02.012
25. Ossenkoppele R, Smith R, Mattsson-Carlgren N, et al. Accuracy of Tau Positron Emission Tomography as a Prognostic Marker in Preclinical and Prodromal Alzheimer Disease: A Head-to-Head Comparison Against Amyloid Positron Emission Tomography and Magnetic Resonance Imaging. JAMA Neurol. Aug 1 2021;78(8):961-971. doi:10.1001/jamaneurol.2021.1858
26. Hanseeuw BJ, Scott MR, Sikkes SAM, et al. Evolution of anosognosia in alzheimer's disease and its relationship to amyloid. Ann Neurol. Feb 2020;87(2):267-280. doi:10.1002/ana.25649
27. Vannini P, Hanseeuw BJ, Gatchel JR, et al. Trajectory of Unawareness of Memory Decline in Individuals With Autosomal Dominant Alzheimer Disease. JAMA Netw Open. Dec 1 2020;3(12):e2027472. doi:10.1001/jamanetworkopen.2020.27472
28. Cacciamani F, Houot M, Gagliardi G, et al. Awareness of Cognitive Decline in Patients With Alzheimer's Disease: A Systematic Review and Meta-Analysis. Front Aging Neurosci. 2021;13:697234. doi:10.3389/fnagi.2021.697234
29. Munro CE, Buckley R, Vannini P, et al. Longitudinal Trajectories of Participant- and Study Partner-Rated Cognitive Decline, in Relation to Alzheimer's Disease Biomarkers and Mood Symptoms. Frontiers in Aging Neuroscience. 2022;13 doi:10.3389/fnagi.2021.806432

3.5 Supplementary Material

Tau and total ADCS ADL-PI scores



Supplemental Figure 1 | Correlations between entorhinal tau (top row) and inferior temporal tau (bottom row) and ADCS ADL-PI total scores, as reported by the participant (left column) and their study partner (right column).

Abbreviations: ADCS ADL-PI, Alzheimer's Disease Cooperative Study Activities of Daily Living Prevention Instrument; PVC, partial volume corrected; SUVr, standard uptake value ratio.

Dichotomized amyloid

Supplemental Table 1 | Coefficients and 95% confidence intervals from models with continuous ADCS ADL-PI scores and dichotomized amyloid.

	Self-report		Study partner-report	
	B [95%CI]	**P**	**B [95%CI]**	**P**
Entorhinal cortex				
Tau	-0.04 [-0.30, 0.22]	0.755	-0.15 [-0.42, 0.11]	0.257
Amyloid	-0.23 [-3.29, 2.83]	0.882	-0.78 [-3.88, 2.31]	0.620
Tau–amyloid interaction	0.01 [-0.26, 0.29]	0.930	0.05 [-0.22, 0.33]	0.698
Inferior temporal cortex				
Tau	-0.18 [-0.50, 0.15]	0.283	-0.34 [-0.66, -0.01]	0.043
Amyloid	-1.70 [-5.74, 2.35]	0.411	-2.48 [-6.56, 1.60]	0.233
Tau–amyloid interaction	0.13 [-0.20, 0.47]	0.437	0.19 [-0.15, 0.53]	0.269

Tau and dichotomized IADL

Supplemental Table 2 | Odds ratio and 95% confidence intervals from models with dichotomized ADCS ADL-PI scores.

	Self-report		Study partner-report	
	OR [95%CI]	**P**	**OR [95%CI]**	**P**
Entorhinal cortex				
Model 1: Tau	1.04 [0.92, 1.17]	0.553	1.10 [0.97, 1.25]	0.131
Model 2: Tau	0.90 [0.39, 2.06]	0.794	0.92 [0.39, 2.19]	0.850
Model 2: Tau–amyloid interaction	1.05 [0.58, 1.92]	0.863	1.10 [0.59, 2.05]	0.762
Model 3: Tau	0.88 [0.57, 1.36]	0.571	1.07 [0.71, 1.62]	0.752
Model 3: Tau–amyloid interaction	1.17 [0.74, 1.83]	0.500	0.92 [0.59, 1.41]	0.688
Inferior temporal cortex				
Model 1: Tau	1.12 [0.96, 1.30]	0.147	1.15 [0.98, 1.35]	0.079
Model 2: Tau	0.68 [0.24, 2.00]	0.487	1.87 [0.61, 5.75]	0.273
Model 2: Tau–amyloid interaction	1.36 [0.64, 2.90]	0.427	0.68 [0.31, 1.50]	0.336
Model 3: Tau	1.13 [0.67, 1.89]	0.644	0.82 [0.49, 1.35]	0.432
Model 3: Tau–amyloid interaction	0.95 [0.56, 1.63]	0.858	1.24 [0.73, 2.11]	0.422

Model 1 includes only tau, model 2 includes continuous tau, continuous amyloid, and a tau–amyloid interaction, model 3 includes continuous tau, dichotomized amyloid, and a tau–amyloid interaction. All models are adjusted for age, sex, and education.
Abbreviations: OR, odds ratio; 95%CI, 95% confidence interval.

Participant and study partner discordance

Supplemental Table 3 | Odds ratio and 95% confidence intervals from models with participant and study partner discordance, divided into self-overreport and self-underreport.

	Self-overreport		Self-underreport	
	OR [95%CI]	**P**	**OR [95%CI]**	**P**
Entorhinal cortex				
Model 1: Tau	1.16 [1.01, 1.33]	0.034	1.22 [1.05, 1.42]	0.011
Model 2: Tau	0.51 [0.20, 1.33]	0.168	0.59 [0.21, 1.67]	0.320
Model 2: Tau–amyloid interaction	1.81 [0.90, 3.65]	0.096	1.75 [0.81, 3.76]	0.155
Model 3: Tau	0.80 [0.64, 1.88]	0.423	1.10 [0.64, 1.88]	0.732
Model 3: Tau–amyloid interaction	1.46 [0.84, 2.54]	0.178	1.12 [0.64, 1.97]	0.682
Inferior temporal cortex				
Model 1: Tau	1.25 [1.05, 1.50]	0.013	1.34 [1.11, 1.62]	0.003
Model 2: Tau	0.48 [0.20, 1.19]	0.114	1.99 [1.14, 3.47]	0.016
Model 2: Tau–amyloid interaction	1.97 [1.05, 3.71]	0.036	0.77 [0.52, 1.14]	0.191
Model 3: Tau	1.30 [0.68, 2.50]	0.434	2.04 [1.03, 4.04]	0.041
Model 3: Tau–amyloid interaction	0.95 [0.48, 1.88]	0.888	0.63 [0.31, 1.29]	0.205
Whole brain				
Model 4: Amyloid	1.98 [0.61, 6.46]	0.256	1.39 [0.36, 5.37]	0.631

Self-overreport refers to participants who reported more IADL difficulty than their study partner, self-underreport refers to participants who reported less IADL difficulty than their study partner. Concordance is the reference group. Model 1 includes only tau, model 2 includes continuous tau, continuous amyloid, and a tau–amyloid interactions, model 3 includes continuous tau, dichotomized amyloid, and a tau–amyloid interaction, and model 4 includes only global cortical amyloid. All models are adjusted for age, sex, and education.

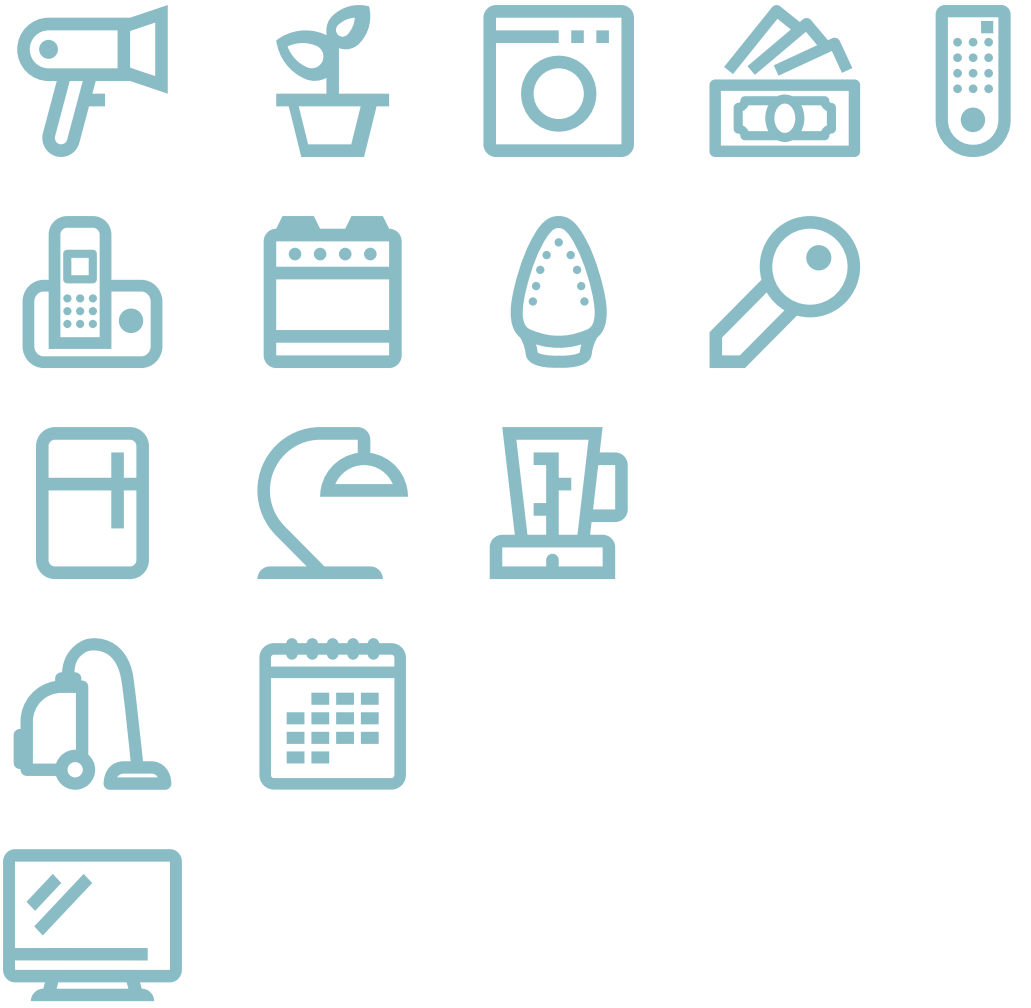
Abbreviations: OR, odds ratio; 95%CI, 95% confidence interval.

PART II

INFLUENCE OF BIAS IN THE MEASUREMENT OF EVERYDAY FUNCTIONING

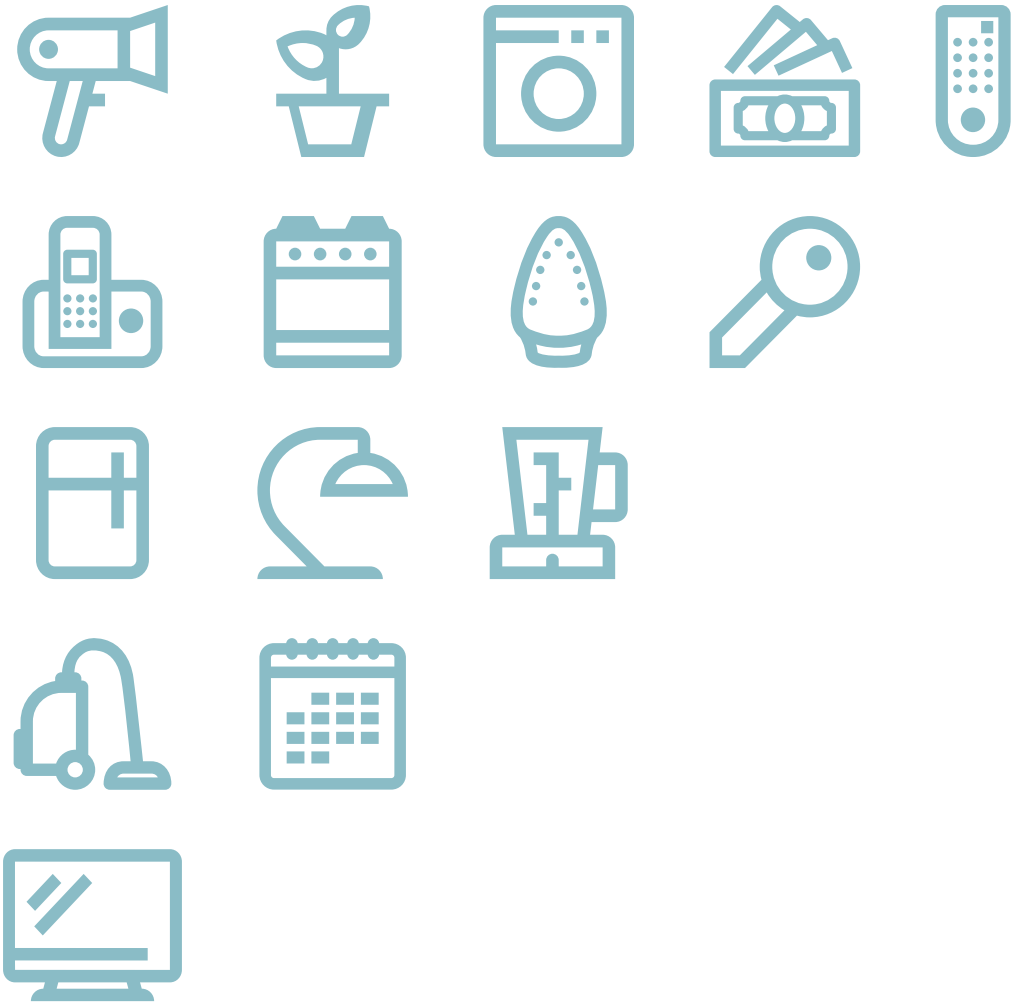
"They mistook my kindness for weakness
*I f****d up; I know that but, Jesus*
Can't a girl just do the best she can?
Catch a wave and take in the sweetness
Think about it, the darkness, the deepness
All the things that make me who I am
And who I am is a big-time believer
That people can change, but you don't have to leave her
When everyone's talking, you can make a stand"

– Lana Del Rey, 'Mariners Apartment Complex' (2019)



CHAPTER 4

MEASURING EVERYDAY FUNCTIONING ACROSS LANGUAGES AND CULTURES



CHAPTER 4.1

THE INFLUENCE OF DIVERSITY ON THE MEASUREMENT OF FUNCTIONAL IMPAIRMENT: AN INTERNATIONAL VALIDATION OF THE AMSTERDAM IADL QUESTIONNAIRE IN EIGHT COUNTRIES

Mark A. Dubbelman, Merike Verrijp, David Facal, Gonzalo Sánchez-Benavides, Laura J.E. Brown, Wiesje M. van der Flier, Hanna Jokinen, Athene Lee, Iracema Leroi, Cristina Lojo-Seoane, Vuk Milošević, José Luís Molinuevo, Arturo X. Pereiro Rozas, Craig Ritchie, Stephen Salloway, Gemma Stringer, Stelios Zygouris, Bruno Dubois, Stéphane Epelbaum, Philip Scheltens, Sietske A.M. Sikkes

Published: May 13, 2020

In: Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring

Abstract

Introduction: To understand the potential influence of diversity on the measurement of functional impairment in dementia, we aimed to investigate possible bias caused by age, gender, education, and cultural differences.

Methods: A total of 3,571 individuals (67.1 ± 9.5 years old, 44.7% female) from The Netherlands, Spain, France, United States, United Kingdom, Greece, Serbia, and Finland were included. Functional impairment was measured using the Amsterdam Instrumental Activities of Daily Living (IADL) Questionnaire. Item bias was assessed using differential item functioning (DIF) analysis.

Results: There were some differences in activity endorsement. A few items showed statistically significant DIF. However, there was no evidence of meaningful item bias: Effect sizes were low (ΔR^2 range 0–0.03). Impact on total scores was minimal.

Discussion: The results imply a limited bias for age, gender, education, and culture in the measurement of functional impairment. This study provides an important step in recognizing the potential influence of diversity on primary outcomes in dementia research.

4.1.1 Introduction

Impairment in cognitively complex "instrumental activities of daily living" (IADLs), such as doing grocery shopping, managing personal finances, and using mobile devices, may be one of the first symptoms of dementia.[1-3] IADL performance is related to quality of life, caregiver burden, and resource utilization.[4] Moreover, IADL impairment in preclinical stages might be a predictor of progression to dementia.[5,6] Therefore, functional impairment in an important and highly relevant outcome measure for clinical practice and clinical trials. In recently drafted industry guidelines, the U.S. Food and Drug Administration (FDA) recommended the use of functional impairment as a measure for effectiveness of treatment and of disease progression.[7] It is a potential global outcome measure in dementia research.[8,9]

Because everyday functioning relates to daily life, IADLs may be especially sensitive to bias caused by various factors, such as age, gender, and cultural differences. Previous studies have shown gender effects on traditional IADL instruments,[9-12] as they include predominantly household activities, which may be performed more often by women. Scientific literature concerning cultural and ethnoracial diversity in the context of dementia is scarce.[13,14] The selection of activities to include in an IADL instrument may be culture specific. For example, in the United States, it is customary to write checks, whereas in The Netherlands, people more often use online banking. Mere translation of an instrument does not always account for national (cross-cultural) disparities,[15,16] and although many functional instruments have been translated into numerous languages, there is no gold standard for cross-cultural adaptation of questionnaires.[17] This emphasizes the importance of investigating potential sources of bias and their influence on item and scale level.

We aimed to study the potential influences of diversity on the measurement of functional impairment using the Amsterdam IADL Questionnaire (A-IADL-Q). Specifically, we investigated item bias caused by various factors: cross-cultural differences (operationalized by using country of residence), age, gender, and education. We obtained data from eight Western countries: The Netherlands, Spain, France, United States, United Kingdom, Greece, Serbia, and Finland.

4.1.2 Methods

The present study included data from 3,571 individuals with a completed A-IADL-Q from memory clinics and cognition studies from eight countries: The Netherlands (Amsterdam Dementia Cohort[18] and European Prevention of Alzheimer's Dementia Longitudinal Study, EPAD[19,20]), Spain

Table 1 | Information about participants, in- and exclusion criteria, and information about the A-IADL-Q administration per included sample.

Study Name	**Amsterdam Dementia Cohort** [113]	**Compostela Aging Study** [49, 116]	**European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD)** [114, 115]	**Alfa+ Study** [117]	**INSIGHT PreAD** [118]	**Butler Alzheimer's Prevention Registry** [119]	**Socrates**	**Greek Association of Alzheimer's Disease and Related Disorders**	**Niš Clinic of Neurology** [120]	**Helsinki Small Vessel Disease Study**	**SAMS Project** [121]
Country	Netherlands	Spain	Spain (n = 218) France (n = 103) Netherlands (n = 88) United Kingdom (n = 71)	Spain	France	United States of America	France	Greece	Serbia	Finland	United Kingdom
Participants included	1,429	600	480	333	308	154	98	61	45	43	22
Age range	25–84 years	50–101 years	51–88 years	49–73 years	70–85 years	58–77 years	46–85 years	65–92 years	26–93 years	66–75 years	65–82 years
Research environment											
Recruitment	Consecutive memory clinic patients	MCI patients referred by GP	Participants from existing study cohorts	Mostly offspring of AD patients	Consecutive memory clinic patients & advertisement recruited	Advertisement recruited	Memory clinic patients	Patients from day center for dementia	Memory clinic patients	Patients with neuroimaging data selected from existing databank	Recruited from dementia research registry, memory clinic patients
Relevant inclusion and exclusion criteria	None	Cognitive complaints without dementia; Age ≥ 50 years	No dementia; Age ≥ 50 years	CN (MMSE ≥ 26, CDR 0); No neurological diseases; Age 45–74 years	CN (MMSE ≥ 27, CDR 0); Amyloid PET at baseline; No episodic memory deficits, no neurological diseases; Not living in nursing home; Age 70–85 years	CN or mild memory loss; No neurological diseases or dementia diagnosis; Age 55–85 years	Dementia-related diagnosis (MMSE ≥ 10); No neurological diseases other than dementia; Age 40–85 years	Dementia-related diagnosis; Reliable informant; No neurological diseases other than dementia; Age ≥ 65 years	CN, MCI, post-stroke cognitive impairment	No major neurological symptoms or psychiatric disease; Independence in basic ADL; No large infarcts, hemorrhages, contusion or tumor on MRI; Age 65–75 years	SCD (ECog ≥ 1.436 and answered "yes" when asked if "concerned they have a memory or other thinking problem"), MCI; Age ≥ 65 years
A-IADL-Q version	Original (n = 730) SV (n = 699)	Original	Original	SV	Original	SV	SV	Original (n = 28) SV (n = 33)	SV	SV	Version adapted from original

Study Name	Amsterdam Dementia Cohort [113]	Compostela Aging Study [49, 116]	European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD) [114, 115]	Alfa+ Study [117]	INSIGHT PreAD [118]	Butler Alzheimer's Prevention Registry [119]	Socrates	Greek Association of Alzheimer's Disease and Related Disorders	Niš Clinic of Neurology [120]	Helsinki Small Vessel Disease Study	SAMS Project 121
Clinical measures											
Participants (%) selected for validation*	1,369 (95.8)	300 (50.0)	480 (100.0)	333 (100.0)	308 (100.0)	154 (100.0)	—	26 (42.6)	45 (100.0)	43 (100.0)	22 (100.0)
Measures available	MMSE, CAMCOG, CDR, GDS	MMSE, CAMCOG, GDS	MMSE, CDR, GDS	MMSE, CDR	MMSE, CDR	MMSE, GDS	None	MMSE, CAMCOG, CDR, GDS	MMSE	MMSE, GDS	GDS

* Participants living in a nursing home were excluded from validation and no clinical measures were obtained for them.

Additional inclusion and exclusion criteria are available upon request.

Abbreviations: MCI mild cognitive impairment; GP general practitioner; AD Alzheimer's disease; CN cognitively normal; MMSE Mini-Mental State Examination; CDR Clinical Dementia Rating; ADL activities of daily living; MRI magnetic resonance imaging; SCD subjective cognitive decline; CAMCOG Cambridge Cognitive Examinations; GDS Geriatric Depression Scale

Compostela Aging Study;[21,22] EPAD; and Alfa+ project[23]), France (investigation of Alzheimer's predictors in subjective memory complainers (INSIGHT-preAD) study;[24] EPAD; and Socrates study), United States (Butler Alzheimer's Prevention Registry),[25] United Kingdom (EPAD; and software architecture for mental health self-management (SAMS) project),[26] Greece (Greek Association for Alzheimer's Disease and Related Disorders), Serbia (Niš Clinic of Neurology),[27] and Finland (Helsinki Small Vessel Disease study).

Participants had some degree of cognitive complaints or had an increased genetic or neurovascular risk for cognitive decline. Participants were recruited from memory clinics, through advertisement, or from existing databanks. Inclusion criteria ranged from being cognitively normal to having a dementia-related diagnosis. Other relevant inclusion and exclusion criteria for each cohort in this study can be found in Table 1. Participants provided written informed consent, and the studies were approved by their institutional review boards, which included, in each, consent for data sharing.

Measures

Amsterdam IADL Questionnaire

The Amsterdam IADL Questionnaire (A-IADL-Q) assesses cognitively complex IADLs that are prone to decline in incipient dementia. It covers a wide range of activities: the original version contains 70 items, while the short version (A-IADL-Q-SV) has 30. Both the original and short versions were used in the included studies. We analyzed both versions, with a special focus on the short version, because all items from the short version are also included in the original and can therefore be compared between all participants.

Unlike many other IADL instruments,[28] the A-IADL-Q has been validated extensively and has been shown to have good internal consistency, validity, and reliability.[29-31] Furthermore, it appears to be independent of age and gender,[30] and sensitive to change over time.[32] The short version was developed to create a more concise measure, as well as to reduce potential cultural bias by only including widely relevant activities.[33] International use of the A-IADL-Q is steadily increasing. All translations have gone through a cross-cultural adaptation process based on procedures described by Beaton et al.[34] in which experts and prospective users were asked to evaluate the translated instrument (a more detailed description of this process can be found in the Supplementary Material).

The questionnaire is scored using item response theory (IRT), as described elsewhere.[29,33] IRT assumes that an instrument measures a latent trait, which is represented in a scale ranging from total absence to abundance of the particular trait.[35] The A-IADL-Q latent trait is "IADL functioning".[30] In IRT, parameters are calculated for each item, which contain information about item response category location (or difficulty, i.e., at which trait level half the population endorses a given response category of an item), as well as slope (or discriminatory ability, i.e., how well an item can distinguish between people with lower and higher levels of the trait).

All A-IADL-Q items have five response categories, ranging from having "no difficulty" in performing an activity to being "unable to perform" an activity due to cognitive problems. IRT-based *T*-scores representing the trait level were calibrated in a memory-clinic population and were centered around a mean of 50 with a standard deviation (SD) of 10. Lower scores indicate more severe functional impairments.

Clinical measures

Mini-Mental State Examination (MMSE, scores range 0–30)[36] and Cambridge Cognition Examination (CAMCOG, scores range 0–107)[37] served as general indicators of cognitive functioning. For both measures, lower scores indicate worse cognition. The Clinical Dementia Rating (CDR)[38] was an indicator of functional status. A global CDR score of 0 represents no dementia, and scores of 0.5 to 3 are related to more advanced stages of dementia (and thus more functional impairment). Finally, the short form Geriatric Depression Scale (GDS, scores range 0–15)[39] was used to assess depressive symptoms. Data were not obtained for all included participants: We excluded individuals living in nursing homes (n = 130) because they have limited IADL independence.

Statistical analyses

We investigated item bias using 'differential item functioning' (DIF) analysis. DIF analysis is a technique for identifying items that have different item locations and/or slopes in different groups. DIF is assumed to occur when the relationship between a test item and the latent trait is not the same across study-irrelevant groups.[35] It is considered a variation in measurement and is therefore undesirable.[40] We studied DIF in the following groups: (1) nationality, using the Dutch cohorts as a reference group, while grouping all other studies by country; (2) men and women; and, based

on median split; (3) young (<67.2 years) and old age (≥67.2 years); and (4) low (<12 years) and high education (≥12 years).

For all DIF analyses, a minimum count of one case in at least two different response categories was required in each group for every item. We used the ordinal logistic regression (OLR) approach, which is often used and can be performed in standard software. OLR has been shown previously to be superior to the Mantel-Haenszel procedure.[41] We used the "lordif" package version 0.3-3 for R, developed by Choi et al.;[40] "lordif" has been used extensively in the literature, ensuring appropriateness and replicability of our procedures. In the OLR approach, a null model and three hierarchically nested models are created and compared for each item. When DIF is present and constant across all levels of the latent trait, it is called uniform DIF. The response categories of an item with uniform DIF are located at a different location in each group.[42] When an item is easier at one level of the trait and more difficult at another level, it is considered to have non-uniform DIF.[42] Items with non-uniform DIF have different discriminatory abilities in each group. Statistically significant DIF was determined on the basis of the likelihood-ratio $\chi 2$-test with an α-level of .01, to avoid type I error, and because multiple nested models are being tested for each item. Because of inflated type I error in OLR DIF analyses,[43] we added a step to establish presence of practically meaningful DIF,[44,45] based on a McFadden's pseudo *R2* ($\Delta R2$) value of .035 or larger. This approach reduces the risk of finding significant but negligible DIF, albeit at the cost of a reduction in power.[43] Furthermore, we used the following effect size criteria to quantify DIF size: $\Delta R2$ values between .035 and .070 for moderate, and above .070 for large DIF.[43] To refine DIF detection and effect size estimates, we then performed Monte Carlo simulations over 1,000 replications in which the detection criteria as well as effect size measures are computed repeatedly over simulated data based on the empirical datasets. The simulated data are generated under the hypothesis that there is no DIF, while keeping the observed group differences in trait levels.

As a means of construct validation, Pearson's *r* for continuous or Kendall's τ correlation coefficients for ordinal-level measures were calculated for the association between A-IADL-Q-SV *T*-scores and age, education level, gender of the participant, cognitive functioning (MMSE and CAMCOG), functional state (CDR), and mood (GDS).

Data were processed in SPSS Statistics version 22[46] and R version 3.6.1.[47]

4.1.3 Results

On average, participants were 67.1 ± 9.5 (mean ± SD) years old. Table 2 shows the demographics and clinical measures of all participants, stratified by country.

The overall mean score on the A-IADL-Q was 58.40 ± 14.2. A-IADL-Q scores per country are shown in Table 2.

Table 2 | Demographics and clinical characteristics for all participants, and grouped per country.

	All	The Nether-lands	Spain	France	United States	United King-dom	Greece	Serbia	Finland
Total n	3,571	1,515	1,151	509	154	93	61	45	43
Female, n (%)[a]	1,597 (44.7)	637 (42.0)	485 (42.1)	262 (51.5)	104 (67.5)	43 (46.2)	18 (29.5)	25 (55.6)	23 (53.5)
Age (years)	67.1 ±9.5	63.8 ±8.5	67.8 ±10.4	73.5 ±6.2	66.7 ±4.5	68.4 ±5.8	80.0 ±6.4	65.4 ±13.1	71.7 ±2.8
Education years	12.2 ±3.9	11.3 ±3.2	12.0 ±4.4	14.0 ±3.7	16.8 ±2.3	13.0 ±3.1	9.5 ±4.3	13.9 ±4.3	12.9 ±5.5
Dementia diagnosis, n (%)[a]	860 (29.9)	647 (47.2)	188 (20.2)	0 (0)	0 (0)	0 (0)	21 (80.8)	4 (8.9)	0 (0)
A-IADL-Q T-score[b]	58.40 ±14.2	51.54 ±11.7	61.82 ±15.2	67.33 ±9.4	67.48 ±3.5	71.16 ±5.1	39.48 ±13.9	61.67 ±8.8	66.30 ±5.2
Clinical measures[a]									
MMSE	26.20 ±4.6	24.22 ±5.0	27.76 ±3.7	28.62 ±1.2	29.35 ±1.0	28.46 ±1.5	19.58 ±4.6	27.49 ±3.6	27.60 ±2.2
CAMCOG	78.57 ±17.3	78.75 ±16.1	80.98 ±19.1	—	—	—	41.62 ±9.7	—	—
CDR, M (IQR)	0 (0–0.5)	0.5 (0–1)	0 (0–0)	0 (0–0)	—	0 (0–0)	2 (0.5–2)	—	—
GDS	3.66±3.6	3.80±3.3	4.09±4.0	4.33±4.2	0.85±1.3	3.52±4.5	2.38±3.1	—	2.10±3.1

All data are displayed as mean ± standard deviation, except as stated otherwise. "—" denotes that data were not available.

[a] Data were not obtained for all participants.

[b] The score shown is based on either the original or short version of the A-IADL-Q, as administered to each participant.

Abbreviations: A-IADL-Q, Amsterdam Instrumental Activities of Daily Living Questionnaire; CAMCOG, Cambridge Cognitive Examinations; CDR, Clinical Dementia Rating; GDS, Geriatric Depression Scale; IQR, interquartile range; M, median; MMSE, Mini-Mental State Examination.

Item endorsement

Generally, item endorsement was comparable between countries, as well as between men and women, younger and older participants, and participants with lower and higher education. Table 3 highlights a few activities in which there were apparent differences. "Minor repairs" was endorsed by a larger percentage of men, as compared to women. Conversely, "using a washing machine" was endorsed more often by women. Participants with a lower education endorsed "withdrawing cash from an ATM" somewhat less often than participants with a higher education. Older participants were less likely to work, compared to younger participants. Participants from Greece, Spain, and Serbia used computers less often than participants from the other countries. Participants from the United States appeared to use public transportation less often than participants from European countries (see Table 3).

Item bias

Due to restricted variability in some items, we were unable to analyze all items. Two hundred seventy-two of 300 items (90.7%) in the A-IADL-Q-SV were analyzed. Of the items analyzed, 26.6% had statistically significant DIF. Effect sizes were very small for all factors (ΔR^2 range .000–.034, see Figure 1). Monte Carlo simulations showed that the mean p-value for the χ^2-statistic across all items varied between comparisons from .006 to .012, which was close to the .01 α-level used to detect DIF. Simulation-based thresholds for effect size ranged from .001 to .018 across all analyses (Figure 1). Lowering the threshold would lead to more items being flagged for DIF. The effect sizes, however, remained very small.

For the original version, 437 of 490 items (89.2%) were analyzed. Of those, 20.4% had statistically significant DIF. The effects for age, gender, and education were again small (ΔR^2 range .000–.032). Four items showed meaningful DIF for nationality with a moderate effect. In Spain, "using the washing machine" (ΔR^2 = .043), "making appointments" (ΔR^2 = .064), and "playing card and board games" (ΔR^2 = .043) were flagged. All three items had uniform DIF: the first item was more difficult for Spanish individuals; the other two were easier, as compared to the Dutch reference group. The fourth item had non-uniform DIF and was found in the French group: "functioning adequately at work" (ΔR^2 = .064). The item appeared to be better at discriminating between people with lower and higher levels of functional impairment in France than in the Netherlands. We used the DIF results to re-estimate the T-scores for Spanish and French participants, thereby correcting

Table 3 | Differences in percentage endorsement of selected activities.

Activity	Country								Age		Sex		Education	
	The Netherlands	Spain	France	United States	United Kingdom	Greece	Serbia	Finland	Young	Old	Men	Women	Low	High
Minor repairs	46.2	57.9	67.4	55.8	62.4	55.7	57.8	72.1	53.6	55.8	**70.9**	**42.4**	52.8	58.8
Washing machine	58.4	70.7	77.0	92.2	75.3	63.9	71.1	81.4	72.4	63.5	**44.0**	**93.1**	65.7	73.1
Withdrawing cash from an atm	69.6	64.7	82.7	74.0	80.6	9.8	55.6	72.1	77.2	62.6	75.8	75.4	**66.2**	**82.0**
Working	52.3	42.4	54.2	66.9	24.7	9.8	53.3	58.1	**61.6**	**36.2**	53.1	50.9	47.4	55.4
Using a computer	**82.0**	**52.7**	**75.2**	**97.4**	**94.6**	**8.2**	**53.3**	**81.4**	82.1	60.3	79.7	73.3	65.3	84.3
Public transportation	**49.7**	**70.5**	**86.2**	**27.9**	**59.1**	**83.6**	**51.1**	**79.1**	58.0	64.5	59.0	66.5	55.4	68.0

Differences of interest between groups within each factor (country, age, sex, and education) are displayed in bold. Endorsement of other activities included in the Amsterdam IADL Questionnaire did not differ as much and these activities are not displayed here.

Figure 1 | DIF effect sizes for country, age, gender, and education in the A-IADL-Q-SV.
Green circles represent the empirically found ΔR^2 effect sizes; blue asterisks represent the 99th percentile ΔR^2 effect sizes from MC simulations. A solid green line is placed at the predetermined threshold for practically meaningful DIF (ΔR^2 = .035); a dashed blue line is placed just above the highest simulated effect size threshold.

for the effect of DIF. In the Spanish group, the mean score decreased by 0.16 points on the *T*-scale, in the French group the mean score decreased by 0.07 points. The largest individual differences in both countries (-1.14 and -1.33, respectively) corresponded to a difference of approximately one tenth of an SD and can therefore be considered negligible. Figure 2 shows the individual score changes after DIF correction in Spain and France. There was no meaningful bias for nationality in the other countries. Simulations showed the mean χ^2-statistic *p*-value across all items varied from .008 to .012. The largest ΔR^2 effect size was .026 (range .000–.026), which corresponds to a negligible effect.

Figure 2 | Scatter plot showing the differences between initial (uncorrected) and DIF-corrected *T*-scores for the A-IADL-Q in the French (red) and Spanish (blue) groups, plotted against the uncorrected *T*-scores.
A dashed line is placed at the mean change in score in the French and Spanish groups. Difference in total score ranges from -0.5 to +1.5 on the *T*-score, corresponding to approximately one tenth of a standard deviation difference. A solid black line is placed at no change.
Abbreviations: A-IADL-Q, Amsterdam Instrumental Activities of Daily Living Questionnaire; DIF, differential item functioning.

A-IADL-Q-SV construct validation

Overall, all correlations were in the same directions and of similar magnitudes as compared to the original validation data from The Netherlands.[30] Age seemed more strongly associated with IADL impairment in Spain (r = -0.47, 95% confidence interval [CI] = -0.51, -0.42), Greece (r = -0.31, 95%CI = -0.52, -0.06), and Serbia (r = -0.48, 95%CI = -0.68, -0.21) than in The Netherlands (r = -0.08, 95%CI = -0.31, -0.02). MMSE scores appeared to be less associated with IADL impairment in France (r = 0.11, 95%CI = 0.02, 0.21), United States (r = 0.12, 95%CI = -0.05, 0.27), and United Kingdom (r = -0.10, 95%CI = -0.33, 0.14), compared to the reference (r = 0.33, 95%CI = 0.28, 0.38). In these countries, the MMSE had a restricted score range. Conversely, MMSE scores were more strongly associated with IADL impairment in Serbia (r = 0.56, 95%CI = 0.32, 0.73). An overview of all correlations can be found in the Supplementary Material.

4.1.4 Discussion

In this study, we demonstrated that the influence of diversity on the measurement of IADL impairment, as measured with the A-IADL-Q, seems minimal. Although we found some differences with regard to activity endorsement between countries, there was no evidence of practically meaningful item bias caused by various factors, including age, gender, education, and culture. These findings, together with the similar associations with demographic, cognitive, and functional measures as found in earlier validation efforts,[30] further support the validity of the A-IADL-Q.

Addressing potential bias caused by various types of diversity is highly relevant in dementia research.[14] With respect to the measurement of functional impairment, there have been contradictory findings, with some studies showing a general comparability of IADLs across cultures and different ethnoracial groups,[8,9] and others reporting differences between cultures, genders, and ages.[48-51] For an optimal comparison of functional outcome in international studies and clinical trials, a valid, cross-culturally adapted instrument is crucial. In the present study, the relevance of addressing potential bias was underscored by the fact that we found some differences in activity endorsement, particularly in activities related to the household and to technology. Gender roles can differ between countries, and they might determine the IADL activities in which one participates. In Mediterranean countries, it seemed people used computers less often than in Northern European countries and America.

In our current sample, the effects of DIF were small and thus did not pass our threshold for practically meaningful DIF. The reason that we found little evidence of meaningful DIF may be attributed to the cross-cultural adaptation process that all translations went through, in which potential cross-cultural differences were identified beforehand, and cultural adaptations were made as necessary. These changes were minor, and we believe the items included should be applicable to Western culture in general. As part of the development of the short version, international experts provided feedback on the cross-cultural comparability of the items,[33] which may explain the absence of practically meaningful item bias for nationality. Because the A-IADL-Q-SV does not appear to have practically meaningful item bias, *T*-scores do not need to be adjusted to be compared across countries, ages, genders, or levels of education. This suggests that the A-IADL-Q yields valid and cross-culturally comparable estimations of functional decline. Previous studies[30,31,33] have already shown that A-IADL-Q scores are independent of age, gender, and education, and our findings corroborate this. This is an important

finding, because other functional instruments do appear to be bias for gender, age, and cultural differences.[48,49]

In the original version, a few items appeared to be biased in Spain and France. "Making appointments" had the largest DIF effect, and a potential explanation is that examples were added in the Spanish translation, because language experts indicated that the proposed translation for the word "appointments" (*citas*) could be interpreted as "(romantic) dates", whereas the intended definition was broader. However, adding examples may actually have restricted the interpretation of the question to the specific examples given, and led to a loss of the broader meaning. The other items with DIF had a smaller effect and no clear reason for the presence of DIF could be discerned. Despite the finding of item bias in the original version, the effect on the total scores was minimal.

The associations between A-IADL-Q-SV scores and demographic, cognitive, and functional measures we found here largely correspond to those previously described for the original version.[30] In Spain, Greece and Serbia, participants were older than average, and associations between age and IADL were stronger. In Spain, an association between age and IADL functioning was found earlier in a group of patients without dementia.[21] In France, the United States and the United Kingdom, the studies recruited mainly cognitively healthy participants, resulting in limited variation in the measure of cognition, and IADL functioning seemed to be less associated with cognitive measures.

An important strength of this study is that we used a data-driven approach to investigate the cross-cultural comparability of IADL. We used DIF, which is a powerful procedure to detect variance in measurement between groups on an item level and was possible as a result of the IRT scoring method. Not only does DIF tell us whether an item may be biased, but it also provides insight into the impact of the bias on the overall scores and it allows for correction. We additionally used simulations to further validate the empirical findings. These advantages allowed us to create a clear picture of possible measurement variance and impact on the instrument. Another strength of the study is that we included data from more than 3,500 individuals from eight countries. People with a wide variety of cognitive impairment-related diagnoses or complaints were included, ranging from subjective cognitive decline to dementia. Furthermore, the age of participants ranged from adulthood to old age. The large sample size and large variety in diagnoses and age contributes to the generalizability of our results and conclusions.

This study also had a few limitations. First, we included data from only eight developed, Western countries. Our findings cannot be generalized to other parts of the world. One study found DIF in an IADL instrument between different Asian cultures.[52] It should also be noted that we use the term "culture" to refer to each country's national culture. Furthermore, we did not have access to information about ethnicity or race. It is currently unclear what the influence of ethnoracial differences is on the measurement of IADLs. Second, our sample comprised mainly highly educated people. The group we defined as having low education still received up to 12 years of education. It is possible that different results would be obtained in samples with less formal education. Third, the sample size was relatively small in Finland, Serbia, and Greece. This may have reduced our power to detect DIF. We tried to address this issue by performing Monte Carlo simulations, which indicated that the predetermined cutoff for practically meaningful DIF may have been somewhat high. More items would show DIF, if the threshold was lowered. However, when considering how these findings influence the total scores, the impact seems minimal and the DIF effect sizes remain small.

The present study is an important first step in recognizing the influence of diversity on the measurement of functional impairment, and future studies should build on these findings. More research is needed to understand the differences between Western and Oriental and other cultures, as well as differences between ethnicities and races.

The A-IADL-Q-SV might be the preferred version for future international use, as it includes only the most broadly relevant everyday activities, does not seem to have meaningful item bias, has good construct validity, and is more pragmatic.

To conclude, we found no indication of the presence of clinically relevant bias caused by several aspects of diversity, including age, gender, education, and cultural differences. This is important, because it further underscores the potential of the A-IADL-Q, and the short version in particular, as an outcome measure of daily functioning in clinical practice and clinical trials.

Declarations

Acknowledgments

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.

The development of the Amsterdam IADL Questionnaire is supported by grants from Stichting VUmc fonds and Innovatiefonds Zorgverzekeraars. The Alzheimer Center Amsterdam is supported by Stichting Alzheimer Nederland and Stichting VUmc fonds. The present study is supported by a grant from Memorabel (#733050205), which is the research program of the Dutch Deltaplan for Dementia. The chair of W.M.F. is supported by the Pasman stichting. The clinical database structure for the Amsterdam Dementia Cohort was developed with funding from Stichting Dioraphte. D.F., C.L.B. and A.X.P.R. are supported by FEDER grant PSI2014-55316-C3-1-R, the Spanish National Research Agency grant PSI2017-89389-C2-1-R, and the Galician Government GI-1807-USC: Ref. ED431-2017/27. G.S.B., J.L.M. and the ALFA+ project have received funding from "la Caixa" Foundation (ID 100010434), under agreement LCF/PR/GN17/50300004 and the Alzheimer's Association and an international anonymous charity foundation through the TriBEKa Imaging Platform project (TriBEKa-17-519007). C.W.R. and the work for EPAD have received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking EPAD grant agreement #115736. S.E. is supported by a joint collaborative grant by the AP-HP and Inria. L.J.E.B., I.L. and G.S. were supported by the UK Engineering and Physical Sciences Research Council #EP/K015796/1. S.Z. is supported by the Robert Bosch Foundation Stuttgart within the Graduate Program People with Dementia in General Hospitals, located at the Network Aging Research, Heidelberg University, Germany. V.M. is supported by the Serbian Ministry of Education, Science and Technological Development, grant OI 173022. A.L. is partially supported by Institutional Development Award Number U54GM115677 from the National Institute of General Medical Sciences of the National Institutes of Health, which funds Advance Clinical and Translational Research (Advance-CTR). The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health. The other authors did not receive funding directly related to this work.

Conflicts of interest

The Amsterdam IADL Questionnaire was developed by S.A.M.S. and P.S., who were involved in the conception of the present study. The other authors report no conflicts of interest.

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. 5th ed. American Psychiatric Association; 2013.
2. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. Autumn 1969;9(3):179-86.
3. Peres K, Helmer C, Amieva H, et al. Natural history of decline in instrumental activities of daily living performance over the 10 years preceding the clinical diagnosis of dementia: a prospective population-based study. J Am Geriatr Soc. Jan 2008;56(1):37-44. doi:10.1111/j.1532-5415.2007.01499.x
4. Giebel CM, Sutcliffe C, Challis D. Activities of daily living and quality of life across different stages of dementia: a UK study. Aging Ment Health. Jan 2015;19(1):63-71. doi:10.1080/13607863.2014.915920
5. Luck T, Riedel-Heller SG, Luppa M, et al. A hierarchy of predictors for dementia-free survival in old-age: results of the AgeCoDe study. Acta Psychiatr Scand. Jan 2014;129(1):63-72. doi:10.1111/acps.12129
6. Weintraub S, Carrillo MC, Farias ST, et al. Measuring cognition and function in the preclinical stage of Alzheimer's disease. Alzheimers Dement (N Y). 2018;4:64-75. doi:10.1016/j.trci.2018.01.003
7. Early Alzheimer's disease: developing drugs for treatment - guidance for industry. 2018.
8. Truscott DJ. Cross-cultural ranking of IADL skills. Ethn Health. Feb 2000;5(1):67-78. doi:10.1080/13557850050007365
9. Nikula S, Jylha M, Bardage C, et al. Are IADLs comparable across countries? Sociodemographic associates of harmonized IADL measures. Aging Clin Exp Res. Dec 2003;15(6):451-9. doi:10.1007/BF03327367
10. Sikkes SAM, De Rotrou J. A qualitative review of instrumental activities of daily living in dementia: what's cooking? Neurodegener Dis Manag. 2014;4(5):393-400. doi:10.2217/nmt.14.24
11. Lechowski L, de Stampa M, Denis B, et al. Patterns of loss of abilities in instrumental activities of daily living in Alzheimer's disease: the REAL cohort study. Dement Geriatr Cogn Disord. 2008;25(1):46-53. doi:10.1159/000111150
12. Vergara I, Bilbao A, Orive M, Garcia-Gutierrez S, Navarro G, Quintana JM. Validation of the Spanish version of the Lawton IADL Scale for its application in elderly people. Health Qual Life Outcomes. Oct 30 2012;10:130. doi:10.1186/1477-7525-10-130
13. Chin AL, Negash S, Hamilton R. Diversity and disparity in dementia: the impact of ethnoracial differences in Alzheimer disease. Alzheimer Dis Assoc Dis. Jul-Sep 2011;25(3):187-95. doi:10.1097/WAD.0b013e318211c6c9
14. Babulal GM, Quiroz YT, Albensi BC, et al. Perspectives on ethnic and racial disparities in Alzheimer's disease and related dementias: Update and areas of immediate need. Alzheimers Dement. Dec 13 2018;doi:10.1016/j.jalz.2018.09.009
15. Mungas D, Reed BR, Farias ST, Decarli C. Age and education effects on relationships of cognitive test scores with brain structure in demographically diverse older persons. Psychol Aging. Mar 2009;24(1):116-28. doi:10.1037/a0013421
16. Parra MA. Overcoming barriers in cognitive assessment of Alzheimer's disease. Dement Neuropsychol. Apr-Jun 2014;8(2):95-98. doi:10.1590/S1980-57642014DN82000002
17. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. J Clin Epidemiol. Apr 2015;68(4):435-41. doi:10.1016/j.jclinepi.2014.11.021

18. van der Flier WM, Scheltens P. Amsterdam Dementia Cohort: Performing Research to Optimize Care. J Alzheimers Dis. 2018;62(3):1091-1111. doi:10.3233/JAD-170850

19. Ritchie CW, Molinuevo JL, Truyen L, Satlin A, Van der Geyten S, Lovestone S. Development of interventions for the secondary prevention of Alzheimer's dementia: the European Prevention of Alzheimer's Dementia (EPAD) project. Lancet Psychiatry. Feb 2016;3(2):179-186. doi:10.1016/S2215-0366(15)00454-X

20. Solomon A, Kivipelto M, Molinuevo JL, Tom B, Ritchie CW, EPAD Consortium. European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD LCS): study protocol. Bmj Open. Dec 2018;8(12)doi:10.1136/bmjopen-2017-021017

21. Facal D, Carabias MAR, Pereiro AX, et al. Assessing everyday activities across the dementia spectrum with the Amsterdam IADL Questionnaire. Curr Alzheimer Res. 2018;15(13):1261-1266. doi:10.2174/1567205015666180925113411

22. Facal D, Juncos-Rabadan O, Guardia-Olmos J, Pereiro AX, Lojo-Seoane C. Characterizing Magnitude and Selectivity of Attrition in a Study of Mild Cognitive Impairment. J Nutr Health Aging. 2016;20(7):722-8. doi:10.1007/s12603-015-0635-4

23. Molinuevo JL, Gramunt N, Gispert JD, et al. The ALFA project: A research platform to identify early pathophysiological features of Alzheimer's disease. Alzheimers Dement (N Y). 2016;2(2):82-92. doi:10.1016/j.trci.2016.02.003

24. Dubois B, Epelbaum S, Nyasse F, et al. Cognitive and neuroimaging features and brain beta-amyloidosis in individuals at risk of Alzheimer's disease (INSIGHT-preAD): a longitudinal observational study. Lancet Neurol. Apr 2018;17(4):335-346. doi:10.1016/S1474-4422(18)30029-2

25. Lee A, Alber J, Monast D, et al. The Butler Alzheimer's Prevention Registry: recruitment and interim outcome. Alzheimers Dement. 2017;13(7):622-623.

26. Stringer G, Brown L, Leroi I, Sikkes SAM. Capturing declining daily activity performance in a technologically-advancing older population: UK cultural validation of the Amsterdam IADL Questionnaire. 2016:

27. Petrovic J, Milosevic V, Zivkovic M, et al. Slower EEG alpha generation, synchronization and "flow"-possible biomarkers of cognitive impairment and neuropathology of minor stroke. PeerJ. 2017;5:e3839. doi:10.7717/peerj.3839

28. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, Scheltens P, Uitdehaag BM. A systematic review of Instrumental Activities of Daily Living scales in dementia: room for improvement. J Neurol Neurosurg Psychiatry. Jan 2009;80(1):7-12. doi:10.1136/jnnp.2008.155838

29. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. Alzheimers Dement. Nov 2012;8(6):536-43. doi:10.1016/j.jalz.2011.08.006

30. Sikkes SAM, Knol DL, Pijnenburg YAL, 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(1):35-41. doi:10.1159/000346277

31. Sikkes SAM, Pijnenburg YAL, 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. Dec 2013;26(4):244-50. doi:10.1177/0891988713509139

32. Koster N, Knol DL, Uitdehaag BM, Scheltens P, Sikkes SAM. The sensitivity to change over time of the Amsterdam IADL Questionnaire((c)). Alzheimers Dement. Oct 2015;11(10):1231-40. doi:10.1016/j.jalz.2014.10.006

33. Jutten RJ, Peeters CFW, Leijdesdorff SMJ, et al. Detecting functional decline from normal aging to dementia: Development and validation of a short version of the Amsterdam IADL Questionnaire. Alzheimers Dement (Amst). 2017;8:26-35. doi:10.1016/j.dadm.2017.03.002
34. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976). Dec 15 2000;25(24):3186-91.
35. Reise SP, Waller NG. Item response theory and clinical measurement. Annu Rev Clin Psychol. 2009;5:27-48. doi:10.1146/annurev.clinpsy.032408.153553
36. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research. 1975;12(3):189-198. doi:10.1016/0022-3956(75)90026-6
37. Huppert FA, Jorm AF, Brayne C, et al. Psychometric properties of the CAMCOG and its efficacy in the diagnosis of dementia. Aging Neuropsychol C. Sep 1996;3(3):201-214. doi:Doi 10.1080/13825589608256624
38. Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. Int Psychogeriatr. 1997;9 Suppl 1:173-6.
39. Yesavage JA, Sheikh JI. Geriatric Depression Scale (GDS) - Recent evidence and development of a shorter version. Clinical Gerontologist. 1986;5(1-2):165-173. doi:10.1300/J018v05n01_09
40. Choi SW, Gibbons LE, Crane PK. lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. Journal of Statistical Software. 2011;39(8)
41. Herrera AN, Gomez J. Influence of equal or unequal comparison group sample sizes on the detection of differential item functioning using the Mantel-Haenszel and logistic regression techniques. Qual Quant. Dec 2008;42(6):739-755. doi:10.1007/s11135-006-9065-z
42. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. Cambridge University Press; 2011.
43. Jodoin MG, Gierl MJ. Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. Appl Meas Educ. 2001;14(4):329-349. doi:10.1207/S15324818AME1404_2
44. Gelin MN, Zumbo BD. Differential item functioning results may change depending on how an item is scored: An illustration with the Center for Epidemiologic Studies Depression Scale. Educ Psychol Meas. Feb 2003;63(1):65-74. doi:10.1177/0013164402239317
45. Rouquette A, Hardouin JB, Vanhaesebrouck A, Sebille V, Coste J. Differential Item Functioning (DIF) in composite health measurement scale: Recommendations for characterizing DIF with meaningful consequences within the Rasch model framework. Plos One. 2019;14(4):e0215073. doi:10.1371/journal.pone.0215073
46. IBM SPSS Statistics for Windows. Version 22. IBM Corp.; 2013.
47. R: A language and environment for statistical computing. Version 3.6.1. 2019.
48. Fleishman JA, Spector WD, Altman BM. Impact of differential item functioning on age and gender differences in functional disability. J Gerontol B Psychol Sci Soc Sci. Sep 2002;57(5):S275-84. doi:10.1093/geronb/57.5.S275

49. Tennant A, Penta M, Tesio L, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care. Jan 2004;42(1 Suppl):I37-48. doi:10.1097/01.mlr.0000103529.63132.77

50. Berezuk C, Zakzanis KK, Ramirez J, et al. Functional Reserve: Experience Participating in Instrumental Activities of Daily Living is Associated with Gender and Functional Independence in Mild Cognitive Impairment. J Alzheimers Dis. 2017;58(2):425-434. doi:10.3233/JAD-161227

51. Sheehan C, Domingue BW, Crimmins E. Cohort Trends in the Gender Distribution of Household Tasks in the United States and the Implications for Understanding Disability. J Aging Health. Dec 2019;31(10):1748-1769. doi:10.1177/0898264318793469

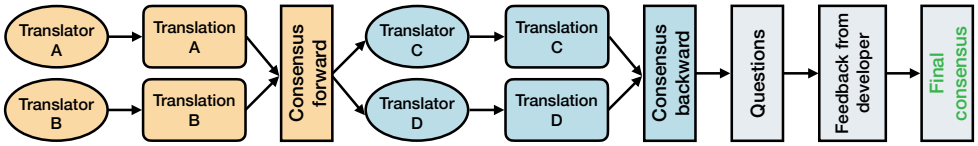
52. Niti M, Ng TP, Chiam PC, Kua EH. Item response bias was present in instrumental activity of daily living scale in Asian older adults. J Clin Epidemiol. Apr 2007;60(4):366-74. doi:10.1016/j.jclinepi.2006.07.012

4.1.5 Supplementary Material

Cross-cultural adaptation

To date, the Amsterdam IADL Questionnaire has been translated into thirteen languages following an extensive cross-cultural adaptation process. Some translations were made by researchers who wanted to use the questionnaire in a language that was not yet available, while others were made on request by ICON plc (https://www.iconplc.com/), a company specialized in the translation of clinical instruments.

The cross-cultural adaptation process was comprised of seven steps. First, two native speakers of the target language independently translated the questionnaire from either one of the two source languages (American English or Dutch) into the target language. Second, the two translations were reconciled into a single 'forward translation'. Any discrepancies between the two translations were discussed and a single translation was chosen. The forward translation was subsequently translated back into the source language by two new individuals. This step was performed to check whether the intended meaning of the instructions, questions and answer options were retained. Additionally, it allowed the developers to review translations in languages they do not speak. If needed, adjustments were made in the forward translation. The fourth step was a discussion of the forward and backward translations among the translators, the developer, and the translation project coordinator. This step should lead to a preliminary consensus translation (see Figure 1).



Supplemental Figure 1 | Translation process for the Amsterdam IADL Questionnaire.

Some activities included in the instrument were deemed to be less relevant for certain countries, e.g., 'preparing sandwiches' for Spain, or 'using the coffee maker' for the United Kingdom. Thus, minor changes were made to reflect the habits in the target population better: 'preparing sandwiches' was adapted to 'preparing a cold meal' in Spain and 'or making a pot of tea' was added to 'using the coffee maker' in the United Kingdom.

Subsequently, an expert committee, consisting of a small number of clinicians and knowledgeable professionals, was invited to review the translation for contents and clarity. The committee was asked to check whether the activities were clearly formulated and whether they correctly depicted the intended concepts. In the penultimate step, the translation project coordinator organized a pilot test with approximately ten caregivers of people with dementia. These caregivers should be native speakers of the target language and should not be experts in the field of questionnaires. In thinking-out-loud cognitive interviews, the caregivers were asked to explain how they would interpret and answer the questions.

In the seventh and final step, a consensus meeting was held with the translation project coordinator and the developer to discuss the feedback from the expert committee and cognitive interviews, and to address any potential alterations to the translation of the items. After completion of this process, the translation was considered to be cross-culturally adapted and suitable for use. Translations performed by ICON plc were in accordance with ISO 17100:2015 regulations, and a linguistic validation certificate was also made available.

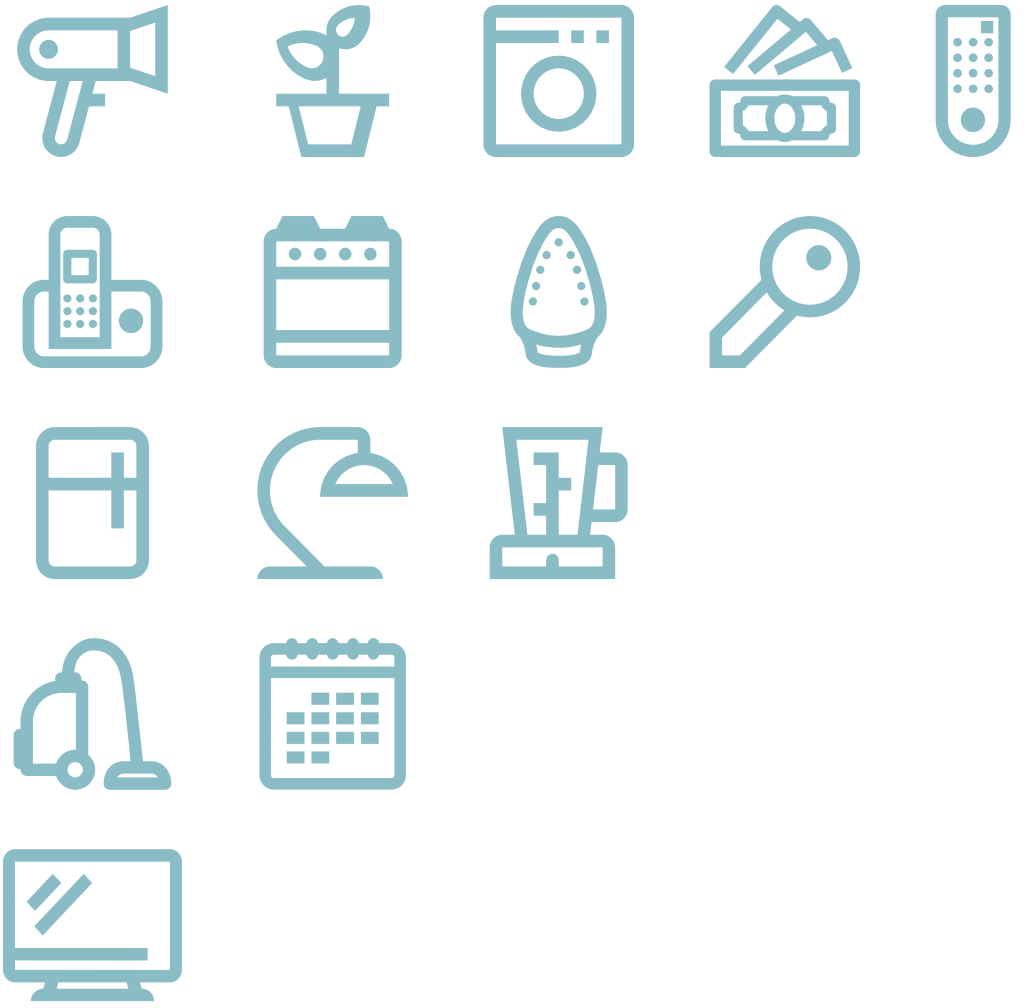
Correlations

Supplemental Table 1 | Pearson's r or Kendall's τ values and 95% confidence intervals for the correlations between A-IADL-Q-SV T-scores and demographic data and cognitive and functional measures, per country.

Country	Age	Edu-cation	Sex	MMSE	Cam-Cog	CDR	GDS
Netherlands	-.08 [-0.13,-0.02]	.09 [.04, .15]	.04 [-.02, .09]	.33 [.28, .38]	.33 [.28, .38]	-.45 [-.49,-.41]	-.11 [-.16,-.05]
Spain	-.47 [-.51,-.42]	.34 [.28, .40]	-.01 [-.08, .06]	.34 [.28, .40]	.50 [.41, .58]	-.13 [-.21,-.05]	-.04 [-.12, .05]
France	.02 [-.07, .10]	.09 [-.01, .18]	-.04 [-.13, .06]	.11 [.02, .21]	—	-.08 [-.17, .02]	-.14 [-.28, .00]
United States	.03 [-.14, .20]	.07 [-.09, .23]	-.06 [-.22, .09]	.12 [-.05, .27]	—	—	-.04 [-.20, .12]
United Kingdom	-.06 [-.26, .15]	.01 [-.20, .21]	-.20 [-.39, .00]	-.10 [-.33, .14]	—	.00 [-.24, .23]	-.13 [-.33, .08]
Greece	-.31 [-.52,-.06]	.06 [-.34, .44]	-.17 [-.52, .24]	.22 [-.18, .56]	.24 [-.17, .57]	-.44 [-.70,-.06]	.03 [-.36, .41]
Serbia	-.48 [-.68,-.21]	.10 [-.20, .38]	-.07 [-.35, .23]	.56 [.32, .73]	—	—	—
Finland	.01 [-.29, .31]	.12 [-.19, .41]	-.15 [-.43, .15]	.31 [.01, .56]	—	—	-.29 [.55, .02]

"—" denotes that data was not available.

Abbreviations: CAMCOG, Cambridge Cognitive Examinations; CDR, Clinical Dementia Rating; GDS, Geriatric Depression Scale; MMSE, Mini-Mental State Examination.



CHAPTER 4.2

CROSS-CULTURAL ADAPTATION AND VALIDATION OF THE AMSTERDAM INSTRUMENTAL ACTIVITIES OF DAILY LIVING QUESTIONNAIRE SHORT VERSION GERMAN FOR SWITZERLAND

Marina Bruderer-Hofstetter, Mark A. Dubbelman, André Meichtry, Florian Koehn, Thomas Münzer, Roos J. Jutten, Philip Scheltens, Sietske A.M. Sikkes, Karin Niedermann

Published: October 2, 2020

In: Health and Quality of Life Outcomes

Abstract

Background: Instrumental activities of daily living (IADL) limitations are associated with reduced health-related quality of life for people with mild cognitive impairment (MCI). For these people, the assessment of IADL is crucial to the diagnostic process, as well as for the evaluation of new interventions addressing MCI. The Amsterdam IADL Questionnaire Short Version (A-IADL-Q-SV) is an established assessment tool with good psychometric properties that has been shown to be robust to cultural differences in Western countries. The aims of this study were to: (1) cross-culturally adapt and validate the A-IADL-Q-SV for the German-speaking population of Switzerland; (2) investigate its cultural comparability; and (3) evaluate further psychometric properties.

Methods: The A-IADL-Q-SV German was pretested on clinicians and participants in a memory clinic setting. The psychometric properties and cultural comparability of the questionnaire were investigated in memory clinic settings including participants with MCI or mild dementia, as well as participants with normal cognition recruited from the community. Item response theory (IRT) was applied to investigate measurement invariance by means of differential item functioning to assess item bias. Additionally, the test–retest reliability on scale level, the construct validity through hypothesis testing and the discriminant validity of the A-IADL-Q-SV German were evaluated.

Results: Ninety-six informants of participants with normal cognition, MCI or mild dementia completed the A-IADL-Q-SV German. The basic assumptions for IRT scoring were met. No meaningful differential item functioning for culture was detected between the Swiss and Dutch reference samples. High test–retest reliability on scale level (ICC 0.93; 95%CI [0.90, 0.96]) was found. More than 75% of the observed correlations between the A-IADL-Q-SV German and clinical measures of cognition and functional status were found to be in the direction and of the magnitude hypothesized. The A-IADL-Q-SV German was shown to be able to discriminate between participants with normal cognition and MCI, as well as MCI and mild dementia.

Conclusions: The A-IADL-Q-SV German is a psychometrically robust measurement tool for a Swiss population with normal cognition, MCI, and mild dementia. Thus, it provides a valuable tool to assess IADL functioning in clinical practices and research settings in Switzerland.

4.2.1 Introduction

Instrumental activities of daily living (IADL) comprise the complex tasks needed to live independently in society.[1] Within the context of cognitive decline, IADL were defined as, "Complex activities with little automated skills for which multiple cognitive processes are necessary".[2]

Mild cognitive impairment (MCI) is a transient health state between normal cognition (NC) and dementia.[3] People with MCI experience cognitive and physical functioning impairments[3] and IADL limitations are frequent.[4] The latter are associated with reduced health-related quality of life[5] and are one of the defining features distinguishing MCI from NC.[6] They are predictive of the future development of dementia, both for people with MCI and NC.[7] Therefore, IADL performance is an important aspect of early cognitive diagnostics.[8]

Researchers are becoming increasingly aware of the importance of assessing IADL performance as a key outcome in intervention trials on older people with MCI and mild dementia (MD).[4] Improvements in IADL performance make a treatment meaningful for patients.[9] Furthermore, besides quality of life and self-efficacy, IADL performance is a prioritized treatment outcome for people with MCI and their caregivers.[10] To adequately assess the efficacy and effectiveness of IADL interventions, and to allow for comparison between studies, assessment tools with good psychometric properties (e.g., reliability, validity, sensitivity to change) are needed. Ideally, they are also robust across different languages and cultures.

To date, no gold standard exists for the assessment of IADL performance. Different methods of measurement are applied, i.e., performance-based assessments, self-rated and/or informant-rated questionnaires.[11] For people with early cognitive decline, informant-based questionnaires are the most accurate and convenient form of assessment.[12] However, the face validity of older, although well-known, questionnaires has been questioned, since they do not include activities with respect to technical appliances (e.g., computer use).[11] Additionally, commonly used IADL questionnaires have poor psychometric properties[13] and lack in sensitivity when classifying healthy aging, MCI and dementia.[2] Several self-reported and informant-reported IADL questionnaires have recently been developed to address these drawbacks. These questionnaires are sensitive to IADL limitations in the early stages of cognitive decline.[4]

The informant-based Amsterdam IADL Questionnaire (A-IADL-Q) was developed to assess IADL functioning.[14] It includes a wide range of IADLs covering all stages of cognitive decline in the setting of memory clinics. The A-IADL-Q has been validated in a Dutch cohort and demonstrated good psychometric properties,[15] as well as diagnostic value.[16] It was shown to be sensitive to capturing changes over time,[17] and also to be robust across cultural differences in a comparison between different Western countries, with regard to culture, sex, age and education.[18] The European Joint Program for Neurodegenerative Diseases Working Group has recommended the use of the A-IADL-Q for research and clinical purposes.[19] The original A-IADL-Q contains 70 items, while the short version (A-IADL-Q-SV) contains 30 items.[20] The questionnaire has been translated into thirteen languages, including German. The translation into German was made by ICON plc, a company specializing in the translation of clinical instruments. The translation process followed the steps recommended by Beaton et al.[21] This involved making two independent forward translations into the target language (i.e., German) followed by reconciliation into one version of the forward translation. Subsequently, two independent backward translations into the source language (i.e., Dutch) were made to check whether the intended meaning of the items, answer options and instructions had been retained. The translation process was finalized by a consensus meeting of the translators, the developer and translation project coordinator.[21] Although clinicians have already reviewed the translated German version, its cross-cultural validity in Switzerland has not yet been established.

Therefore, the aims of this study were to: (1) Adapt and validate the A-IADL-Q-SV German version cross-culturally, in order to be able to assess IADL performance in Switzerland of community-dwelling elderly people with NC, MCI and MD; (2) Further evaluate specific psychometric properties (i.e., measurement invariance, test–retest reliability, construct validity and interpretability).

4.2.2 Methods

Design

To obtain a final version of the A-IADL-Q-SV German version, we firstly pre-tested the translated questionnaire on clinicians and participants in a memory clinic setting to assess the comprehensibility of the translation, highlight any items that may be inappropriate at a conceptual level, and identify any other issues that may cause confusion, e.g., unclear wording.[21,22] This final version of the A-IADL-Q-

SV German was then evaluated in an observational study with two measurement time points. Data from the first measurement time point were used to investigate measurement invariance, construct validity and discriminant validity. Data from both the first and second measurement time points were used to investigate test–retest reliability.

The study was approved by the responsible ethics committee (EKOS, BASEC-NR. 2017-02200) and was conducted in accordance with the Declaration of Helsinki.

The A-IADL-Q-SV

The A-IADL-Q-SV contains 30 items and requires about 10 to 15 minutes for completion.[20] The questionnaire is adaptive and computerized, although it can also be administered on paper (with additional instructions necessary). In this study, the paper version of the questionnaire was used. All items are rated on a five-point scale, ranging from 'no difficulty' to 'unable to perform'; scoring is based on item response theory (IRT).[14,20] The A-IADL-Q and A-IADL-Q-SV have been found to meet all the basic assumptions of IRT scoring, based on a graded response model: (1) unidimensionality, which implies that one underlying latent trait determines the items (in this case, IADL functioning); (2) local independence, meaning the independence of item responses, conditional on the latent trait; and (3) monotonicity, meaning the probability of endorsing higher item categories as the trait level increases.[14,20] The IRT latent trait levels were transformed into a *T*-score that was calibrated to a memory clinic population, with a range from approximately 20 to 80, a mean of 50 and a standard deviation of 10, with higher *T*-scores indicating better IADL functioning.[14] The A-IADL-Q-SV was translated into German; work on this translation was not published before. All 30 items of the German questionnaire are the same as in the original version and are described in the Supplementary Material Table 1. The A-IADL-Q-SV German can be obtained from the developers after registration, and is free for use in all public health and non-profit agencies (https://www.alzheimercentrum.nl/professionals/amsterdam-iadl/).

Participants and sample size

Community-dwelling older persons of age >60 years and with NC, MCI, or MD, together with their informants, were included in the study. Informants could be relatives, close friends, or caregivers, who interacted closely enough with

the participant to be able to respond to the questionnaire. Exclusion criteria for participants were: 'moderate to severe' cognitive decline, based on the Mini-Mental State Examination (MMSE; <20) for participants with MCI or MD, and the modified Telephone Interview for Cognitive Status (TICS-m; <32) for participants with MD; cognitive decline due to causes other than Alzheimer's disease or vascular dementia (e.g., neurological diseases, trauma and people diagnosed with depression, alcohol, or drug misuse). Participants with probable MCI or MD were recruited from two memory clinics in the German-speaking region of Switzerland (Geriatrische Klinik St. Gallen; Psychiatrie St. Gallen Nord, Wil). General practitioners refer people with potential MCI or MD to a memory clinic for clarification of their cognitive complaints (i.e., dementia screening) as part of standard care. During these screening visits, a member of our study team gave people verbal and written information on the study, answered pending questions and obtained written informed consent. Participants with NC were recruited from the local community via flyers and advertisements distributed by the Pro Senectute St. Gallen organization and the Association of active older persons in the city and region of St. Gallen. Interested persons were prompted to contact the study team by e-mail or telephone. A member of the study team then provided verbal information to these interested persons, answered pending questions, and scheduled a call to check the eligibility criteria (e.g., TICS-m).

The targeted sample size to execute the cognitive debriefing/pretest was five clinicians and a minimum of five informants from people with MCI or MD to complete the A-IADL-Q-SV, with the option to recruit additional informants until no new issues or comments were raised. The targeted sample size for the evaluation of the A-IADL-Q-SV German version was 100 participants, based on the proposed COSMIN recommendations.[23] Firstly, a sample size of 50 participants is recommended for test–retest analyses, including the calculation of intraclass correlation coefficients (ICCs) (two measurements, targeted ICC of 0.8 with width 0.2 of the 95% confidence interval).[23] Secondly, a minimum of 50 participants is required (larger samples are recommended, e.g., 100 participants) for the investigation of the cross-cultural validity based on hypothesis testing by means of correlations.[23]

Procedures: Cognitive debriefing/pretest

Initially, five clinicians from a memory clinic were asked to give feedback on the A-IADL-Q-SV German. Issues discussed included answer options, activities or

sentences, and the grade of difficulty. As a result, small adjustments were made and documented. Such adjustments included e.g., the correction of spelling mistakes and grammatical inaccuracies, and specification of items (e.g., item 24: 'operating devices' into 'operating electronic devices').

Eight informants of people with MCI or MD completed the A-IADL-Q-SV German. The thinking-out-loud method was used, where informants were asked to write down their comments and issues on the relevance of each item, the applicability/meaning of the activities in Switzerland and the understandability of the questions. The results were reviewed to identify the necessity for translation modification (e.g., rewording of items/response options). Additionally, the completed questionnaires were examined to detect high levels of missing items or single responses. Minor adjustments were again made to the questionnaire and fully discussed with the developer. Points of discussion included the specification of items, e.g., item 20 'work' was supplemented with the specification 'paid or unpaid'; for item 11 'household appliances' the possibility of complementing it with examples was discussed but rejected because it may have influenced participants' responses. Accordingly, a final version of the A-IADL-Q-SV German was obtained.

Procedures: Validation and test–retest reliability of the A-IADL-Q-SV German

Measurements were performed in the memory clinics during the standard cognitive testing sessions for participants with MCI or MD. Each participant underwent an extensive cognitive screening procedure, including clinical and neuropsychological assessments, following international standards for dementia diagnosis.[24] During the same sessions, the informants completed the A-IADL-Q-SV German.

Interested participants from the community were contacted by telephone to check the inclusion and exclusion criteria. Thereafter, a cognitive impairment screening was performed, using the modified Telephone Interview for Cognitive Status (TICS-m).[25] An education-adapted score of ≥32 out of 50 points was required to qualify as not being subject to cognitive decline.[26] Their informants also completed the A-IADL-Q-SV German.

All informants were asked to complete the questionnaire a second time some 2 to 4 weeks later. Due to this short time interval, it was assumed that the cognitive

status remained stable and that a deterioration in the IADL performance was very unlikely.[27]

Additional clinical assessments

The following additional clinical assessments were used in this study:

The Mini-Mental State Examination (MMSE)—assesses global cognition (score range 0–30), with higher scores indicating better cognitive performance.[28] The MMSE is the most widely used global cognitive screening tool in clinical and research settings with sound psychometric properties.[19]

The Clinical Dementia Rating (CDR)—an assessment to stage the severity of dementia (score range 0–3), with higher scores indicating more severe stages of dementia.[29] The CDR is a recommended staging scale of dementia with high inter-rater reliability, good discriminant and concurrent validity.[30]

The Informant Questionnaire for Cognitive Decline in the Elderly (IQCODE)—assesses cognitive decline based on questions regarding cognitive performance (score range 1–5), with higher scores indicating worse performance.[31] The IQCODE is widely used as a screening test for dementia. It has been shown to measure a single factor of cognitive decline with high reliability and correlates with a wide range of cognitive tests.[32]

The Lawton Brody IADL scale—assesses performance in eight domains of IADLs (score range 0–8 for women; 0–5 for men), with higher scores indicating better performance.[33] To achieve comparability between subjects regardless of gender, in this study the scores were dichotomized into impaired = 1 (i.e., at least one considered activity with impairment) and not impaired = 0. The Lawton Brody IADL scale is one of the most frequently used IADL tools, with high reliability estimates. However, it has limitations due to content aspects (e.g., face validity), possibly due to its long existence.[2,19]

The Depression in old Age Scale (DIA-S)—is a relatively new screening tool to measure depression (score range 0–10); with scores >4 indicating probable pathological depression.[34] The DIA-S has been shown to have high discriminative power in terms of internal consistency and specificity compared to the Geriatric Depression Scale.[35]

Analysis

All analyses used the statistical software R (version 3.6.3)[36] and Mplus (version 7);[37] the α-level was set to 0.05.

Differences in the demographic characteristics of the included participants from the different settings (i.e., memory clinic setting, community) were investigated using Welch two sample t-test or Pearson's $\chi 2$-test, as appropriate.

The original A-IADL-Q-SV was fitted to a full graded response model on the basis of approximate marginal maximum likelihood estimation.[20] Unidimensionality of the A-IADL-Q-SV German was examined using confirmatory factor analysis (CFA) through investigating the factor structure (one-factor model).[14,20] Model fit to the full graded response model of the A-IADL-Q-SV German was evaluated with the comparative fit index (CFI > 0.90) and root mean square error of approximation (RMSEA < 0.05), as described elsewhere.[20] To further examine unidimensionality, we calculated a difference approximation to the second-order derivatives along the scree plot based on eigenvalue decomposition on the matrix of robust Spearman correlations between the items.[38] The resulting acceleration approximation indicates points of abrupt change along the scree plot, and the number before the point with the maximum acceleration value indicates the number of latent dimensions.[38] We assessed local independence by inspecting the residual correlation matrices, and considered residual correlations > 0.25 as indicative of potentially problematic item pairs,[20] and evaluated the monotonicity assumption using Mokken scale analysis.[39] Subsequently, measurement invariance was examined by means of differential item functioning (DIF) analysis for culture, comparing Swiss-German participants with the Dutch reference sample. The reference sample encompassed participants from the Amsterdam Dementia Cohort (n = 699).[40] No DIF, i.e., measurement invariance, in this context means that the items function identically in culturally different samples.[41] Uniform DIF is defined as a consistent difference between groups across the latent trait level, in this case IADL functioning. Non-uniform DIF occurs when an item is easier or more difficult for one group compared to the other at the same level of the latent trait.[23] Sufficient item endorsement, defined as at least two selected response categories per item, was required for DIF analysis.[18] The DIF analysis was based on ordinal logistic regressions: for every item, a null model and three hierarchically nested models were created and compared. Statistically significant DIF was determined based on the likelihood-ratio chi-square test with an alpha level of 0.01. To detect practically meaningful DIF, a cut-off on the change in McFadden's

pseudo *R2* of ≥0.035 was used.[42] We then performed Monte Carlo simulations over 100 replications to refine detection criteria as well as effect size measures. These are computed repeatedly over simulated data based on the empirical datasets.[43] For the DIF analyses, we used the 'lordif' package version 0.3-3, developed by Choi et al.[43]

Test–retest reliability was investigated on the scale level of the *T*-scores based on intraclass correlation coefficient (ICC) (ICC3,1, two-way mixed effects consistency model, single measurement),[58] overall and separately for the groups of participants with MCI/MD and with NC. The standard error of measurement (SEM) was calculated as the square root of the residual variance of the model and graphically depicted by a Bland and Altman plot.[23] Additionally, the smallest detectable change (SDM) was calculated using the formula $\pm 1.96 \times \sqrt{2} \times SEM$.[23] For interpretation of the SEM, it was compared to the total range of the *T*-scores (i.e., 20 to 80). Based on previous research on the A-IADL-Q[16,17,20] an SEM < 6 was interpreted as acceptable.

Construct validity was assessed by examination of Spearman's correlations between the A-IADL-Q-SV German and age, education, the MMSE, CDR, IQCODE, Lawton and Brody IADL Scale and DIA-S. Based on the results from previous studies on the A-IADL-Q, the hypotheses were stated quite specifically (Table 2).[15,17,20]

Discriminant validity was investigated to ascertain whether the A-IADL-Q-SV German version was able to discriminate between the three diagnostic groups of NC, MCI, MD. Differences in the *T*-scores between these groups were investigated using the Kruskall-Wallis rank sum test, followed by post hoc pairwise Wilcoxon tests. The Bonferroni-Holm method was applied to correct for multiple testing.

4.2.3 Results

In total, 96 community-dwelling elderly people were included, 56 (58%) from memory clinics and 40 (42%) from the community. The mean age of participants was 73.5 years (range 60–86 years); 44 (46%) were female; and for 93 (97%) of the participants the duration of their relationship with their informant was >10 years. Participants recruited from memory clinics were older, had a lower level of education and were more impaired on the A-IADL-Q-SV German than participants recruited from the community. Informants of the participants from memory clinics were less often a spouse and more often children. They lived apart from their informants more often compared to the participants recruited from the community and their informants. Details of demographic and clinical characteristics are summarized in Table 1.

Table 1 | Participant and tinformant characteristics.

	Participants (*n* = 96)				Informants (*n* = 96)		
		Memory clinic (*n* = 56)	Community (*n* = 40)	P-value	Memory clinic (*n* = 56)	Community (*n* = 40)	P-value
Age	74.08 (6.77)	77.04 (6.13)	69.95 (5.34)	< .001	64.73 (15.3)	64.50 (11.66)	.900
range	60–89	62–89	60–83		30–89	34–83	
Female (%)	44 (45.8)	25 (44.6)	19 (47.5)	.900	38 (67.8)	27 (67.5)	1.000
Level of education[a]							.
1	7 (7%)	6 (10%)	1 (2.5%)	< .010	5 (11%)	2 (5%)	400
2	54 (57%)	38 (67.5%)	17 (42.5%)		32 (57%)	19 (47.5%)	
3	24 (25%)	10 (18%)	14 (35%)		15 (27%)	16 (40%)	
4	5 (5%)	0 (0%)	5 (12.5)		3 (5%)	3 (7.5%)	
5	5 (5%)	2 (3.5%)	3 (7.5%)		1 (2%)	0 (0%)	
Diagnosis							
NC	43 (45%)	3 (6%)	40 (100%)		—	—	
MCI	27 (28%)	27 (48%)	0 (0%)				
MD	26 (27%)	26 (46%)	0 (0%)				
Relationship spouse	—	—	—		32 (57%)	31 (77.5%)	< .010
Duration (>10 years)	—	—	—		55 (98%)	38 (95%)	.700
Living together	—	—	—		32 (57%)	30 (75%)	< .010
A-IADL-Q-SV							
T-score	59.89 (9.29)	54.71 (8.37)	67.13 (4.41)	< .001			
Latent trait score	-0.99 (0.93)	-0.47 (0.84)	-1.71 (0.44)	< .001			
Clinical measures							
MMSE		25.05 (2.94)	—				
CDR, median (IQR)		0.5 (0.5)	—				
IQCODE		3.69 (0.51)	—				
Lawton Brody[b]		39 (74%)	—				
DIA-S, median (IQR)		2 (3)	—				
mTICS		—	37.0 (3.75)				

Values are means (standard deviation), medians (interquartile range) or frequencies. *P*-values based on Welch two sample *t*-tests or Pearson's chi-square tests.

[a] Level of education in accordance with the international standard classification of education: ISCED (1 = ISCED 2, 2 = ISCED 3–5, 3 = ISCED 6, 4 = ISCED 7, 5 = ISCED 8).

[b] Dichotomized impaired/non-impaired.

Abbreviations: A-IADL-Q-SV, Amsterdam Instrumental Activities of Daily Living Questionnaire Short Version; MMSE, Mini-Mental State Examination; CDR, Clinical Dementia Rating; IQCODE, Informant Questionnaire for Cognitive Decline in the Elderly; IQR, interquartile range; DIA-S, Depression in old age scale; mTICS, modified Telephone Interview for Cognitive Status; MCI, mild cognitive impairment; MD, mild dementia; NC, normal cognition

Measurement invariance

We checked the basic assumptions for IRT scoring. The Supplementary Material 3.2.1: Table 1 provides the graded response model estimates for item parameters and item information values in the reference sample. The CFI showed a good model fit (0.95), but the RMSEA (0.11, 95% CI [0.10, 0.12]) was indicative for borderline poor model fit. Several items had high inter-item correlations, probably due to restricted response variation. All items loaded significantly on the IADL factor (one factor model), confirming unidimensionality. Furthermore, the maximum acceleration value from the scree plot was at the first factor, confirming unidimensionality. No items violated the monotonicity assumption. A few item pairs showed a potential local dependence, possible due to restricted variability in item responses; details on these item pairs are presented in the Supplementary Material.

Figure 1a shows the distributions of the trait (theta), Figure 1b depicts the test characteristic curves for all items, and Figure 1c the test characteristic curves for the items with DIF for the Swiss sample and the Dutch reference sample. All items were sufficiently endorsed by both groups. The results from the likelihood-ratio chi-square tests indicated three items with statistically significant DIF: item 2 'doing the shopping'; item 20 'working'; item 23 'printing documents'; the item characteristic curves for these items are depicted in the Supplementary Material. Items 2 and 23 showed uniform DIF, with item 2 being easier and item 23 being more difficult in the Swiss sample compared to the Dutch reference sample. Item 20 showed non-uniform DIF. However, effect sizes (change in McFadden's pseudo R^2) were negligible (i.e., $R^2 < 0.035$; for item 2 $R^2 = 0.008$, item 20 $R^2 = 0.02$, item 23 $R^2 = 0.015$), suggesting that there was no practically meaningful item bias. All chi-squared values and ΔR^2 values for the logistic regressions obtained from the empirical data used for the DIF analyses are presented in Supplementary Material. Monte Carlo simulations confirmed that the a priori cut-offs we used, were appropriate. The Monte Carlo simulations-based cut-off for chi-squared p-values and ΔR^2 values can be found in Supplementary Material. We corrected for DIF by means of a re-estimation of the *T*-scores in the Swiss sample based on the DIF results. The mean *T*-score increased by 0.38 points, and the largest individual change was an increase of 2.17, corresponding to approximately one-fifth of a SD change.

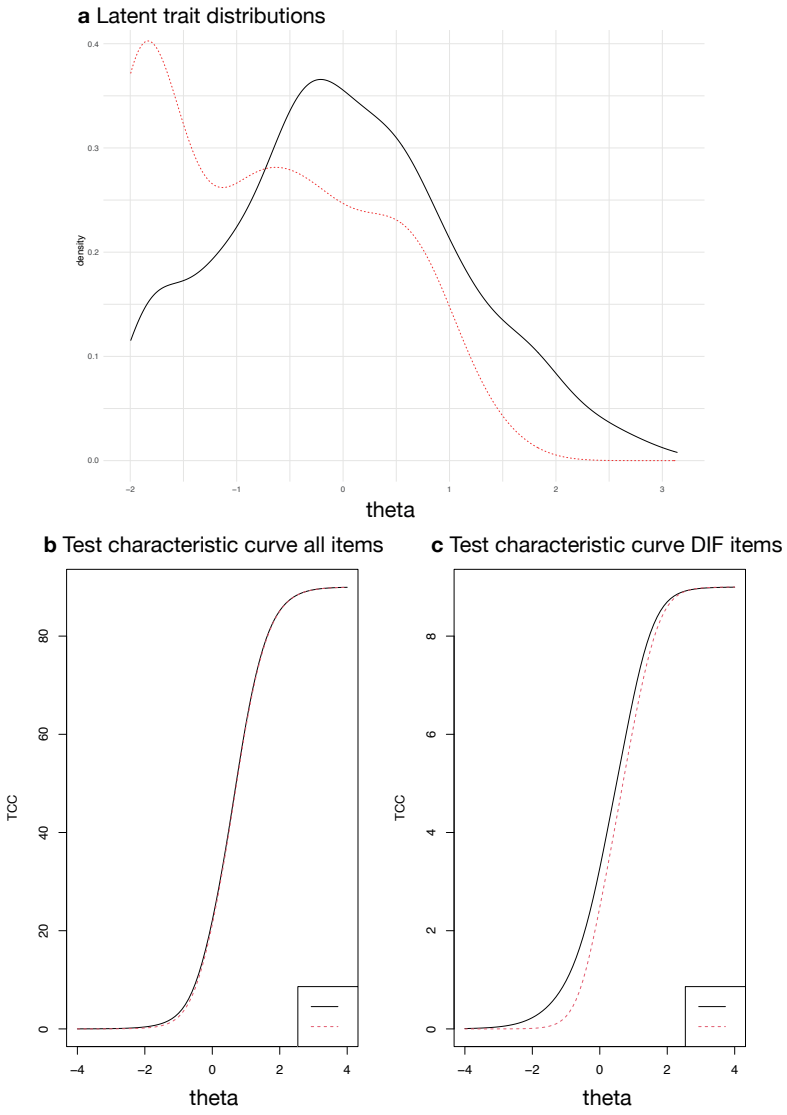


Figure 1 | Latent trait distributions and test characteristic curves.
The red dashed lines show the Swiss population and the solid black lines the Dutch reference group. Panel a: Latent trait distributions for the Dutch reference sample and the Swiss sample. Panel b: Test characteristic curves including all items for the Dutch reference sample and the Swiss sample. Panel c: Test characteristic curves including the three items with differential item functioning for the Dutch reference sample and the Swiss sample.

Test–retest reliability and measurement error

Of the included 96 informants, 82 (85%) completed the A-IADL-Q-SV German for the second time, with a median of 23 days between the two measurement time points; two questionnaires were excluded because a different informant had completed the second questionnaires, resulting in the inclusion of 80 questionnaires in the analysis. An overall ICC of 0.93 (95%CI = [0.90, 0.96]) was observed. The SEM, by means of classical test theory, was 2.4 and the smallest detectable change was 6.6 (95%CI = [5.3, 7.9]). The range of the *T*-score was 39.7. The corresponding Bland and Altman plot is depicted in Figure 2. The mean difference between the two measurements was 0.4 (95%CI = [-0.4, 1.2], p = 0.29); the lower limit of agreement was -6.2 (95%CI = [-7.5, -4.9]) and the upper limit of agreement 7.0 (95%CI = [5.7, 8.3]). The Bland and Altman plot shows that the data for the group of participants with NC (higher level of IADL functioning) has less variance. Furthermore, residual analysis showed that the data did not conform to model assumptions (i.e., homoscedasticity and normal distribution of residuals).

The separate ICCs for the subgroups of participants with MCI/MD (n=41) and with NC (n=39) were also estimated. For the MCI/MD subgroup, an ICC of 0.86 (95%CI = [0.77, 0.91]) was observed, compared to the NC subgroup with an ICC of 0.92 (95%CI = [0.86, 0.95]). Subsequently, the SEM, SDC and Bland and Altman analyses for the subgroups of participants with MCI/MD and NC participants were investigated separately. The SEM in the MCI/MD subgroup was 3 and the SDC 8.4 (95%CI = [6.1, 8.4]). The Bland and Altman plot for the MCI/MD subgroup is depicted in Figure 3a. There was no evidence of violation of model assumptions. The mean difference of the two measurements was 1.1 (95%CI = [-0.2, 2.5], p = 0.93); the lower limit of agreement was -7.2 (95%CI = [-9.6, -4.9]) and the upper limit of agreement 7.0 (95%CI = [7.2, 11.9]). As in the NC subgroup, approximative normality of differences based on residual analyses could not be confirmed, so the *T*-scores were transformed into rankits (i.e., standard normal deviates of the corresponding rank).127 The SEM based on the rankit-transformed T-scores was 0.46 and the SDC 1.3 (95%CI = [-1.3, 1.2]). The corresponding Bland and Altman plot is depicted in Figure 3b. The mean difference of the rankits of the two measurements was -0.05 (95%CI = [-0.3, 0.2], p = 0.17); the lower limit of agreement was -1.3 (95%CI = [-1.7, -0.96]) and the upper limit of agreement was 1.2 (95%CI = [0.9, 1.6]).

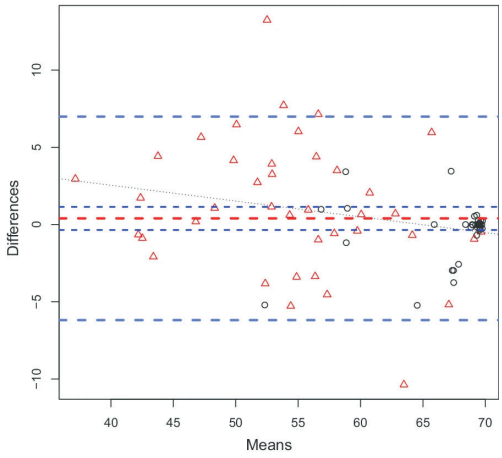


Figure 2 | Bland and Altman plot including all participants.
The X-axis shows the means of the *T*-scores of the two measurement time points and the Y-axis the difference in means of the *T*-scores between the two measurements. The horizontal red dashed line represents the mean difference, the dark blue dashed lines the 95% confidence interval of the mean difference, the blue dashed lines the lower and upper limits of agreement, and the black dotted line the regression line between the mean of the *T*-scores and difference in the means of the *T*-scores. Triangles represent participants with mild cognitive impairment or mild dementia and circles participants with normal cognition.

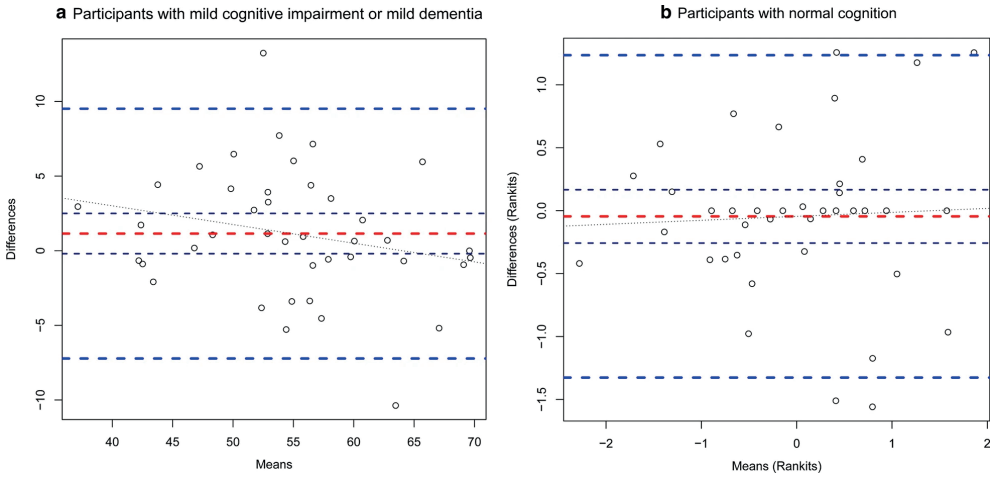


Figure 3 | Bland and Altman plot of the subgroups.
The horizontal red dashed line represents the mean difference, the dark blue dashed lines the 95% confidence interval of the mean difference, the blue dashed lines the lower and upper limits of agreement, and the black dotted line the regression line between the mean and difference in the means. In panel a, the X-axis shows the mean of the *T*-scores of the two measurement time points and the Y-axis the difference in means of the *T*-scores between the two measurements. The figure in panel b is based on the rankits of the *T*-score in participants with normal cognition. In panel b, the X-axis shows the means of rankits of the *T*-scores of the two measurement time points and the Y-axis the difference in means of rankits of the *T*-scores between the two measurements.

Construct validation of the A-IADL-Q-SV German

Point estimates of the observed correlations between the A-IADL-Q-SV German and education, the CDR, IQCODE, Lawton Brody scale, and MMSE were in the direction and of the magnitude hypothesized. Age was more strongly associated with the A-IADL-Q-SV German than hypothesized (-0.39, 95%CI [-0.60, -0.15]) and point estimates for depression were in the opposite direction. All hypothesized and observed correlations are summarized in Table 2.

Table 2 | Construct validation Spearman's correlation coefficients of *T*-scores of the A-IADL-Q-SV German with clinical measures

Measure	Hypothesized correlations		n	Observed correlations [95%CI]
	Direction	Range		
Age	–	0.0–0.2	56	-0.39 [-0.60, -0.15]
Education[a]	+	0.0–0.2	56	0.07 [-0.19, 0.33]
Everyday functioning				
CDR	–	0.2–0.4	56	-0.35 [-0.56, -0.09]
IQCODE	–	0.4–0.7	53	-0.69 [-0.81, -0.51]
lawton Brody scale[b]	–	0.4–0.7	53	-0.41 [-0.61, -0.16]
Cognitive function: MMSE	+	0.2–0.4	56	0.38 [-0.13, -0.58]
Depression: DIA-S	–	0.0–0.2	53	0.01 [-0.03, 0.27]

[a] Level of education in accordance with the international standard classification of education: ISCED (1 = ISCED 2, 2 = ISCED 3–5, 3 = ISCED 6, 4 = ISCED 7, 5 = ISCED 8).

[b] Dichotomized impaired = 1/non-impaired = 0.

Abbreviations: CDR, Clinical Dementia Rating; DIA-S, Depression in old age scale; IQCODE, Informant Questionnaire for Cognitive Decline in the Elderly; MMSE, Mini-Mental State Examination

Discriminant validity

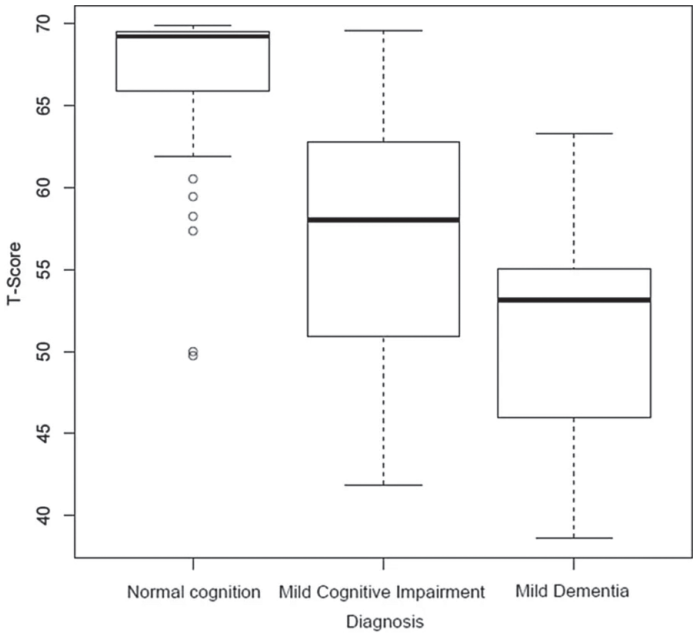


Figure 4 | *T*-scores of the three diagnostic groups.

Figure 4 shows the mean of the *T*-scores for participants with NC as 67 (range 50–70), those with MCI as 57 (range 42–70) and those with MD as 51 (range 39–63). Homogeneity of variances could not be assumed based on Levene's test for homogeneity of variances (F-value 6.54, df=2, p = .002) and Bartlett's test of homogeneity of variances (Bartlett's K-squared = 10, df=2, p = .008). Therefore, non-parametric analyses were performed. The results derived from the Kruskal–Wallis rank sum test indicated that the location parameters of the *T*-scores between the three diagnostic groups differed (Kruskal–Wallis' chi-square = 49, df=2, p < .001). Post-hoc pairwise comparisons using Wilcoxon rank sum tests revealed the following differences: NC versus MCI (p < .001); NC versus MD (p < .001); and MCI versus MD (p < .05).

4.2.4 Discussion

The results of the cross-cultural adaptation and validation indicated that the A-IADL-Q-SV German retained the measurement properties, i.e., there was no evidence of measurement variance by means of DIF, good construct validity, discriminant validity and test–retest reliability of the original version in a Swiss-German population of elderly people with NC, MCI, and MD. Therefore, the A-IADL-Q-SV German has been shown to be a psychometrically robust measurement instrument to assess IADL in elderly people within the range of no cognitive impairment to mild dementia. It is also comparable across countries.

In terms of measurement invariance by means of DIF, all basic assumptions for IRT scoring were met. This is in line with previous research on the A-IADL-Q-SV, indicating that the questionnaire measures one construct (i.e., IADL functioning).[20] The high inter-item correlations, which may have influenced the model fit indices, might be a reflection of the inclusion of less impaired participants in the Swiss-German sample compared to the Dutch reference sample. Our sample included participants from NC to MD, compared to the Dutch reference sample that included only memory clinic patients, who were generally more IADL impaired. In the Swiss-German sample, a high proportion of people rated most of the items with 'no problems', which may have inflated the inter-item correlations. A few item pairs (1%) showed larger correlation residuals than 0.25. This may indicate that the local independence of these items is compromised. The large residuals may also be caused by the fact that the sample was relatively homogeneous with regards to their level of overall functional impairment. This caused a limited variability in selected item responses. As most residuals are only marginally above the cut-off, and because other analyses show that the original IRT model fits and provides reliable estimates of everyday functioning, we are confident that the IRT model appears to fit in the Swiss sample.

The results of DIF analysis based on empirical data using pre-defined cut-offs indicated that the A-IADL-Q-SV German was robust to differences between the Swiss and Dutch cultures. Due to the small sample size in the Swiss sample, we additionally used Monte Carlo simulations to obtain the 99th percentile of the most extreme chi-squared and ΔR^2 values under the assumption that there is no DIF. The Monte Carlo simulations thus provide more precise cut-offs for the chi-squared test and ΔR^2 values. The items flagged for DIF using a priori thresholds matched the items flagged using the Monte Carlo thresholds, providing support for the a priori thresholds. The findings of the DIF analysis agree with a previous investigation on item bias in eight Western countries, which indicated that the A-IADL-Q-SV was robust to cultural differences, as well as to age, sex and education differences.[18]

In terms of test–retest reliability on scale level, our results indicated a good to excellent ICC based on a two-way mixed effects consistency model overall, as well as in the MCI/MD subgroup and NC subgroup. Previous research investigated test–retest reliability of the A-IADL-Q on the item level and revealed high test–retest reliability.[14] However, test–retest reliability of the A-IADL-Q-SV on scale level has

not been investigated previously. Nonetheless, the results of test–retest reliability on scale level of the A-IADL-Q-SV are relevant for clinical and research purposes. Both aim to use the questionnaire as an outcome measure, since the total score is interpreted.[41]

The SEM overall, as well as in the MCI/MD and NC subgroups, calculated by means of classical test theory, is implied to be acceptable with reference to the range of the *T*-scores. Measurement error was also investigated with Bland and Altman analyses. We observed that the data for the subgroup of participants with NC did not conform to the assumptions of the model. Therefore, we transformed the *T*-scores into rankits to rerun the analysis. The results of the Bland and Altman analyses overall and in the subgroups indicated that a change in the *T*-score of more than eight points might be interpreted as a real change.[23]

Construct validity in terms of hypothesis testing was shown, with more than 75% of the stated hypotheses being confirmed.[41] The hypotheses were specifically stated based on previous research on the A-IADL-Q[15] and A-IADL-Q-SV.[20,44] The correlations between the A-IADL-Q-SV and the clinical measures of cognition and functioning were in the magnitude and direction as hypothesized and are, therefore, in line with previous studies.[15,20] However, we observed a moderate correlation between the A-IADL-Q-SV German and age, whilst the original A-IADL-Q and A-IADL-Q-SV observed small correlations.[15,18,20] Another study of the A-IADL-Q-SV in Spain also observed a moderate correlation with age.[44] The findings in our study might be explained by the significantly higher age of participants with MCI and MD (and hence with significantly more IADL limitations) than participants with NC. A study investigating age as a source of item bias on the A-IADL-Q-SV found that the *T*-scores were not influenced by age.[18]

Furthermore, a positive correlation between the A-IADL-Q-SV German and the DIA-S was observed, which stands in contrast to our hypothesis and previous research.[15,20] This may be due to the different measurement instruments used to assess depression. The DIA-S was developed to assess depression in accordance with the diagnostic criteria of depression and, therefore, includes different items than those on the Geriatric Depression Scale (GDS). Only a moderate correlation was observed between the DIA-S and GDS.[34] However, since the observed correlation in our study between depression and IADL limitation was small, and in line with the literature,[15,20] it may be concluded that IADL limitation, as measured by the A-IADL-Q-SV, is not influenced by depression.

In terms of discriminant validity, our results indicate that the A-IADL-Q-SV German was able to discriminate between participants with NC, MCI, and ND. The interpretation of the *T*-scores observed in our study fitted well with the interpretation scheme. In fact, a previous study investigating the diagnostic value of the A-IADL-Q found a cut-off of 51.4 to differentiate between people with dementia and people without dementia,[16] corresponding almost perfectly to the mean *T*-score found in our MD group.

Limitations

We acknowledge some limitations to our study. A major limitation of this study may be the sample size. In terms of test–retest reliability based on ICCs and estimates of measurement error, the number of participants was relatively small in the two subgroups. This is reflected by the 95%CI of the ICCs of the subgroups (width > 0.2) and the change of the limits of agreement between the overall sample and the subgroup of participants with MCI/MD. With respect to the investigation of construct validity based on hypothesis testing, the small sample (n = 56) may have produced wide confidence intervals. A larger sample would have provided more precise estimates of the correlations. Furthermore, the overall sample size may have been too small to detect subtle measurement variance with DIF analysis. However, the ordinal logistic regression approach used in our study has previously been shown to be capable of detecting DIF when the reference sample is large, even when the focus sample is smaller.[45] Nonetheless, the generalizability of our results may be limited due to the restricted sample size.

Participants with NC were recruited from the community, while participants with MCI and MD were recruited from memory clinics associated with geriatric institutions, using a convenience sampling strategy. This may have produced bias that is reflected in the differences in demographics.

Cognitive status for participants with NC was investigated solely using TICS-m, a telephone screening for cognitive decline. Therefore, the possibility of participants with so-called subjective cognitive decline also being included in this group cannot be ruled out.

Information on participants' comorbidities was collected restrictively, meaning that the chance of comorbidities having influenced our results also cannot be excluded. However, due to its scoring structure, the A-IADL-Q-SV considers only

those limited activities related to cognitive problems. Furthermore, participants with comorbidities known to have an influence on cognitive function were excluded (i.e., clinical depression, drug and alcohol abuse, as well as neurological diseases, such as Parkinson's disease, stroke or traumatic brain injuries). Finally, data on the factors known to influence IADL functioning were collected, i.e., age, sex, level of education and living situation. As a result, we are convinced that the A-IADL-Q-SV T-scores correctly represent the level of IADL functioning, controlled for, e.g., physical impairments.

A subgroup of cognitively healthy participants was included in the test–retest analysis and in the investigation of measurement error. This inclusion of less impaired participants may have inflated the overall ICC, because the heterogeneity of the overall sample was increased. Consequently, the test–retest reliability and measurement error in the subgroups were also investigated separately. However, the inclusion of such participants was relevant for our study, because the decline in IADL functioning from a previously measured level often predates cognitive decline.[8]

Finally, our sample was not severely impaired and does not reflect the full dementia spectrum. Future investigations of the A-IADL-Q-SV German should use larger samples and include younger patients with MCI or a dementia diagnosis, as well as participants at the later stages of cognitive decline, i.e., moderate and severe dementia.

Conclusion

The cross-culturally validated A-IADL-Q-SV German has retained the psychometric properties (i.e., measurement invariance, test–retest reliability, construct validity and discriminant validity) of the original version. This study implies that the A-IADL-Q-SV German is a promising tool for use in clinical practice to investigate IADL functioning in elderly people with normal cognition, mild cognitive impairment, and mild dementia. It is also useful for research purposes and allows international comparisons to be made.

Declarations

Acknowledgments

The authors would like to thank all the participants and their informants for their participation in this study. The authors are also grateful to Dr. Stephanie Kaiser for

her help in data acquisition, and to Thomas Diener and Dr. André Straessle for their help in recruiting participants from the community.

Authors' contributions

M.B. participated in the study design and conception, data acquisition and analysis, wrote the first draft of the manuscript and revised new drafts. M.A.D. and A.M. participated in the data analysis and manuscript drafting. F.K. participated in the data acquisition and manuscript drafting. T.M. participated in the study design and conception and manuscript drafting. R.J.J. and P.S. participated in the data acquisition. S.A.M.S. participated in the study design and conception, data acquisition and analysis and manuscript drafting. K.N. participated in the study design and conception and revised new drafts of the manuscript. All authors read and approved the final manuscript.

Funding

This study was partially funded by the Foundation for Physiotherapy Science, % Inselspital, University Hospital Berne, Switzerland. The Foundation had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki and the principles of Good Clinical Practice. The respective ethical committee (EKOS, BASEC-NR. 2017-02200) approved the study protocol. All participants gave written informed consent to participate in the study.

Competing interests

The authors declare that they have no competing interests.

References

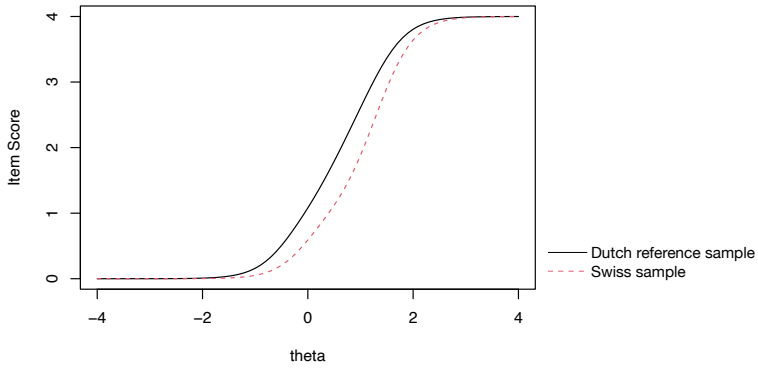
1. Farias ST, Lau K, Harvey D, Denny KG, Barba C, Mefford AN. Early Functional Limitations in Cognitively Normal Older Adults Predict Diagnostic Conversion to Mild Cognitive Impairment. J Am Geriatr Soc. Jun 2017;65(6):1152-1158. doi:10.1111/jgs.14835
2. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, Scheltens P, Uitdehaag BM. A systematic review of Instrumental Activities of Daily Living scales in dementia: room for improvement. J Neurol Neurosurg Psychiatry. Jan 2009;80(1):7-12. doi:10.1136/jnnp.2008.155838
3. Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild cognitive impairment: a concept in evolution. J Intern Med. Mar 2014;275(3):214-28. doi:10.1111/joim.12190
4. Jekel K, Damian M, Wattmo C, et al. Mild cognitive impairment and deficits in instrumental activities of daily living: a systematic review. Alzheimers Res Ther. 2015;7(1):17. doi:10.1186/s13195-015-0099-0
5. Ginsberg TB, Powell L, Emrani S, Wasserman V, Higgins S, Chopra A. Instrumental activities of daily living, neuropsychiatric symptoms, and neuropsycholgical impairment in mild cognitive impairment. J Am Osteopath Assoc. 2019;119(2):96-101.
6. Sachdev PS, Mohan A, Taylor L, Jeste DV. DSM-5 and Mental Disorders in Older Individuals: An Overview. Harv Rev Psychiatry. Sep-Oct 2015;23(5):320-8. doi:10.1097/HRP.0000000000000090
7. Gold DA. An examination of instrumental activities of daily living assessment in older adults and mild cognitive impairment. J Clin Exp Neuropsychol. 2012;34(1):11-34. doi :10.1080/13803395.2011.614598
8. 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. Oct 1 2012;2(5):483-491. doi:10.2217/nmt.12.55
9. Lautenschlager NT, Cox K, Cyarto EV. The influence of exercise on brain aging and dementia. Biochim Biophys Acta. Mar 2012;1822(3):474-81. doi:10.1016/j.bbadis.2011.07.010
10. Barrios PG, Gonzalez RP, Hanna SM, et al. Priority of Treatment Outcomes for Caregivers and Patients with Mild Cognitive Impairment: Preliminary Analyses. Neurol Ther. Dec 2016;5(2):183-192. doi:10.1007/s40120-016-0049-1
11. Lindbergh CA, Dishman RK, Miller LS. Functional Disability in Mild Cognitive Impairment: A Systematic Review and Meta-Analysis. Neuropsychol Rev. Jun 2016;26(2):129-59. doi:10.1007/s11065-016-9321-5
12. Sikkes SAM, De Rotrou J. A qualitative review of instrumental activities of daily living in dementia: what's cooking? Neurodegener Dis Manag. 2014;4(5):393-400. doi:10.2217/nmt.14.24
13. Kaur N, Belchior P, Gelinas I, Bier N. Critical appraisal of questionnaires to assess functional impairment in individuals with mild cognitive impairment. Int Psychogeriatr. Sep 2016;28(9):1425-39. doi:10.1017/S104161021600017X
14. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. Alzheimers Dement. Nov 2012;8(6):536-43. doi:10.1016/j.jalz.2011.08.006
15. Sikkes SAM, Knol DL, Pijnenburg YAL, 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(1):35-41. doi:10.1159/000346277

16. Sikkes SAM, Pijnenburg YAL, 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. Dec 2013;26(4):244-50. doi:10.1177/0891988713509139
17. Koster N, Knol DL, Uitdehaag BM, Scheltens P, Sikkes SAM. The sensitivity to change over time of the Amsterdam IADL Questionnaire((c)). Alzheimers Dement. Oct 2015;11(10):1231-40. doi:10.1016/j.jalz.2014.10.006
18. Dubbelman MA, Verrijp M, Facal D, et al. The influence of diversity on the measurement of functional impairment: An international validation of the Amsterdam IADL Questionnaire in eight countries. Alzheimers Dement (Amst). 2020;12(1):e12021. doi:10.1002/dad2.12021
19. Costa A, Bak T, Caffarra P, et al. The need for harmonisation and innovation of neuropsychological assessment in neurodegenerative dementias in Europe: consensus document of the Joint Program for Neurodegenerative Diseases Working Group. Alzheimers Res Ther. Apr 17 2017;9(1):27. doi:10.1186/s13195-017-0254-x
20. Jutten RJ, Peeters CFW, Leijdesdorff SMJ, et al. Detecting functional decline from normal aging to dementia: Development and validation of a short version of the Amsterdam IADL Questionnaire. Alzheimers Dement (Amst). 2017;8:26-35. doi:10.1016/j.dadm.2017.03.002
21. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976). Dec 15 2000;25(24):3186-91.
22. Wild D, Grove A, Martin M, Eremenco SL, McElroy S, Verjee-Lorenz A. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. Value Health. Mar-Apr 2005;8(2):94-104. doi:10.1111/j.1524-4733.2005.04054.x
23. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. Cambridge University Press; 2011.
24. Burge M, Bieri G, Bruhlmeier M, Colombo F, Demonet JF, Felbecker A. Recommendations of Swiss Memory Clinics for the diagnosis of dementia. Praxis. 2018;107(8):435-51.
25. Lacruz M, Emeny R, Bickel H, Linkohr B, Ladwig K. Feasibility, internal consistency and covariates of TICS-m (telephone interview for cognitive status-modified) in a population-based sample: findings from the KORA-Age study. Int J Geriatr Psychiatry. Sep 2013;28(9):971-8. doi:10.1002/gps.3916
26. Duff K, Shprecher D, Litvan I, Gerstenecker A, Mast B. Correcting for demographic variables on the modified telephone interview for cognitive status. Am J Geriatr Psychiatry. Dec 2014;22(12):1438-43. doi:10.1016/j.jagp.2013.08.007
27. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64(1):96-106.
28. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research. 1975;12(3):189-198. doi:10.1016/0022-3956(75)90026-6
29. Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. Int Psychogeriatr. 1997;9 Suppl 1:173-6.
30. Olde Rikkert MG, Tona KD, Janssen L, et al. Validity, reliability, and feasibility of clinical staging scales in dementia: a systematic review. Am J Alzheimers Dis Other Demen. Aug 2011;26(5):357-65. doi:10.1177/1533317511418954
31. Ehrensperger MM, Berres M, Taylor KI, Monsch AU. Screening properties of the German IQCODE with a two-year time frame in MCI and early Alzheimer's disease. Int Psychogeriatr. Feb 2010;22(1):91-100. doi:10.1017/S1041610209990962

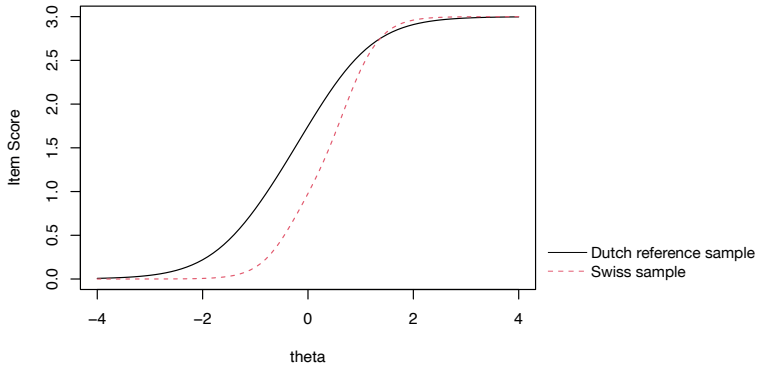
32. Jorm AF. The Informant Questionnaire on cognitive decline in the elderly (IQCODE): a review. Int Psychogeriatr. Sep 2004;16(3):275-93. doi:10.1017/S1041610204000390
33. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. Autumn 1969;9(3):179-86.
34. Heidenblut S, Zank S. Development of a new screening instrument for geriatric depression. The depression in old age scale (DIA-S). Z Gerontol Geriatr. Jun 2010;43(3):170-6. Entwicklung eines neuen Depressionsscreenings fur den Einsatz in der Geriatrie. Die "Depression-im-Alter-Skala" (DIA-S). doi:10.1007/s00391-009-0067-z
35. Heidenblut S, Zank S. Screening for depression with the Depression in Old Age Scale (DIA-S) and the Geriatric Depression Scale (GDS15): diagnostic accuracy in a geriatric inpatient setting. GeroPsych J Gerontopsychol Geriatr Psychiatry. 2014;27(1):41-9. doi:10.1024/1662-9647/a000101
36. R: A language and environment for statistical computing. Version 3.6.3. 2020.
37. Muthén LK, Muthén BO. Mplus User's Guide. 7th ed. Muthén & Muthén; 2012.
38. Raîche G, Walls TA, Magis D, Riopel M, Blais JG. Non-graphical solutions for Cattell's Scree test. Methodology. 2013;9(1):23-9.
39. van der Ark LA, Croon MA, Sijtsma K. Mokken Scale Analysis for Dichotomous Items Using Marginal Models. Psychometrika. Jun 2008;73(2):183-208. doi:10.1007/s11336-007-9034-z
40. van der Flier WM, Scheltens P. Amsterdam Dementia Cohort: Performing Research to Optimize Care. J Alzheimers Dis. 2018;62(3):1091-1111. doi:10.3233/JAD-170850
41. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. Jan 2007;60(1):34-42. doi:10.1016/j.jclinepi.2006.03.012
42. Jodoin MG, Gierl MJ. Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. Appl Meas Educ. 2001;14(4):329-349. doi:10.1207/S15324818AME1404_2
43. Choi SW, Gibbons LE, Crane PK. lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. Journal of Statistical Software. 2011;39(8)
44. Facal D, Carabias MAR, Pereiro AX, et al. Assessing everyday activities across the dementia spectrum with the Amsterdam IADL Questionnaire. Curr Alzheimer Res. 2018;15(13):1261-1266. doi:10.2174/1567205015666180925113411
45. Herrera AN, Gomez J. Influence of equal or unequal comparison group sample sizes on the detection of differential item functioning using the Mantel-Haenszel and logistic regression techniques. Qual Quant. Dec 2008;42(6):739-755. doi:10.1007/s11135-006-9065-z

4.2.5 Supplementary Material

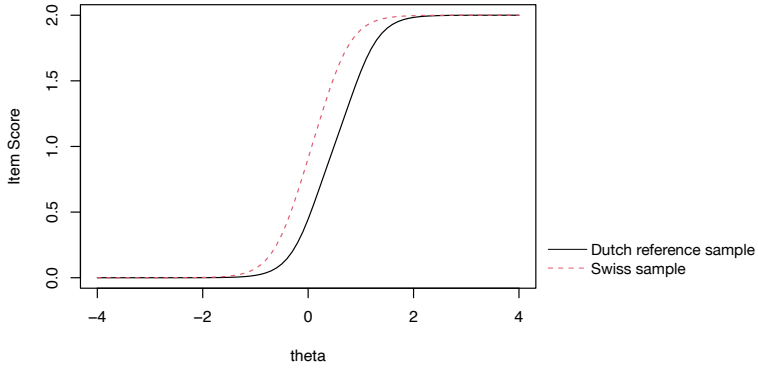
a) Item characteristic curve, item 2



b) Item characteristic curve, item 20



c) Item characteristic curve, item 23



Supplemental Figure 1 | Item characteristic curves.

Supplemental Table 1 | GRM item parameters and item information values.

Item no.	Item	Item parameters					Item informa-tion
		α	β1	β2	β3	β4	
Q1	Carrying out household duties	1.75±0.09	-0.31±0.06	0.83±0.14	1.82±1.04	2.81±8.15	4.61
Q2	Doing the shopping	2.35±0.12	-0.33±0.05	0.67±0.11	1.42±0.95	1.96±4.90	6.10
Q3	Buying the correct articles	2.06±0.11	-0.15±0.05	0.85±0.17	1.32±0.78	1.45±2.23	4.11
Q4	Cooking	2.30±0.12	-0.20±0.05	0.75±0.14	1.28±0.79	1.73±2.82	5.40
Q5	Preparing sandwich meals	1.94±0.13	1.05±0.06	1.84±1.64	2.49±8.40	2.89±30.47	4.14
Q6	Making minor repairs to the house	2.58±0.13	-0.48±0.05	0.32±0.05	0.76±0.26	1.01±0.73	5.62
Q7	Operating domestic appliances	2.29±0.13	0.32±0.05	1.27±0.55	1.99±3.98	2.58±22.59	5.85
Q8	Operating the microwave	2.01±0.12	0.69±0.05	1.36±0.67	1.88±2.57	2.08±7.32	3.79
Q9	Operating the coffee maker	1.90±0.13	1.05±0.06	2.01±1.99	2.49±9.67	2.71±26.54	3.78
Q10	Operating the washing machine	1.91±0.12	0.91±0.06	1.73±1.20	2.12±4.65	2.18±15.26	3.36
Q11	Paying bills	3.50±0.19	-0.01±0.04	0.60±0.18	0.90±0.83	1.13±2.26	7.76
Q12	Using a mobile phone	2.02±0.11	-0.04±0.05	0.90±0.19	1.64±1.17	2.37±6.58	5.03
Q13	Managing the paperwork	3.32±0.18	-0.54±0.04	0.34±0.04	0.72±0.34	0.96±1.05	8.06
Q14	Using electronic banking	2.96±0.16	-0.01±0.04	0.66±0.17	0.94±0.72	1.25±1.88	6.35
Q15	Using a PIN code	1.93±0.12	0.63±0.05	1.44±0.65	1.87±2.55	2.14±6.59	3.73
Q16	Obtaining money from an ATM	2.13±0.14	0.90±0.05	1.51±1.11	1.75±3.50	1.88±7.37	3.56
Q17	Paying using cash	1.98±0.13	0.96±0.06	1.74±1.36	2.17±5.51	2.56±15.43	3.97
Q18	Making appointments	2.01±0.11	-0.33±0.05	0.50±0.08	1.20±0.46	1.64±1.63	4.49
Q19	Filling in forms	2.63±0.13	-0.62±0.05	0.43±0.05	0.91±0.39	1.27±1.26	6.42
Q20	Working	1.52±0.09	-1.06±0.08	-0.05±0.06	0.47±0.12	0.74±0.29	2.83
Q21	Using a computer	2.68±0.14	-0.33±0.05	0.56±0.09	1.07±0.62	1.60±2.77	6.84

Item no.	Item	Item parameters					Item informa-tion
		α	β1	β2	β3	β4	
Q22	Emailing	2.68±0.15	0.07±0.04	0.79±0.22	1.11±0.91	1.30±2.23	5.43
Q23	Printing documents	2.48±0.15	0.26±0.04	0.86±0.28	1.15±0.95	1.31±2.12	4.53
Q24	Operating devices	2.71±0.15	-0.01±0.04	1.10±0.45	1.85±4.96	2.29±27.15	7.43
Q25	Operating the remote control	1.61±0.10	0.37±0.06	1.53±0.48	2.53±3.46	3.20±18.10	3.85
Q26	Playing card and board games	1.23±0.08	-0.29±0.07	1.12±0.18	1.89±0.81	2.30±2.21	2.46
Q27	Driving a car	1.53±0.09	-0.22±0.06	0.83±0.14	1.24±0.53	1.52±1.14	2.83
Q28	Using a sat-nav system	2.13±0.12	-0.13±0.05	0.75±0.15	1.09±0.62	1.40±1.39	4.28
Q29	Using public transportation	2.11±0.12	0.21±0.05	1.05±0.30	1.58±1.39	1.84±4.06	4.37
Q30	Being responsible for own medication	1.50±0.09	0.15±0.06	1.17±0.26	2.09±1.36	2.84±6.17	3.39

Supplemental Table 2 | Investigation of local independence.

Item 1	Item 2	Residual
Q8	Q13	-0.279
Q9	Q16	-0.258
Q10	Q16	-0.367
Q14	Q17	-0.252
Q15	Q23	-0.273

Supplemental Table 3 | Differential item functioning based on empirical data.

Item	χ^2 *p*-value	McFadden's ΔR^2
Q1	0.239	0.001
Q2	0.001	0.008
Q3	0.356	0.002
Q4	0.357	< 0.001
Q5	0.844	< 0.001
Q6	0.632	< 0.001
Q7	0.288	0.001
Q8	0.921	< 0.001
Q9	0.597	< 0.001
Q10	0.535	0.002
Q11	0.942	< 0.001
Q12	0.926	< 0.001
Q13	0.272	0.002
Q14	0.072	0.005
Q15	0.404	< 0.001
Q16	0.097	0.007
Q17	0.802	< 0.001
Q18	0.422	< 0.001
Q19	0.133	< 0.001
Q20	< 0.001	0.019
Q21	0.265	0.001
Q22	0.067	0.003
Q23	0.002	0.015
Q24	0.105	0.003
Q25	0.183	0.002
Q26	0.137	0.002
Q27	0.905	< 0.001
Q28	0.364	< 0.001
Q29	0.888	< 0.001
Q30	0.320	< 0.001

χ^2 and McFadden's ΔR^2 values as obtained in differential item functioning (DIF) analyses from the empirical data. Items flagged for DIF are displayed bold in blue. Empirical cutoffs were set a priori at $\alpha < .01$ for statistically significant DIF, and $\Delta R^2 > .035$ for clinically meaningful DIF.

Supplemental Table 4 | Differential item functioning based on Monte Carlo simulations.

Item	χ^2 *p*-value	McFadden's ΔR^2
Q1	0.001	0.006
Q2	0.001	0.007
Q3	0.001	0.007
Q4	0.004	0.007
Q5	0.016	0.009
Q6	0.009	0.004
Q7	0.006	0.007
Q8	0.001	0.009
Q9	0.034	0.006
Q10	0.001	0.010
Q11	0.001	0.006
Q12	0.002	0.006
Q13	0.010	0.004
Q14	0.023	0.005
Q15	0.016	0.005
Q16	0.001	0.016
Q17	0.001	0.010
Q18	0.015	0.004
Q19	0.033	0.003
Q20	0.004	0.005
Q21	0.016	0.004
Q22	0.002	0.006
Q23	0.007	0.006
Q24	0.017	0.004
Q25	0.006	0.006
Q26	0.028	0.004
Q27	0.008	0.004
Q28	0.003	0.006
Q29	0.014	0.005
Q30	0.002	0.006

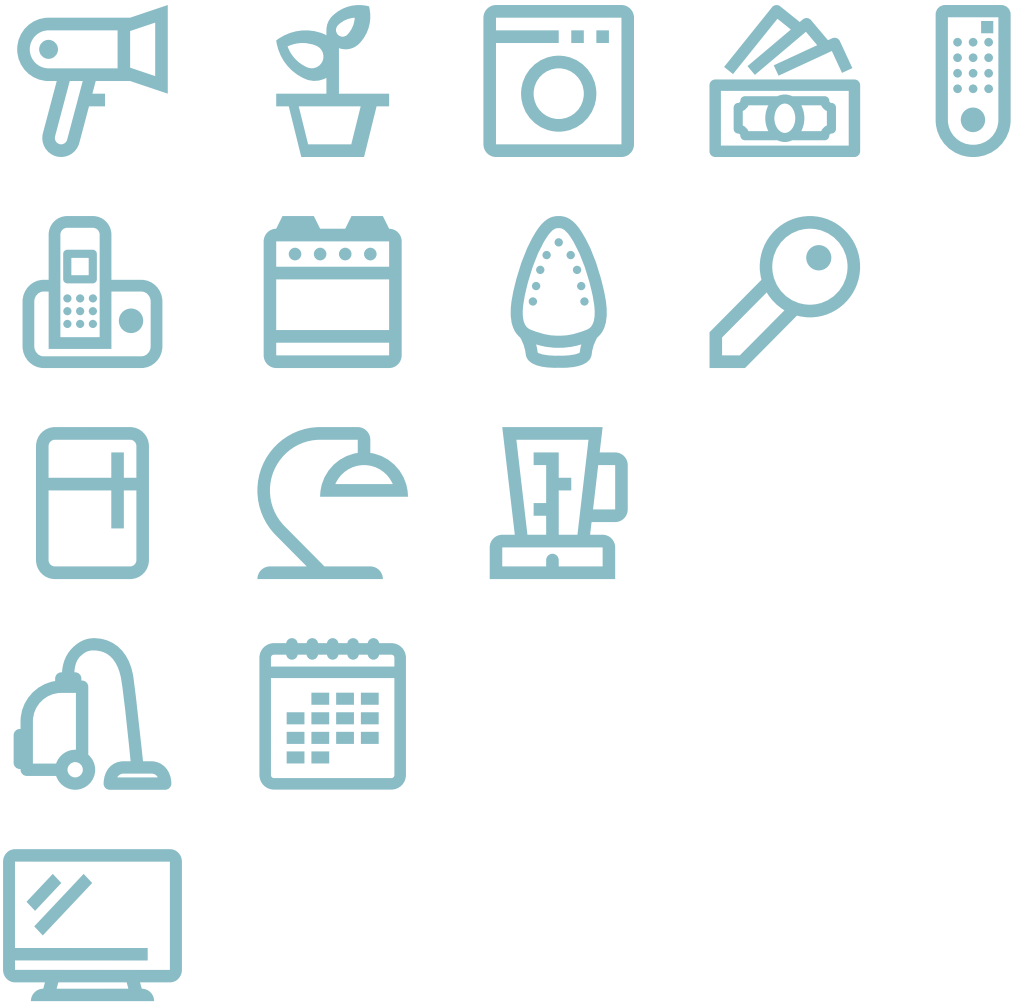
X^2 and McFadden's ΔR^2 values as obtained in differential item functioning (DIF) analyses from the empirical data. Items flagged for DIF are displayed bold in blue. Empirical cutoffs were set a priori at $\alpha < .01$ for statistically significant DIF, and $\Delta R^2 > .035$ for clinically meaningful DIF.

PART III

MEASURING CHANGES IN EVERYDAY FUNCTIONING

"A diferença entre o 'ser', estado permanente,
e o 'estar', estado passageiro.
Eu sou. Sem pedir licença"

— Aurora Negra (2020



CHAPTER 5

COMPARING AND ALIGNING SCORES BASED ON ITEM RESPONSE THEORY AND CLASSICAL TEST THEORY FOR THE AMSTERDAM INSTRUMENTAL ACTIVITIES OF DAILY LIVING QUESTIONNAIRE: A LONGITUDINAL STUDY OF COGNITIVELY NORMAL AND IMPAIRED OLDER ADULTS

Mark A. Dubbelman, Merel C. Postema, Roos J. Jutten, John E. Harrison, Craig W. Ritchie, André Aleman, Frank-Jan de Jong, Wiesje M. van der Flier, Philip Scheltens, Sietske A.M. Sikkes

Under review

Abstract

Purpose: We aimed to compare and align item response theory (IRT) with classical test response theory (CTT) scores of impairment in everyday functioning, as measured with the Amsterdam Instrumental Activities of Daily Living Questionnaire (A-IADL-Q).

Methods: 2,295 participants were included, of whom *n* = 2,032 (89%) were cognitively normal, *n*=93 (4%) had subjective cognitive decline, *n* = 79 (3%) had mild cognitive impairment and *n* = 91 (4%) had dementia. We compared IRT-based and CTT-based score distributions and discriminative ability between diagnostic groups using linear regressions and investigated floor and ceiling effects. We compared change over time between scoring methods using linear mixed models with random intercepts and slopes for time in a subsample of *n* = 1,145 (62%) who were followed for 1.3 ± 0.6 years. To align IRT-based and CTT-based scores on a single scale, we simulated 150,000 A-IADL-Q responses.

Results: At baseline, IRT-based and CTT-based scores were highly correlated (*r* = -0.92) and differed between diagnostic groups (all *p* < .001). Concerning changes over time, IRT-based scores declined significantly among cognitively normal individuals (unstandardized coefficient (B) = -0.15, 95% confidence interval (95%CI) = [-0.28, -0.03]), whereas CTT-based scores did not (B = 0.20, 95%CI = [-0.02, 0.41]). In the other groups, both scores showed similar change over time. CTT-based scores were successfully aligned to the IRT scale.

Conclusion: Both scoring methods showed similar results, but IRT-based scores seemed slightly more sensitive than CTT-based scores in assessing early functional changes. Alignment of the scoring methods on the same scale allows comparison of results, eliminates the need for separate cutoffs and interpretation guidelines, and facilitates use in research and clinical practice.

5.1 Introduction

Alzheimer's disease (AD) is characterized by a progressive loss of cognitive and functional abilities.[1] Early in the disease, individuals may develop difficulties with cognitively complex activities, or so-called 'instrumental activities of daily living' (IADLs),[2-4] such as managing paperwork and making appointments. The level of IADL functioning represents a clinically relevant outcome, even in early stages of AD. There are various instruments for measuring changes in IADL functioning in the context of AD, including patient- or (observer-)reported outcome measures (PROMs).[5] PROMs assessing IADL functioning are frequently used in longitudinal studies, including in clinical trials.

Classical test theory (CTT) holds that an observed total score is the sum of a person's true score (or ability) and random error of measurement, which is assumed to be equal across all individuals, regardless of their ability. A limitation of this assumption is that individuals with extreme levels of an ability are likely to be measured less precisely because most items of a questionnaire generally cover average ability. This is a relevant limitation in the assessment of IADL functioning, as many individuals have extreme ability (no impairment in IADL), as indicated by the presence of ceiling effects in many IADL instruments.[6]

Item response theory (IRT), also known as 'modern test theory' because it builds on and follows CTT, accounts for this and assumes that measurement error varies across the scale. IRT employs mathematical models to describe the relationship between a person's true ability on a construct that is not directly observable and the probability of the person giving a certain response to an individual item measuring that ability.[7,8] Unidimensional IRT models, such as the graded response model, assume that a single latent trait underlies all items of an instrument, i.e., that the construct represents one dimension, but it does not assume a linear relationship between a person's ability and the item responses. Rather, it accounts for varying properties of items, as the discrimination and location parameters.

While both methods are complimentary in use,[9] IRT holds several advantages over CTT. Importantly, it has been suggested that IRT-based scores may be less biased than CTT-based scores when estimating change in a construct,[10-12] even potentially increasing responsiveness.[13] Furthermore, because item parameters provide information on each item's difficulty and discriminatory ability, IRT allows for computerized adaptive testing (CAT) which may substantially reduce the burden of completing questionnaires.

One example of an IADL PROM based on IRT, is the Amsterdam IADL Questionnaire (A-IADL-Q), which was developed in 2012[14] to measure difficulties performing complex everyday activities due to cognitive decline. The A-IADL-Q has since been extensively validated, showing good psychometric properties,[15-19] and particularly sensitivity to changes over time.[20] The A-IADL-Q can be scored using both IRT and CTT. IRT scoring is most often used for research purposes because it uses activity-level information and therefore allows for more precise measurement of functional ability. CTT scoring is typically used in clinical settings for practical reasons, such as it being more straightforward to calculate the scores.

We aimed to compare the IRT and CTT scoring methods of the A-IADL-Q in a predominantly cognitively normal sample. First, we aimed to compare score distributions and floor and ceiling effects between diagnostic groups. Second, we intended to analyze change over time in both scores. We hypothesized that the IRT-based scores (a) will discriminate well between people who are cognitively normal and those with subjective cognitive decline, and (b) will show a similarly strong signal when assessing changes over time as CTT-based scores. The third and final aim was to align the two scoring methods, with the ultimate goal of moving towards a single IRT metric, allowing for uniform scoring and a single set of interpretation guidelines.

5.2 Methods

Participants

We included participants from various study samples: the European Prevention for Alzheimer's Disease consortium Longitudinal Cohort Study (EPAD-LCS),[21] the Capturing Changes in Cognition (Catch-Cog) study,[22] and the SCIENCe cohort.[23] EPAD-LCS included participants from numerous study sites across Europe and was designed to reflect a trial-ready population. Participants were 50 years and older.[21] Participants in Catch-Cog were recruited at multiple sites in The Netherlands and Scotland, were 65 years or older, and had diagnoses of subjective cognitive decline (SCD), mild cognitive impairment (MCI) or mild dementia at study inclusion.[22] Patients who visited the outpatient memory clinic of the Alzheimer Center Amsterdam and received a diagnosis of SCD were included in the SCIENCe cohort.[23] In addition to the inclusion criteria of these studies, we selected individuals who had at least one completed A-IADL-Q, with item-level data available. When multiple assessments were available, we included all assessments in the three years from baseline.

Measures: the Amsterdam Instrumental Activities of Daily Living Questionnaire (A-IADL-Q)

The Amsterdam IADL Questionnaire (A-IADL-Q) is a proxy-reported outcome measure assessing difficulties in the performance of various cognitively complex activities.[14] It has been extensively validated.[15,17,18,20] All 30 items included in the short form[16] are scored on a Likert-type scale ranging from 0 (the person has no difficulty performing an activity) to 4 (the person is unable to perform an activity). Items are not scored—and thus considered missing—when: (a) the participant never performed the activity; (b) the inability to perform the activity was due to some reason other than cognitive impairment; or (c) the study partner did not know whether the participant performed the activity in the four weeks preceding the assessment. An example item with the scoring logic is displayed in Figure 1.

Did he/she manage his/her paperwork?
This question relates to the past 4 weeks.

- ○ Yes
  - *If yes,*
  - Did he/she find it more difficult to manage his/her paperwork than he/she had in the past?
    - ○ No (0)
    - ○ Yes, slightly more difficult (1)
    - ○ Yes, more difficult (2)
    - ○ Yes, much more difficult (3)
    - ○ Yes, he/she is no longer able to perform this task (4)
- ○ No
  - *If no,*
  - He/she did not manage his/her paperwork for the following reason;
    - ○ He/she was no longer able to do so due to difficulties with his/her memory, planning, or thinking (4)
    - ○ He/she was no longer able to do so due to his/her physical problems (-)
    - ○ He/she has never done that before (-)
    - ○ Other, please state (-)
- ○ Don't know
  - *Continue to next question* (-)

Figure 1 | Example item of the A-IADL-Q, with the scoring logic displayed on the right.
(-) denotes that the response is scored as missing.

Using the graded response model,[24] an IRT-based 'theta' was defined with a mean of 0 and a standard deviation of 1 in the memory clinic. We subsequently linearly transformed the theta to a *T*-score, multiplying it by 10 and adding 50 (*T*-score 50 ± 10). Higher scores represent higher levels of everyday functioning. The IRT item parameters have been published previously by Jutten et al.[16]

Because missing item responses are introduced to the questionnaire by design, the CTT-based total score is not a sum score, but rather the average of all completed items (i.e., the sum divided by the number of non-missing items). The average is multiplied by 25, so scores range from 0 to 100. A score of 0 represents no impairment and a score of 100 represents severe impairment in everyday functioning.

Statistical analyses

Group differences in IRT-based and CTT-based A-IADL-Q total scores between the diagnostic groups were tested using analysis of variance and chi-squared tests, as appropriate, and *p*-values were adjusted using Tukey's Honest Significant Difference test. We also counted how many individuals had the lowest and highest scores (floor and ceiling effects) with both scoring methods.

Next, we wanted to compare the ability of IRT-based and CTT-based A-IADL-Q scores to capture change over time. In the subsample of participants with longitudinal A-IADL-Q data, we ran two linear mixed models, one with IRT-based and one with CTT-based A-IADL-Q scores as the dependent variable. Time was the main independent variable and we included random intercepts and random slopes for time, as well as a diagnostic group by time interaction. The models were adjusted for age, sex, and education. Effect sizes were calculated by dividing the unstandardized time coefficient by the sum of the square roots of the variances of all intercepts and slopes, as well as the residual variance.[25] Furthermore, we computed an unadjusted mean-to-standard deviation ratio (MSDR) of change. The MSDR is calculated by dividing the mean change from baseline to the last visit by the standard deviation of that change. Larger MSDRs represent larger effects.

All analyses were performed in R version 4.1.2,[26] using the 'lme4' package version 1.1-27 for the linear mixed models.[27]

Aligning scoring methods

The final step was to align CTT-based scores to the IRT-based *T*-scale. Because of designed missingness and use of an average CTT-based score, a direct scale alignment approach, such as the Lord-Wingersky algorithm,[28,29] was not possible. Thus, we aligned the scales by aligning response patterns for both IRT and CTT scoring based on a large dataset of randomly generated responses to the A-IADL-Q. Each unique response pattern results in a unique IRT-based score, whereas multiple

response patterns may result in the same CTT-based score. Hence, the number of unique IRT-based scores (6^{30} unique response patterns; each of the 30 items can have a score between 0 and 4, or can be missing) is far greater than the number of unique CTT-based scores (*n* = 377).

We randomly generated a sufficiently large number of responses (*n* = 150,000) to cover the entire spectrum of the scale, 100,000 of which were unweighted (i.e., all response categories were equally likely to be endorsed; 20% of responses in each category). To increase the number of responses at the extremes of the scale, and improve our precision, we also generated 25,000 responses where the probability of endorsing the lowest response option was higher (80% of responses in the lowest category, 5% in each of the other categories) and 25,000 responses where the probability of endorsing the highest response option was higher (80% of responses in the highest category, 5% in each of the other categories). Missingness was imposed according to missingness observed in the real-life A-IADL-Q data. That is, we obtained the percentage of missing answers for each item from the real-life baseline data in the present sample and applied these percentages to the simulated item options.

Because the goal is to establish an IRT-based *T*-score that corresponds best to the CTT-based score, we recorded the minimum, maximum, mean and standard deviation of all IRT-based scores that corresponded to each unique CTT-based score. The mean IRT-based score for each unique CTT-based score was taken as the aligned score and these were used to create an alignment table. By means of a validation of the alignment, we calculated Person's correlation coefficient between the actual IRT-based score and the aligned score in our dataset. Finally, we recorded the variance in both scores, taking the expected value of the squared deviation of the mean of both scores.

5.3 Results

We included 2,295 participants (66.6 ± 7.7 years old, 54% female, median education 15 years), most of whom (*n* = 2,031; 89%) were cognitively normal at inclusion. Of the remaining participants, 93 (35%) were diagnosed with subjective cognitive decline, 79 (30%) with mild cognitive impairment, and 91 (35%) with mild dementia. All diagnostic groups differed from each other in terms of age, with participants with mild cognitive impairment being the oldest, followed by participants with dementia, normal cognition, and subjective cognitive decline (all adjusted *p* <

.01). Sex distributions also differed between the groups ($p < .001$): participants with normal cognition or dementia were more often female than others. Education differed significantly between participants with normal cognition and participants with subjective cognitive decline or dementia (both adjusted $p < .001$), but not between the other groups. Table 1 shows the baseline characteristics of the sample for the entire group, as well as each diagnostic group separately.

Table 1 | Baseline characteristics.

	All	Normal cognition	Subjective cognitive decline	Mild cognitive impairment	Mild dementia
N	2,294	2,031 (88.5)	93 (4.1)	79 (3.4)	91 (4.0)
Age in years	66.6 ± 7.7	66.2 ± 7.5	63.5 ± 7.5	73.7 ± 7.9	70.9 ± 8.8
Female sex, n (%)	1,244 (54.2)	1,138 (56.0)	36 (37.9)	29 (36.7)	41 (45.1)
Education years, M (IQR)	15 (12–17)	15 (12–17)	13 (10–17)	14 (12–16)	13 (10–16)
A-IADL-Q scores					
IRT, mean ± SD	65.7 ± 6.9	67.3 ± 4.1	59.8 ± 8.8	54.3 ± 7.4	44.9 ± 8.6
IRT, range	29.8–70.0	42.1–70.0	35.3–69.8	31.6–69.7	29.8–68.6
IRT, n (%) at ceiling[a]	54 (2.4)	54 (2.7)	0 (0.0)	0 (0.0)	0 (0.0)
CTT, mean ± SD	3.9 ± 11.8	1.3 ± 4.8	10.3 ± 15.7	17.2 ± 16.3	43.2 ± 25.5
CTT, range	0–87	0–53.3	0–71.5	0–78.3	0–87
CTT, n (%) at floor	1,622 (70.7)	1,589 (78.2)	27 (29.0)	5 (6.3)	1 (1.1)

[a] The ceiling for the IRT-based scores was determined as 70.0: the highest scores achieved rounded to one decimal.

Abbreviations: A-IADL-Q, Amsterdam Instrumental Activities of Daily Living Questionnaire; CTT, classical test theory; IRT, item response theory; IQR, interquartile range; M, median; SD, standard deviation.

Comparing scoring methods

IRT-based and CTT-based scores correlated strongly (Pearson's r = -0.92, 95% confidence interval (95%CI) = [-0.93, -0.91]). In cross-sectional comparisons, both IRT and CTT-based A-IADL-Q scores were different between all groups (all adjusted $p < .001$). Figure 2 shows the baseline distributions of both scores for the different diagnostic groups. A total of 1,622 individuals (70.7%) had a CTT-based score of 0, indicating no impairment. Only 54 individuals (2.4%) had an IRT-based score of 70.0, which was the highest score reached in our sample, indicating no impairment. Further, while there were individuals in all diagnostic groups with a CTT-based score at the floor, only cognitively normal individuals reached the ceiling of IRT-based scores.

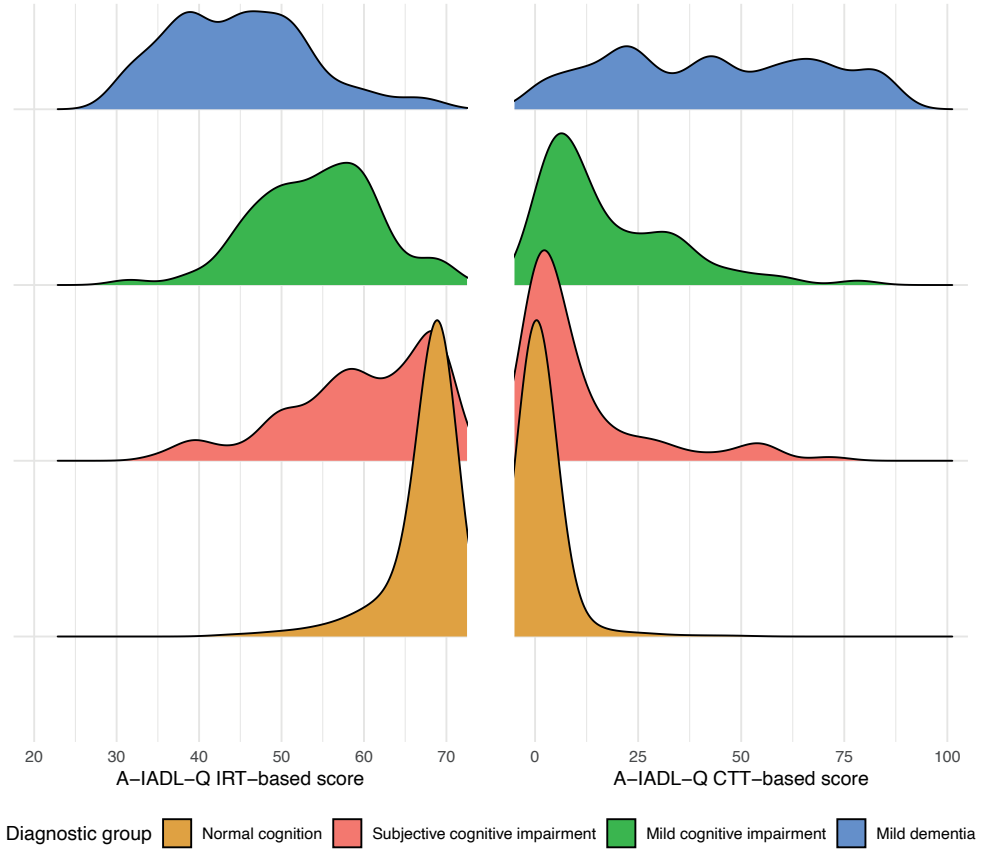


Figure 2 | Baseline distributions of A-IADL-Q IRT-based (left) and CTT-based (right) scores, stratified by diagnostic group.
Abbreviations: A-IADL-Q, Amsterdam Instrumental Activities of Daily Living Questionnaire; CTT, classical test theory; IRT, item response theory.

Change over time

A total of 1,415 individuals (61.7%) had longitudinal data available, with a median of two visits (interquartile range 2–3) per person and a mean of 1.35 ± 0.63 years of follow-up. Linear mixed models showed that both scoring techniques showed change over time in everyday functioning in the whole sample. Table 3 shows the unstandardized coefficients along with effect sizes for IRT and CTT-based scores in the total sample and in the different diagnostic groups. IRT-based scores deteriorated modestly but significantly over time in cognitively normal older adults (B = -0.15, 95%CI = [-0.28, -0.03]), while although CTT-based scores changed

in the expected direction, this change did not reach significance (B = 0.20, 95%CI = [-0.02, 0.41]). Both IRT and CTT-based scores improved in participants with SCD (B = 1.33, 95%CI = [0.78, 1.89] and B = -1.83, 95%CI = [-2.82, -0.83], respectively). In MCI and dementia, both IRT and CTT-based scores deteriorated significantly (see Table 3). Effect sizes were similar between the scoring methods in the whole group but were larger for CTT-based scores in more advanced disease stages. Figure 3 shows the change in the different scoring techniques over time, per diagnosis.

Table 3 | Estimated yearly change (slopes) in IRT and CTT-based scores in the total sample and in different groups.

	IRT		CTT	
	Coefficient	**Effect size**	**Coefficient**	**Effect size**
Total group	-0.26 [-0.38, -0.13]	-0.24	0.68 [0.43, 0.93]	0.22
NC	-0.15 [-0.28, -0.03]	-0.02	0.20 [-0.02, 0.41]	0.02
SCD	1.33 [0.78, 1.89]	0.21	-1.83 [-2.82, -0.83]	-0.15
MCI	-1.44 [-2.29, -0.58]	-0.23	4.62 [3.09, 6.16]	0.37
Dementia	-3.46 [-4.21, -2.71]	-0.55	9.33 [7.97, 10.69]	0.74

Time coefficients are shown with the 95% confidence interval, and were adjusted for baseline age, sex and education. Note: IRT and CTT scales are mirrored to one another.
Abbreviations: CTT, classical test theory; IRT, item response theory; NC, normal cognition; MCI, mild cognitive impairment; SCD, subjective cognitive decline.

In the total group, the mean-to-standard deviation ratio (MSDR) was marginally larger for the CTT score (MSDR = 0.12) than for the IRT score (MSDR = -0.08). A similar pattern was observed in cognitively normal participants (CTT MSDR = 0.10, IRT MSDR = -0.08) and participants with mild cognitive impairment (CTT MSDR = 0.39, IRT MSDR = -0.31). Among participants with subjective cognitive decline, the CTT MSDR was somewhat smaller than the IRT MSDR (-0.23 vs. 0.28), and this was also the case in patients with mild dementia (CTT MSDR = 0.59, IRT MSDR = -0.62).

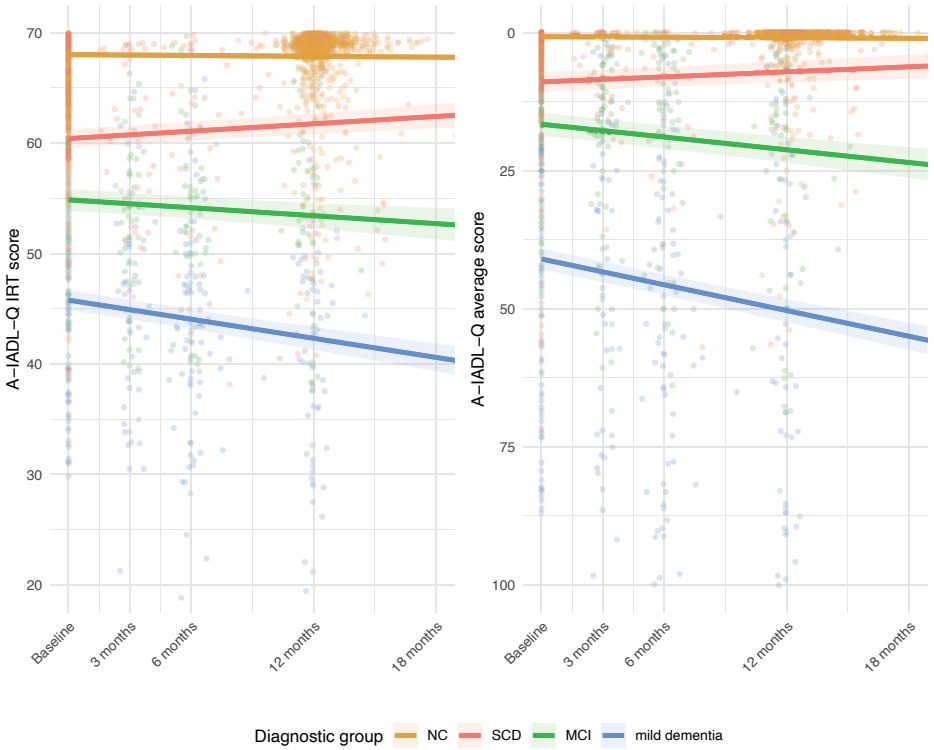


Figure 3 | Trajectories of A-IADL-Q scores over time, by diagnostic group.
Left panel: IRT-based scores, right panel: CTT-based scores. The Y-axis in the right panel is inversed so both scoring methods are in the same direction (i.e., a downward pointing line indicates a decline in everyday functioning).

Aligning scoring methods

The simulations resulted in 150,000 unique IRT-based scores and 372 unique CTT-based scores. The mean and range of IRT-based scores were computed for each unique CTT-based score, and this mean was aligned with the CTT-based score. An alignment table containing the IRT-based scores corresponding to each CTT-based score can be found in the Supplementary Material.

In our real-life dataset, IRT-based scores that were calculated directly on the actual data were highly correlated with the aligned IRT-based scores (Pearson's r = 0.986, 95%CI = [0.984, 0.987]). Figure 4 shows that the aligned IRT-based scores (dark colors) had a reduced variation (variance = 41.5) as compared to the actual IRT-based scores (light colors; variance = 48.1), but that, overall, they had the same distributions.

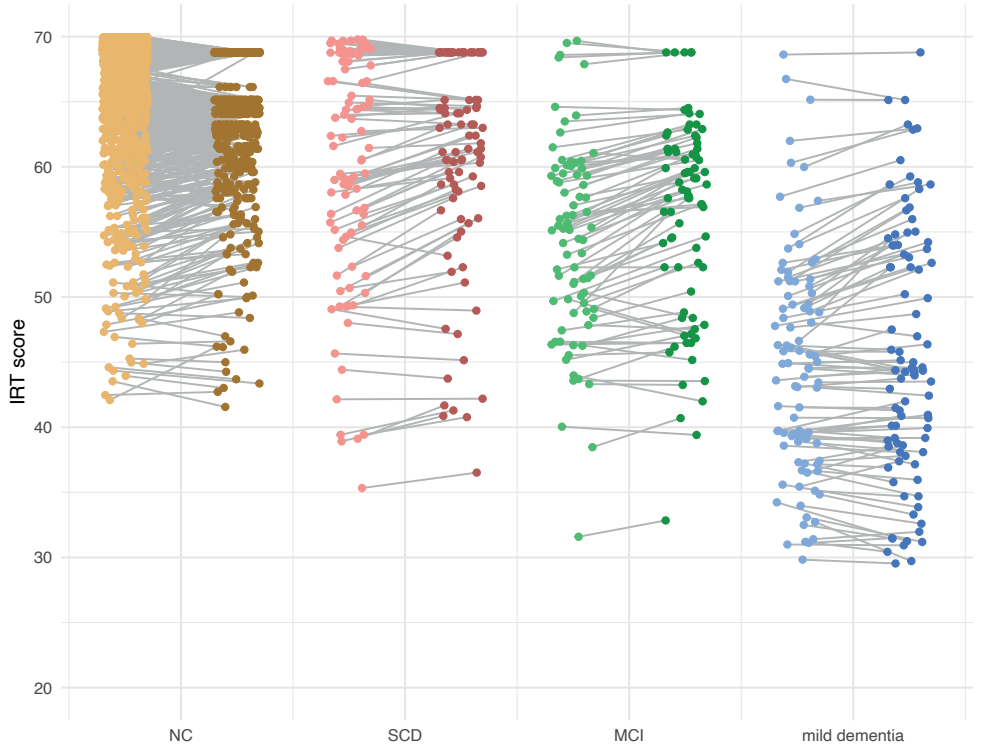


Figure 4 | Actual IRT-based scores (light colors) and aligned IRT-based scores (dark colors), stratified by diagnostic group.
Abbreviations: IRT, item response theory; MCI, mild cognitive impairment; NC, normal cognition; SCD, subjective cognitive decline.

5.4 Discussion

In this study, we set out to compare and align IRT and CTT-based total scores for the A-IADL-Q and found that they were highly correlated with each other. IRT-based scores had much less of a ceiling effect than CTT-based scores, particularly in cognitively normal participants. In longitudinal analyses, effect sizes of change over time in both scores were comparable. These findings allowed us to successfully align IRT-based and CTT-based scores to place them on the same scale.

Previous studies have provided extensive validation of the A-IADL-Q, including good content and construct validity, diagnostic accuracy, and responsiveness.[17,18,20,30] The questionnaire is used internationally and has been culturally adapted; no differential item functioning has been found between countries.[15,31,32] All these studies have consistently made use of the IRT-based scores. This raises the question

of whether CTT-based scores, which are used in clinical practice, are comparable to the IRT-based scores. Indeed, an important advantage of CTT-based scores over IRT-based scores, is that they can be calculated more easily, especially when the questionnaire is administered on paper. We showed that IRT and CTT-based scores correlated almost perfectly, which is a well-established finding for other outcome measures.[9] A caveat to consider was the existence of a floor effect in the CTT-based scores, with more than two thirds of all participants scoring the lowest possible score, indicating no impairment. This effect was much smaller for IRT-based scores, where only two-and-a-half percent of participants scored at the extreme end of the scale indicating no impairment. These findings suggest that CTT-based scores can justifiably be aligned with IRT-based scores. IRT-based scores are favored, especially in populations where the extremes of the scale are more frequently endorsed (i.e., where functional impairment is limited).

Building on that finding, we hypothesized that IRT-based scores would be better able to detect changes in A-IADL-Q scores at the extremes of the scale, too. There is a body of evidence suggesting that IRT-based scores outperform CTT-based scores in longitudinal analyses in terms of consistency of findings,[10-12] especially in small samples and in the face of missing data.[33] While marginal, IRT-based scores showed a significant decline over time among cognitively normal individuals, whereas the change in CTT-based scores was not significant. Interestingly, in individuals with subjective cognitive decline, we observed changes in the opposite direction, where functioning seemed to improve over time. This improvement was observed in both IRT and CTT-based scores. Improvements in functioning are not commonly reported, but it is possible that reassurance after the initial memory clinic visit may have alleviated some concerns in this group of patients. At more advanced disease stages, IRT and CTT-based scores were comparable. These findings may imply that IRT and CTT-based scores are more interchangeable in clinical practice and in clinical trials focused on impaired individuals, whereas IRT-based scores seem to be marginally more precise in prevention trials and research on early stages of Alzheimer's disease and related disorders.

A challenge that lies in using two different scales, is that any developments for the improvement of clinical usability need to be made twice. For example, when researching cutoffs to distinguish normal from impaired everyday functioning, two different cutoffs would need to be determined so they can be applied to both scales. Furthermore, misinterpretation of findings is possible because the two scales

are each other's inverse. Because IRT and CTT-based scores have such a strong correlation, we can align the two on a single scale. Clinicians and researchers using the CTT-based scores should in the future use the IRT-based score. We provided a cross-walk table that can be used to convert CTT-based scores to IRT-based *T*-scores.

Our study highlights methodological complexity when computing CTT to IRT cross-walk tables. There are various methods for linking CTT to IRT-based scores,[34] one of which is the Lord-Wingersky algorithm,[28] used in the PROsetta Stone project.[35] The Lord-Wingersky algorithm relies on the IRT scoring parameters and was designed to link different instruments onto a single scale, but may also be used to align different scoring techniques of the same instrument on the same scale. The latter works only with sum scores, but because of the design of the A-IADL-Q, in which responses that are not relevant for measuring the underlying construct of everyday functioning are considered missing, CTT-based scores are an average score. Hence, we could not use this algorithm to align the CTT and IRT-based scores. Therefore, we opted to use simulations instead, with the drawback that we could not simulate all possible item response combinations due to computational constraints with the enormous number of possible response patterns. Still, we found that the IRT-based scores derived from alignment, that is, CTT-based scores that were converted onto the IRT scale, correlated very highly with IRT-based scores as computed using the IRT parameters. This provides a validation of our alignment method and suggests that the CTT-based scoring may be applied in contexts where the more elaborate IRT scoring is not readily available or practicable, and that these CTT-based scores can subsequently be placed on the IRT scale. An important advantage of aligning the two scales is that previously published IRT score cutoffs can be applied to CTT-based scores as well. Hence, we recommend users of the CTT-based scores to use the cross-walk tables to convert their scores into IRT-based *T*-scores.

A strength of this study was the inclusion of a large sample of patients from various European sites who were followed over time, spanning the disease spectrum from cognitively normal to mild dementia. A limitation was the unequal distribution across diagnostic groups, with specifically the mild cognitive impairment group being underrepresented. This may have limited our power to detect changes in the more advanced disease stages. Another limitation was the fact that we were not able to use the more elegant Lord-Wingersky algorithm for aligning IRT and CTT-based scores and had to use a non-exhaustive simulation approach. However, with

the large number of simulations we achieved a good alternative and had many more unique response patterns than in our real-life dataset.

In conclusion, IRT-based scores for the A-IADL-Q have advantages including the lack of a ceiling effect, possibility of computerized adaptive testing, and slightly superior responsiveness in early disease stages. With the alignment of the IRT and CTT-based scores, made possible by the similarities in measurement properties of the two scoring methods, scores can now be placed on the same scale, regardless of calculation method. This eliminates the need for separate cutoffs and interpretation guidelines, thus facilitating use in clinical practice.

References

1. Scheltens P, De Strooper B, Kivipelto M, et al. Alzheimer's disease. The Lancet. 2021;doi:10.1016/s0140-6736(20)32205-4
2. Marshall GA, Aghjayan SL, Dekhtyar M, et al. Activities of daily living measured by the Harvard Automated Phone Task track with cognitive decline over time in non-demented elderly. J Prev Alzheimers Dis. 2017;4(2):81-86. doi:10.14283/jpad.2017.10
3. Marshall GA, Sikkes SAM, Amariglio RE, et al. Instrumental activities of daily living, amyloid, and cognition in cognitively normal older adults screening for the A4 Study. Alzheimers Dement (Amst). 2020;12(1):e12118. doi:10.1002/dad2.12118
4. Dubbelman MA, Jutten RJ, Tomaszewski Farias SE, et al. Decline in cognitively complex everyday activities accelerates along the Alzheimer's disease continuum. Alzheimers Res Ther. 2020;doi:10.1186/s13195-020-00706-2
5. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, Scheltens P, Uitdehaag BM. A systematic review of Instrumental Activities of Daily Living scales in dementia: room for improvement. J Neurol Neurosurg Psychiatry. Jan 2009;80(1):7-12. doi:10.1136/jnnp.2008.155838
6. Sikkes SAM, De Rotrou J. A qualitative review of instrumental activities of daily living in dementia: what's cooking? Neurodegener Dis Manag. 2014;4(5):393-400. doi:10.2217/nmt.14.24
7. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. Med Care. Sep 2000;38(9 Suppl):II28-42. doi:10.1097/00005650-200009002-00007
8. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol. Nov 2010;63(11):1179-94. doi:10.1016/j.jclinepi.2010.04.011
9. Reise SP, Waller NG. Item response theory and clinical measurement. Annu Rev Clin Psychol. 2009;5:27-48. doi:10.1146/annurev.clinpsy.032408.153553
10. Gorter R, Fox JP, Apeldoorn A, Twisk J. Measurement model choice influenced randomized controlled trial results. J Clin Epidemiol. Nov 2016;79:140-149. doi:10.1016/j.jclinepi.2016.06.011
11. Gorter R, Fox JP, Twisk JW. Why item response theory should be used for longitudinal questionnaire data analysis in medical research. BMC Med Res Methodol. Jul 30 2015;15:55. doi:10.1186/s12874-015-0050-x
12. Jabrayilov R, Emons WHM, Sijtsma K. Comparison of Classical Test Theory and Item Response Theory in Individual Change Assessment. Appl Psychol Meas. Nov 2016;40(8):559-572. doi:10.1177/0146621616664046
13. Kosinski M, Bjorner JB, Ware JE, Batenhorst A, Cady RK. The responsiveness of headache impact scales scored using 'classical' and 'modern' psychometric methods: A re-analysis of three clinical trials. Qual Life Res. Dec 2003;12(8):903-912. doi:10.1023/A:1026111029376
14. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. Alzheimers Dement. Nov 2012;8(6):536-43. doi:10.1016/j.jalz.2011.08.006
15. Dubbelman MA, Verrijp M, Facal D, et al. The influence of diversity on the measurement of functional impairment: An international validation of the Amsterdam IADL Questionnaire in eight countries. Alzheimers Dement (Amst). 2020;12(1):e12021. doi:10.1002/dad2.12021
16. Jutten RJ, Peeters CFW, Leijdesdorff SMJ, et al. Detecting functional decline from normal aging to dementia: Development and validation of a short version of the Amsterdam IADL Questionnaire. Alzheimers Dement (Amst). 2017;8:26-35. doi:10.1016/j.dadm.2017.03.002

17. Sikkes SAM, Knol DL, Pijnenburg YAL, 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(1):35-41. doi:10.1159/000346277

18. Sikkes SAM, Pijnenburg YAL, 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. Dec 2013;26(4):244-50. doi:10.1177/0891988713509139

19. Jutten RJ, Dicks E, Vermaat L, et al. Impairment in complex activities of daily living is related to neurodegeneration in Alzheimer's disease-specific regions. Neurobiol Aging. Nov 26 2018;75:109-116. doi:10.1016/j.neurobiolaging.2018.11.018

20. Koster N, Knol DL, Uitdehaag BM, Scheltens P, Sikkes SAM. The sensitivity to change over time of the Amsterdam IADL Questionnaire((c)). Alzheimers Dement. Oct 2015;11(10):1231-40. doi:10.1016/j.jalz.2014.10.006

21. Solomon A, Kivipelto M, Molinuevo JL, Tom B, Ritchie CW, EPAD Consortium. European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD LCS): study protocol. Bmj Open. Dec 2018;8(12)doi:10.1136/bmjopen-2017-021017

22. Jutten RJ, Harrison J, de Jong FJ, et al. A composite measure of cognitive and functional progression in Alzheimer's disease: Design of the Capturing Changes in Cognition study. Alzheimers Dement (N Y). Jan 2017;3(1):130-138. doi:10.1016/j.trci.2017.01.004

23. Slot RER, Verfaillie SCJ, Overbeek JM, et al. Subjective Cognitive Impairment Cohort (SCIENCe): study design and first results. Alzheimers Res Ther. Aug 7 2018;10(1):76. doi:10.1186/s13195-018-0390-y

24. Samejima F. Estimation of Latent Ability Using a Response Pattern of Graded Scores. Psychometrika. 1969;34(1):1-&. doi:10.1007/BF03372160

25. Westfall J, Kenny DA, Judd CM. Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. J Exp Psychol Gen. Oct 2014;143(5):2020-45. doi:10.1037/xge0000014

26. R: A language and environment for statistical computing. Version 4.1.3. 2022. https://www.R-project.org/

27. Bates D, Machler M, Bolker BM, Walker SC. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. Oct 2015;67(1):1-48. doi:10.18637/jss.v067.i01

28. Thissen D, Pommerich M, Billeaud K, Williams VSL. Item Response Theory for Scores on Tests Including Polytomous Items with Ordered Responses. Appl Psych Meas. 1995;19(1):39-49. doi:10.1177/014662169501900105

29. Cai L. Lord-Wingersky Algorithm Version 2.0 for Hierarchical Item Factor Models with Applications in Test Scoring, Scale Alignment, and Model Fit Testing. Psychometrika. Jun 2015;80(2):535-59. doi:10.1007/s11336-014-9411-3

30. Verrijp M, Dubbelman MA, Visser LNC, et al. Everyday Functioning in a Community-Based Volunteer Population: Differences Between Participant- and Study Partner-Report. Front Aging Neurosci. 2021;13:761932. doi:10.3389/fnagi.2021.761932

31. Bruderer-Hofstetter M, Dubbelman MA, Meichtry A, et al. Cross-cultural adaptation and validation of the Amsterdam Instrumental Activities of Daily Living questionnaire short version German for Switzerland. Health Qual Life Outcomes. Oct 2 2020;18(1):323. doi:10.1186/s12955-020-01576-w

32. Facal D, Carabias MAR, Pereiro AX, et al. Assessing everyday activities across the dementia spectrum with the Amsterdam IADL Questionnaire. Curr Alzheimer Res. 2018;15(13):1261-1266. doi:10.2174/1567205015666180925113411

33. Wang M, Reeve BB. Evaluations of the sum-score-based and item response theory-based tests of group mean differences under various simulation conditions. Stat Methods Med Res. Dec 2021;30(12):2604-2618. doi:10.1177/09622802211043263

34. Schalet BD, Lim S, Cella D, Choi SW. Linking Scores with Patient-Reported Health Outcome Instruments:A VALIDATION STUDY AND COMPARISON OF THREE LINKING METHODS. Psychometrika. Sep 2021;86(3):717-746. doi:10.1007/s11336-021-09776-z

35. Choi SW, Lim S, Schalet BD, Kaat AJ, Cella D. PROsetta: An R Package for Linking Patient-Reported Outcome Measures. Appl Psychol Meas. Jul 2021;45(5):386-388. doi:10.1177/01466216211013106

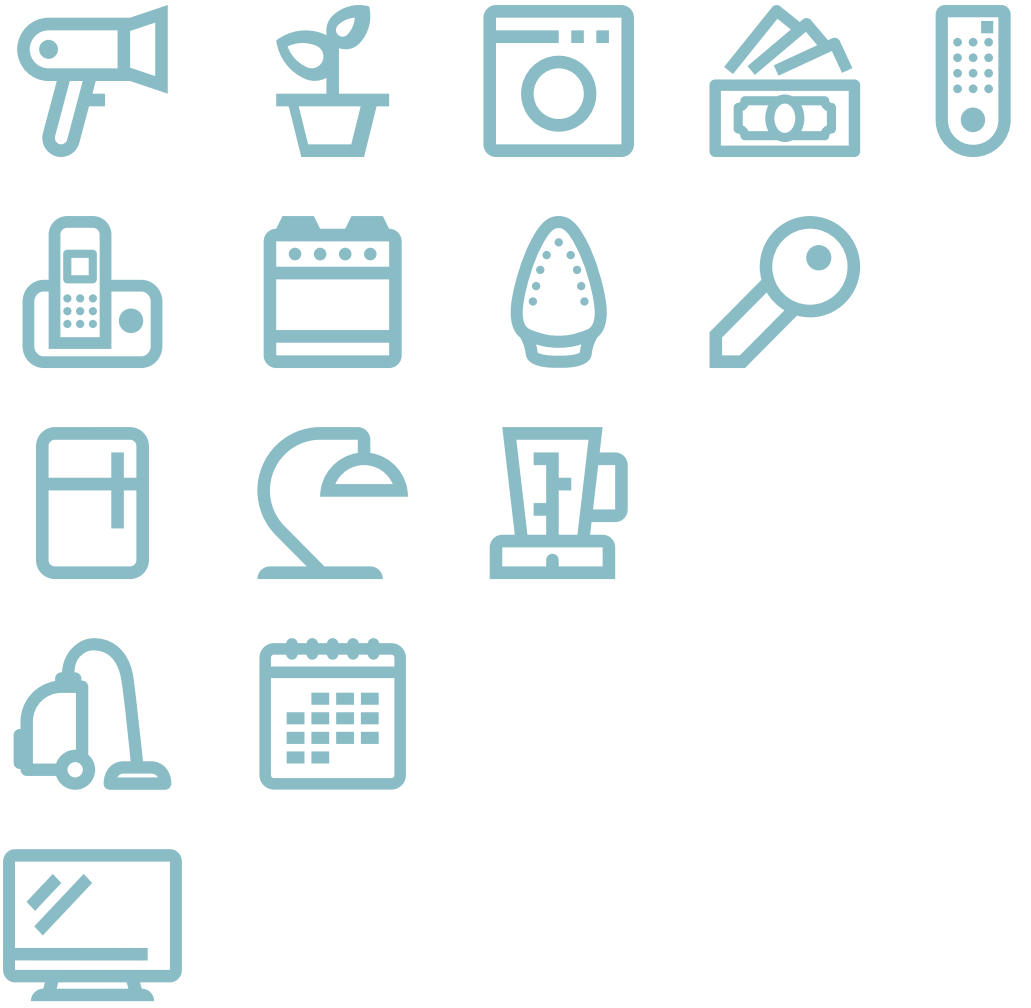
5.5 Supplementary Material

Below is a summarized version of the cross-walk table that can be used to transform CTT-based scores into IRT-based *T*-scores. The full table has 379 rows and is therefore not shown here, but it can be made available upon request.

Supplemental Table 1 | Summarized cross-walk table.

CTT-based score	IRT-based *T*-score		Times simulated	Combinations possible
	Aligned score	Range		
0	68.79	67.38–69.85	124	30
5	61.59	58.08–65.05	183	6
10	58.64	55.25–62.85	375	6
15	56.54	50.85–60.25	318	6
20	54.62	50.50–57.67	144	6
25	52.29	48.48–55.97	188	30
30	48.82	46.47–52.27	26	6
35	46.61	44.25–49.60	140	6
40	45.13	42.19–48.03	484	6
45	43.80	40.94–46.93	995	6
50	42.43	39.17–45.92	5,677	30
55	41.10	37.93–43.59	1,001	6
60	39.90	37.24–42.59	474	6
65	38.52	36.44–41.08	161	6
70	36.96	35.12–38.95	31	6
75	34.29	31.15–37.30	192	30
80	32.55	29.34–36.13	143	6
85	30.40	26.72–33.00	291	6
90	28.16	24.40–31.88	356	6
95	25.44	22.50–29.74	167	6
100	19.46	18.52–21.09	104	30

The value in the 'Aligned score' column represents the IRT-based *T*-score that is aligned with the CTT-based score; it is the mean of all simulated IRT-based *T*-score for the given CTT-based score. The 'Range' column displays the range in simulated IRT-based *T*-scores for each given CTT-based score. The 'Times simulated' column shows the number of times response patterns resulted in each given CTT-based score. The 'Combinations possible' shows how many unique response patterns can result in each given CTT-based score.



CHAPTER 6

TRAJECTORIES OF DECLINE IN COGNITIVELY COMPLEX EVERYDAY ACTIVITIES ON THE ALZHEIMER'S DISEASE CONTINUUM

Mark A. Dubbelman, Roos J. Jutten, Sarah E. Tomaszewski Farias, Rebecca E. Amariglio, Rachel F. Buckley, Pieter Jelle Visser, Dorene M. Rentz, Keith A. Johnson, Michael J. Properzi, Aaron Schultz, Nancy Donovan, Jennifer R. Gatchel, Charlotte E. Teunissen, Bart N.M. van Berckel, Wiesje M. van der Flier, Reisa A. Sperling, Kathryn V. Papp, Philip Scheltens, Gad A. Marshall, Sietske A.M. Sikkes

Published: October 29, 2020

In: Alzheimer's Research and Therapy

Abstract

Background: Impairment in daily functioning is a clinical hallmark of dementia. Difficulties with "instrumental activities of daily living" (IADL) seem to increase gradually over the course of Alzheimer's disease (AD), before dementia onset. However, it is currently not well established how difficulties develop along the preclinical and prodromal stages of AD. We aimed to investigate the trajectories of decline in IADL performance, as reported by a study partner, along the early stages of AD.

Methods: In a longitudinal multicenter study, combining data from community-based and memory clinic cohorts, we included 1,555 individuals (mean age 72.5 ± 7.8 years; 50% female) based on availability of amyloid biomarkers, longitudinal IADL data, and clinical information at baseline. Median follow-up duration was 2.1 years. All amyloid-positive participants (n = 982) were classified into the National Institute on Aging–Alzheimer's Association (NIA-AA) clinical stages ranging from preclinical AD (1) to overt dementia (4+). Cognitively normal amyloid-negative individuals (n = 573) served as a comparison group. The total scores of three study partner-reported IADL questionnaires were standardized.

Results: The rate of decline in cognitively normal (stage 1) individuals with and without abnormal amyloid did not differ (p = .453). However, from stage 2 onwards, decline was significantly faster in individuals on the AD continuum (B [95%CI] = -0.32 [-0.55, -0.09], p = .007). The rate of decline increased with each successive stage: one standard deviation (SD) unit per year in stage 3 (-1.06 [-1.27, -0.85], p < .001) and nearly two SD units per year in stage 4+ (-1.93 [-2.19, -1.67], p < .001). Overall, results were similar between community-based and memory clinic study cohorts.

Conclusions: Our results suggest that the rate of functional decline accelerates along the AD continuum, as shown by steeper rates of decline in each successive NIA-AA clinical stage. These results imply that incremental changes in function are a meaningful measure for early disease monitoring. Combined with the low-cost assessment, this advocates for the use of these functional questionnaires for capturing the effects of early AD-related cognitive decline on daily life.

6.1 Introduction

Alzheimer's disease (AD) pathology, consisting of amyloid-beta plaques, tau neurofibrillary tangles, and neurodegeneration, develops for numerous years before leading to hallmark clinical signs of cognitive and functional impairment.[1] The earliest clinical signs of AD appear to be the subjective experience of memory decline,[2,3] followed by subtle changes in higher-order cognitive functioning. The effects of cognitive decline in daily life can be captured in the performance of cognitively complex "instrumental activities of daily living" (IADL), such as cooking, managing personal and financial paperwork, and keeping appointments.

Impairment in daily functioning is traditionally described as occurring relatively late in the AD disease trajectory, i.e., as a core characteristic of dementia. However, increasing evidence demonstrates that everyday functioning declines gradually over the years preceding clinical diagnosis of dementia. This has been shown in a number of cross-sectional studies,[4-12] as well as a few longitudinal studies.[13-17] Measuring IADL functioning is important, as it is a clinically relevant outcome measure,[18] affecting not only the patient, but also their support system by increasing financial[19] and caregiver burden.[20] Moreover, IADL measures have strong ecological validity, and they are related directly to daily life. As such, they are valuable in both clinical practice and research.

Individuals with abnormal amyloid are in the Alzheimer's continuum, and they can be classified into six stages based on clinical symptom severity, according to the U.S. National Institute on Aging–Alzheimer's Association (NIA–AA) research framework.[1] People in stage 1 (the preclinical phase) do not report a decline in cognition and perform normally on cognitive tests. Stages 2 and 3 (the prodromal phase) are characterized by a self-reported decline in cognition but normal performance on cognitive tests, and the emergence of the first objectified cognitive impairments, respectively. Stages 4 through 6 represent overt dementia with increasing severity of both cognitive and functional impairment. The question remains at what point along the disease trajectory changes in IADL functioning actually start to occur, and how this decline in function develops along the AD continuum.

In this study, we focus especially on the earliest stages (1–3), as we hypothesize that a decline in function may already be present here. We aimed to determine how IADL functioning progresses along the AD continuum, as well as to identify the stage in which decline in functioning is accelerated compared to amyloid-negative, cognitively normal controls. Finally, we aimed to investigate specific activities in

more detail, to determine whether there were any differences in the advent of problems and rate of decline in relatively easy and relatively complex IADLs.

6.2 Methods

Study cohorts and selection criteria

We selected subjects from six cohorts: Harvard Aging Brain Study (HABS, *n* = 259), Alzheimer's Disease Neuroimaging Initiative (ADNI, *n* = 829), National Alzheimer's Coordinating Center (NACC, *n* = 201), Amsterdam Dementia Cohort (ADC, *n* = 178), European Medical Information Framework (EMIF)-AD PreclinAD Study (*n* = 73), and EMIF-AD 90+ Study (*n* = 15).

Specific procedures have been described for each cohort in detail elsewhere.[21-26] Briefly, HABS is a prospective community-based cohort that consists of individuals aged 65 years and older, who are considered cognitively normal at study inclusion.[24] ADNI is a multicenter longitudinal cohort study. For the present study, we obtained baseline and follow-up data acquired for ADNI-GO and ADNI-2.[23] The NACC database contains mostly memory-clinic referred subjects with additional community recruitment.[22] The ADC is a memory-clinic cohort comprised of patients of the Alzheimer Center Amsterdam.[21] In the EMIF-AD PreclinAD Study, cognitively normal subjects aged 60 years and older were included.[25] The EMIF-AD 90+ Study focused on people aged 90 years and older who were either cognitively normal or who had some cognitive impairment.[26]

For the present study, subjects were selected based on (1) availability of amyloid biomarkers at baseline, (2) sufficient information to determine NIA–AA clinical staging at baseline, and (3) availability of longitudinal IADL data, defined as having at least one follow-up assessment. All data used in this study were collected between June 2002 and July 2019.

All studies were approved by ethical review boards, and all subjects provided written informed consent for the use of their data for research purposes, in accordance with the Declaration of Helsinki.

Amyloid

Amyloid status was assessed at baseline using either amyloid positron emission tomography (PET) imaging or cerebrospinal fluid (CSF) using local procedures,

such as described in more detail elsewhere.[25-30] PET scans, using ^{11}C-Pittsburgh compound-B (PiB) in HABS, ^{18}F-florbetapir in ADNI, and one of ^{11}C-PiB, ^{18}F-flutemetamol, ^{18}F-florbetapir, or ^{18}F-florbetaben in the ADC and EMIF cohorts, were judged either using standard uptake volume ratios (ADNI, NACC), distribution volume ratios (HABS), or visual rating by independent nuclear medicine physicians (ADC, both EMIF studies). For CSF, local cutoffs for amyloid positivity were used (NACC, ADC). Where both PET and CSF were available for the same individual (n = 66), PET results were favored. Both amyloid-positive and amyloid-negative individuals were included. Additional details about amyloid assessment can be found in the Supplementary Material.

Clinical stages

Amyloid-positive individuals were categorized into four clinical stages according to the NIA–AA framework,[1] based on baseline measures of subjective cognitive complaints, cognitive performance, and global functional impairment. This procedure and the measures used are described in detail by Jutten et al.[31] Briefly, we considered a visit to a memory clinic, or a positive response to a subjective cognitive decline questionnaire as an indication of subjective complaints. Cognitive performance was determined using the scores on a general cognitive screener and a story or list learning task. Finally, functional impairment was determined using a global dementia rating scale. The IADL instruments used as outcomes were not used to determine the stages. Baseline stages are defined as follows: (1) no complaints and no cognitive deficits, objectified using standard neuropsychological testing; (2) subjective complaints but no objectified cognitive deficits; (3) mild objectified cognitive deficits; and (4+) clinically manifest dementia. We did not distinguish between the NIA–AA stages 4, 5 and 6, as the focus of the current investigation was on the preclinical (1) and prodromal stages (2–3).

Cognitively normal amyloid-negative individuals without cognitive complaints or objectified deficits were included as available from the same cohorts, as a comparison group.

IADL measures

Three study partner-reported IADL instruments were used: the Functional Activities Questionnaire (FAQ), Everyday Cognition (ECog), and Amsterdam IADL Questionnaire (A-IADL-Q).

The FAQ is a 10-item scale.[32] Each item is rated from 0 (no difficulty or independent) to 3 (dependent), as compared to performance one month earlier. We summed all items to compute a total score, ranging from 0 to 30, with higher scores indicating more functional dependence. The ECog is a questionnaire comprised of 39 items reflecting cognitively complex everyday activities across six subscales, including memory, language, and executive functioning.[33,34] All items are rated from 1 (no change in function compared to ten years ago) to 4 (consistently much worse function than ten years ago). Total scores are a weighted average ranging from 1 to 4, with higher scores indicating more problems in everyday functioning. The A-IADL-Q is aimed at assessing cognitively complex, relevant everyday activities.[35] It has been extensively validated.[36-40] Item scores range from 0 (no difficulty performing the activity) to 4 (unable to perform the activity), comparing current performance to the past. Total scores are calculated using item response theory and have a mean score of 50 with a standard deviation (SD) of 10 in a memory clinic population. Higher scores indicate better functioning.

Harmonizing IADL measurements

FAQ and ECog raw total scores were inverted so that higher scores represent better functioning. Individual instrument total scores were converted to Z-scores using the baseline mean and SD of the entire amyloid-negative subsample. Next, a single Z-score was created by pooling the individual instrument Z-scores into one. In instances where individuals had both a completed FAQ and ECog, we first averaged the Z-scores of the FAQ and ECog, before combining them into the final Z-score. The final IADL Z-score is thus a standardized measure of IADL performance, with higher scores representing better functioning. A one-unit difference in the IADL Z-score represents a change of one SD in functioning among cognitively normal, amyloid-negative individuals.

Furthermore, we harmonized items that referenced the same activities and were shared between the instruments. To illustrate how specific IADLs develop over time, we selected two of these activities on opposite ends of the spectrum of IADL complexity: one relatively easy item (“preparing hot beverages”), and one relatively complex item (“managing the paperwork”). The selection was made a priori on the basis of A-IADL-Q item parameters, as presented by Jutten et al.[37] The easier item may not be impaired until a relatively high level of overall IADL impairment has been reached, whereas the more complex item may already be impaired at a lower level of IADL impairment. The harmonized items are shown in Table 1.

Table 1 | Harmonization of items and response options from the FAQ, ECog, and A-IADL-Q.

Harmonization	FAQ	ECog	A-IADL-Q
Cohort(s)	ADNI, NACC	ADNI, HABS	ADC, EMIF pre-AD and 90+
Item content			
Hot beverages	Heating water, making a cup of coffee, turning off the stove	—	Using the coffee maker
Paperwork	Assembling tax records, business affairs, or other papers	Keeping financial records organized	Managing the household paperwork
Response options			
Normal (4)	Normal (0)	Better or no change (1)	No more difficult (0)
Slightly worse (3)	Has difficulty, but does by self (1)	Questionable/ occasionally worse (2)	Slightly more difficult (1)
Worse (2)	—	Consistently a little worse (3)	More difficult (2)
Much worse (1)	Requires assistance (2)	—	Much more difficult (3)
Unable (0)	Dependent (3)	Consistently much worse (4)	No longer able to perform this task (4)

Abbreviations: ADC, Amsterdam Dementia Cohort; ADNI, Alzheimer's Disease Neuroimaging Initiative; A-IADL-Q, Amsterdam Instrumental Activities of Daily Living Questionnaire; ECog, Everyday Cognition; EMIF, European Medical Information Framework; FAQ, Functional Activities Questionnaire; HABS, Harvard Aging Brain Study; NACC, National Alzheimer's Coordinating Center.

Statistical analyses

Linear or logistic regressions were used to investigate baseline group differences between the amyloid-negative group and each of the four NIA–AA stages. Significance was set at $p < 0.01$. To analyze change over time in IADL functioning, linear mixed models (LMMs) with random intercepts and slopes were run using the "lme4" package version 1.1-27[41] for R. LMMs are a powerful method for analyzing change over time when handling unbalanced data, including inconsistent time intervals between follow-up measurements and missing data.[41] We fitted models in which the IADL Z-score was the dependent variable, and time in years was the main independent variable. Interactions between stage and time were included to determine slopes for each stage, treating the amyloid-negative group as 'stage 0' for convenience. Adjustments for clustering within study cohorts, as well as for age

at baseline, sex, and education, were also included. Unstandardized estimates and 95% confidence intervals are reported for fixed effects. Finally, we ran sensitivity analyses to investigate potential differences between community-based and memory clinic studies, as well as the influence of each cohort. We ran ordinal logistic mixed-effects models on the two activities, similar to the main analyses. All analyses were run in R version 4.0.2[42] and Stata version 14.[43]

6.3 Results

Sample characteristics

A total of 1,555 individuals were included (age 72.5±7.8 years old; 49.8% female), of whom 982 were amyloid positive. Mean age did not differ between amyloid positive and amyloid negative individuals ($p = .619$). Amyloid positive individuals had received fewer years of education and had lower MMSE scores at baseline (both $p < .001$) than amyloid negative individuals. Table 2 displays the baseline characteristics of the amyloid negative and amyloid positive groups. Characteristics per cohort can be found in the Supplementary Material.

All amyloid-positive individuals were classified into one of the NIA–AA clinical stages at baseline: 120 individuals (12%) were in stage 1, 160 (16%) in stage 2, 464 (47%) in stage 3, and the remaining 238 (24%) in stage 4+. Individuals in stages 1 and 2 were older than those in stages 3 and 4, had more years of education, and were more likely to be female (Table 2).

Table 2 | Baseline demographics.

	Total group (*n* = 1,555)	Amyloid negative (*n* = 573)	Amyloid positive (*n* = 982)	Stage 1 (*n* = 120)	Stage 2 (*n* = 160)	Stage 3 (*n* = 464)	Stage 4+ (*n* = 238)	Post hoc group differences
Years of follow-up, mean (range)	2.79 (0.25–7.09)	3.31 (0.46–7.07)	2.49 (0.25–7.09)	2.94 (0.56–6.38)	2.96 (0.46–7.06)	2.69 (0.25–7.09)	1.36 (0.26–4.45)	Amyloid negative, stages 1, 2 > 3 > 4
Age	72.45 ± 7.8	72.82 ± 7.1	72.24 ± 8.2	74.72 ± 6.5	74.19 ± 7.7	71.79 ± 8.0	70.55 ± 9.1	Stage 1, 2 > amyloid negative > stage 3, 4
Female, n (%)	763 (50)	311 (55)	452 (47)	66 (56)	84 (54)	196 (43)	106 (45)	Amyloid negative, stages 4, 2, 1 > 3
Education years	15.73 ± 3.3	15.94 ± 3.1	15.05 ± 3.4	15.94 ± 3.2	15.68 ± 3.2	15.13 ± 3.3	14.01 ± 3.7	Amyloid negative, stages 1, 2 > 3 > 4
MMSE	27.08 ± 3.8	29.11 ± 1.1	25.62 ± 4.3	29.06 ± 1.0	28.45 ± 1.7	26.43 ± 2.7	20.08 ± 4.2	Amyloid negative, stage 1 > 2 > 3 > 4
Cohorts, n (%)								
HABS	259 (17)	194 (34)	65 (6)	38 (31)	27 (17)	—	—	
ADNI	829 (53)	323 (55)	506 (52)	49 (42)	72 (45)	288 (62)	98 (41)	
NACC	201 (13)	—	201 (21)	29 (25)	23 (14)	84 (18)	65 (27)	
ADC	178 (11)	—	178 (18)	—	15 (9)	87 (19)	76 (32)	
EMIF-AD	73 (5)	53 (9)	20 (2)	2 (2)	18 (11)	—	—	
EMIF-90+	15 (1)	3 (1)	12 (1)	2 (2)	5 (3)	5 (1)	—	

All are displayed as mean ± standard deviation, except as stated otherwise. Missing: Sex (*n* = 22), education (*n* = 13), MMSE (*n* = 192), age (*n* = 14). Group differences between amyloid negative and each of the four NIA-AA stages are based on linear (for all but sex) or logistic (for sex) regressions.

Abbreviations: ADC, Amsterdam Dementia Cohort; ADNI, Alzheimer's Disease Neuroimaging Initiative; EMIF, European Medical Information Framework; HABS, Harvard Aging Brain Study; MMSE, Mini-Mental State Examination; NACC, National Alzheimer's Coordinating Center.

Overall IADL functioning trajectories

At baseline, 1,077 participants completed the ECog, 1,025 completed the FAQ, and 266 completed the A-IADL-Q. The correlation between the ECog and FAQ was r = .83 (95% confidence interval (95%CI) = [.81, .85], n = 819).

At baseline, amyloid negative (mean (M) ± standard deviation (SD) = 0.05 ± 0.9) and amyloid positive individuals in stage 1 (0.18 ± 0.6) had similar levels of IADL functioning, on average (p = .631). Those in stages 2 (-0.60 ± 1.6), 3 (-3.76 ± 3.3), and 4+ (-8.75 ± 4.3) each had lower baseline functioning. IADL functioning remained fairly stable over time in cognitively normal amyloid negative individuals (B = -0.08, 95%CI = [-0.28, 0.14], p = .453). In contrast, a substantial decline in IADL functioning was found in the amyloid positive group as a whole (-0.95, 95%CI = [-1.20, -0.69], p < .001; see Table 3).

We found that, as a group, individuals in stage 1 showed a small, non-significant decline in IADL functioning over time (B = -0.12, 95%CI = [-0.37, 0.13], p = .342). The rate of decline was only marginally larger than in the amyloid negative group, and this difference was also not significant. Individuals in stage 2 declined significantly (B = -0.32, 95%CI = [-0.55, -0.09], p = .007), as did individuals in stages 3 (B = -1.06, 95%CI = [-1.27, -0.85], p < .001) and 4+ (B = -1.93, 95%CI = [-2.19, -1.67], p < .001). Moreover, when comparing the slopes in all stages and amyloid negative controls with each other, there was a significant time × stage interaction for all stages, except the first stage (Table 3). The rate of decline accelerated with each successive stage (stage 1, -0.12; stage 2, -0.32; stage 3, -1.06; stage 4+, -1.93; Table 3), compared to amyloid-negative individuals.

Figure 1 displays the individual trajectories and group slopes of IADL decline for each stage. As can be seen in Figure 1, there was a large variability in slopes between individuals in the AD continuum.

We additionally investigated the trajectories for community-based and memory clinic study cohorts and found that the results were largely similar, except that in community-based studies, the decline observed in stage 2 was not significantly different from the change in amyloid negatives. The results from these analyses can be found in the Supplementary Material.

Table 3 | Linear mixed model results of change over time in IADL functioning at baseline for amyloid negatives and amyloid positives, divided into the NIA-AA stages.

Group	b	95%CI	P value
Intercepts			
Amyloid negative	-1.48	[-3.46, 0.31]	—[a]
Amyloid positive	-7.31	[-9.88, -4.74]	< .001[a]
Stage 1	-0.80	[-2.64, 1.04]	.631[a]
Stage 2	-1.65	[-3.48, 0.17]	.005[a]
Stage 3	-4.67	[-6.43, -2.92]	< .001[a]
Stage 4+	-9.64	[-11.41, -7.88]	< .001[a]
Intercepts			
Amyloid negative	-0.08	[-0.28, 0.14]	.453
Amyloid positive	-0.94	[-1.20, -0.69]	< .001
Stage 1	-0.12	[-0.37, 0.13]	.342
Stage 2	-0.32	[-0.55, -0.09]	.007
Stage 3	-1.06	[-1.27, -0.85]	< .001
Stage 4+	-1.93	[-2.19, -1.67]	< .001

Shown here are unstandardized betas, adjusted for clustering within study, as well as for baseline age, gender, and years of education. The betas represent *Z*-score intercepts and yearly change (stage and time by stage interactions).

[a] Compared to amyloid negative group.

Abbreviations: IADL, instrumental activities of daily living; 95%CI, 95% confidence interval.

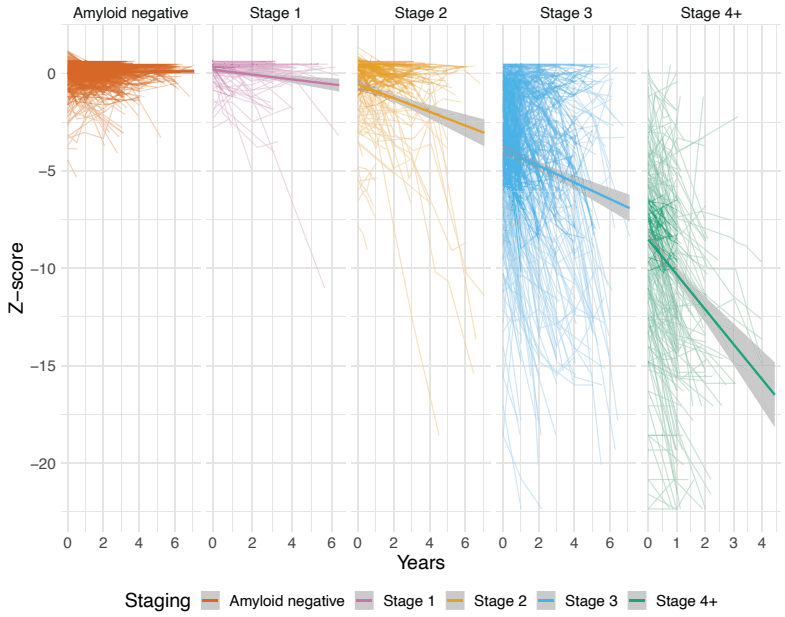


Figure 1 | Individual and group average trajectories per clinical stage for the global IADL Z-score.

The trajectories show that, at the group level, there is no decline in amyloid negative individuals, but it does appear to be present in the earliest AD stages, and it increases with each subsequent stage. A one-unit change in the Z-score represents one standard deviation in the amyloid negative group.

Activity-specific trajectories

Both the relatively easy "preparing hot beverage" (B = -0.68, 95%CI = [-0.80, -0.57]) and the more complex "managing paperwork" (B = -0.66, 95%CI = [-0.75, -0.57]) showed a similar, significant decline in the amyloid-positive group as a whole. Compared to the amyloid-negative group, individuals in stage 2 declined significantly faster on preparing hot beverages ($p < .001$), whereas on managing paperwork, even individuals in stage 1 declined significantly faster ($p = .007$). For both activities, there were no significant differences in rate of decline between stages 2 and 3. Those in stage 4+ declined the fastest on both activities. Individual item responses are shown in Figure 2.

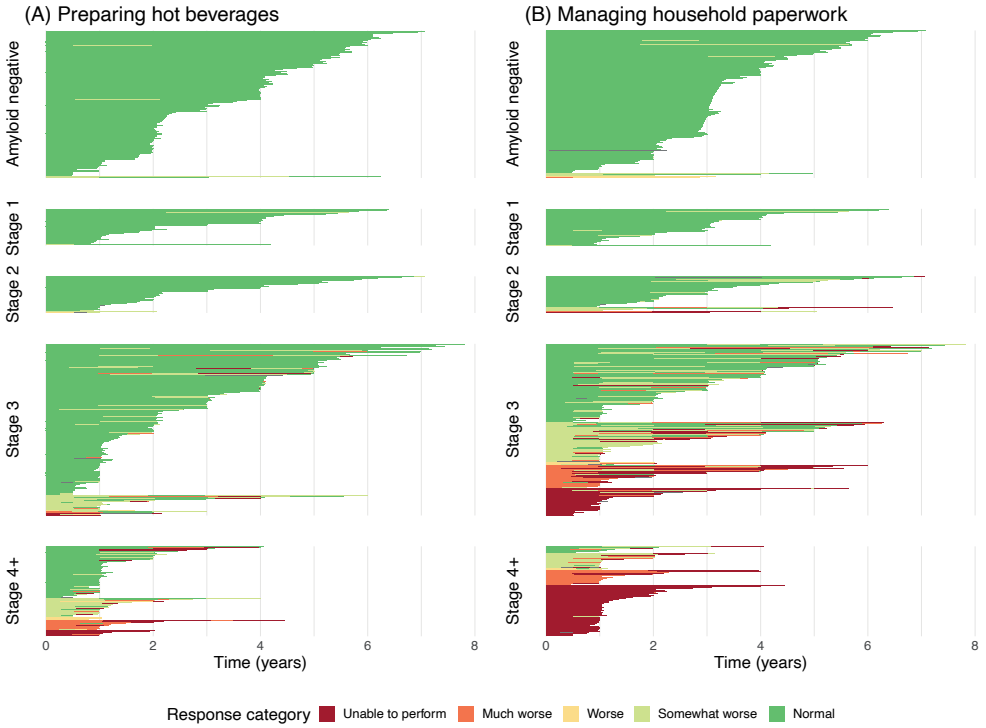


Figure 2 | Individual response categories on two pooled activities: (a) "preparing hot beverages" and (b) "managing the paperwork".
Each horizontal line represents an individual (on the y-axis), with longer lines representing longer follow-up (time in years, on the x-axis). The lines are colored based on the level of difficulty the individual had over the course of their follow-up, ranging from dark green (normal performance) to dark red (unable to perform). Individuals are grouped by NIA-AA clinical stage.

6.4 Discussion

We demonstrated that the decline in daily functioning accelerates as AD progresses along the continuum from preclinical to symptomatic. This decline was distinct from the functional change observed in amyloid-negative, cognitively normal individuals. Furthermore, our data suggest that more complex activities, such as managing the paperwork, are especially sensitive to the earliest cognitive changes, showing a decline in the early prodromal stage.

Functional impairment has long been considered a defining feature of the transition from mild cognitive impairment to the dementia stage of AD.[44,45] Accumulating evidence over the past decade shows that difficulties in cognitively complex activities may be seen in cognitively normal individuals who later progress to dementia.[46-48] This might indicate that these individuals have a lower level of functioning to start with, or it might suggest a decline in the pre-dementia stages. These findings have not previously been investigated in the context of the newly proposed NIA–AA stages. The staging criteria propose that detectible but mild functional impairment may be found in stage 3 and beyond, when performance on objective cognitive tests becomes impaired.[1] We present evidence that decline in functional impairment may already be present in earlier stages. The decline we observed in stage 1 did not differ from the change in amyloid-negative individuals and may have been too subtle to be distinguished from normal aging-related decline in everyday functioning. Others have previously stated that sensitive measures of cognitive and functional impairment are needed to monitor disease progression in early stages, or to evaluate drug effectiveness in the context of AD clinical trials.[49,50] More convincingly, we observed a decline in stage 2, before cognitive decline can be objectively measured using traditional cognitive tests. When compared to cognitively normal amyloid-negative individuals, there was a faster decline in all subsequent stages from stage 2 onward. In community-based studies, individuals in stage 2, which corresponds approximately to the concept of subjective cognitive decline (SCD), did not decline at a significantly different rate than amyloid-negative, cognitively normal individuals. It is possible that stage 2 individuals who have not visited a memory clinic may be in some way different from those who have. Slot and colleagues[3] have previously shown that people with SCD who visited a memory clinic had an increased risk of progressing to dementia, compared to those who were included in community-based studies.

Our findings demonstrate that functional decline co-occurs with the earliest changes in cognition in the context of AD, revealing the importance of assessing daily functioning in addition to cognitive functioning, particularly in early stages. Decline in cognition is assumed to cause functional impairment, not vice versa. However, many frequently used cognitive measures might not be sensitive enough to detect subtle cognitive changes.[51-53] Our results justify combining sensitive IADL measures with sensitive cognitive tests for detecting such changes. Study partner-reported functional questionnaires have additional advantages in that they are easy to administer, have good ecological validity, and are strongly related to quality of life.[18,52] As such, our findings implicate an important benefit of including the measurement of everyday functioning in early AD stages for the evaluation of disease progression and potential intervention effectiveness, in addition to providing potential starting points for early non-pharmacological interventions targeting cognitive functioning.

Limitations

This study had a few limitations. Three questionnaires assessing slightly different aspects of cognitively complex everyday functioning and in reference to differing time frames were combined, and total scores were placed on a single scale by computing Z-scores for each instrument and merging them into a single score. Of all included activities, only a handful overlapped between all three questionnaires. Overall, however, they provide information about the same construct: higher-order cognitive functioning in everyday life, which was partly evidenced by the high correlation between two of the three measures. This justifies the combination of total scores into a single functional measure. Future undertakings could adapt a more sophisticated linking method, e.g., by using item response theory, giving more weight to questionnaires with favorable psychometric properties. A second limitation was that amyloid positivity was assessed using different techniques (i.e., PET and CSF) and using local cut-offs, so that an individual found positive in one cohort might not have been found positive in another. The average follow-up time was approximately three years. For the early stages (1 and 2), three years is a relatively short period of time, as preclinical AD duration is estimated to be about 10 years.[54] Further, we did not include longitudinal assessment of cognition and can therefore not be sure whether participants progressed from one clinical stage to the next. In consequence, it should be taken into account that our trajectories of change might not reflect each stage's entire duration. Finally, our study sample

was comprised of convenience samples with relatively highly educated and mostly Caucasian participants. This has potentially caused a sample bias, and our findings may therefore not be directly applied to the global population.

Strengths

An important strength of this study was the large number of amyloid-positive individuals with a large age range and representing the entire AD continuum, who were followed over time and recruited in different study settings from both the USA and the Netherlands. By combining data from different cohorts, we aimed to overcome at least in part the sample bias. We ran sensitivity analyses (in the Supplementary Material) and found that results were robust when removing either one of the cohorts, suggesting that the results are not driven by a single measure or cohort, supporting the robustness of our findings. Another strength was our approach to define clinical stages of severity, by using a careful operationalization of the NIA–AA clinical staging scheme and grouping individuals into four different clinical stages, which is a more refined method than relying solely on diagnostic status.[55] Additionally, IADL functioning as determined by the three questionnaires was not part of the staging criteria used in the current study, which has been a confound in many previous studies that divided groups into MCI and dementia. However, it must be noted that the staging was not completely independent of IADL as a construct, and that clinicians who determined disease symptom severity may not always have been blinded to the IADL scores, which may have influenced their classification. Our inclusion of an amyloid-negative comparison group indicates that the decline in IADL is disease-specific and not a general aging effect.

Future research should include the other two major components of the NIA–AA model of AD, tau and neurodegeneration, to further investigate the relationship between function and AD pathology. Future studies should also incorporate and combine longitudinal clinical staging, so the continuous progression of cognitive and functional performance along the AD clinical spectrum can be investigated. Because functional impairment is not unique to AD, future research should replicate our study in other neurodegenerative diseases, to investigate the relationship between other types of neurodegeneration and IADL functioning. Furthermore, as we noticed a lot of intra- and inter-individual variability in the change over time, it would be interesting to delve into these individual differences in future studies. Finally, we currently do not know when changes in functioning actually affect

a person's ability to function independently. As such, investigating the clinical meaningfulness of these changes would be an important future endeavor.

Conclusion

To conclude, our findings suggest that increased difficulties with cognitively complex everyday activities may constitute a useful marker of early cognitive decline in the pre-dementia stage of AD. Thus, the assessment of these complex activities may provide valuable information about the severity of cognitive symptoms, especially when measured longitudinally. Incorporating IADL measures alongside cognitive tests would allow for within-individual everyday decline to be gauged in a cost- and time-effective way. We therefore recommend including a measure of functional difficulties in clinical trials at the stage of preclinical AD, as well as in clinical practice.

Declarations

Acknowledgments

The authors thank all patients and participants in the studies, as well as everyone involved in data collection and data sharing.

Part of the data used in preparation of this article were obtained from the Alzheimer Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: https://adni.loni.usc.edu/wp/content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Authors' contributions

M.A.D. designed and conceptualized study, analyzed and interpreted the data, wrote the manuscript. He had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. R.J.J. designed and conceptualized study, supervised the project and performed multiple revisions and integration of comments for the manuscript. S.E.T.F. collected and provided data; revised the manuscript. R.E.A. collected and provided data, gave methodological advice, and revised the manuscript. R.F.B. gave methodological advice and revised the manuscript. P.J.V. collected and provided

data and revised the manuscript. D.M.R. collected and provided data and revised the manuscript. K.A.J. collected and provided data and revised the manuscript. M.J.P. assisted with data-management and revised the manuscript. A.S. assisted with data-management and revised the manuscript. N.D. assisted with data-management and revised the manuscript. J.R.G. assisted with data-management and revised the manuscript. C.E.T. collected and provided data and revised the manuscript. B.N.M.B. collected and provided data and revised the manuscript. W.M.F. collected and provided data and revised the manuscript. R.A.S. collected and provided data and revised the manuscript. K.V.P. collected and provided data and revised the manuscript. P.S. collected and provided data and revised the manuscript. G.A.M. designed and conceptualized study, supervised the project, made multiple revisions, and integrated comments for the manuscript. S.A.M.S. designed and conceptualized study, supervised the project, made multiple revisions, and integrated of comments for the manuscript. The authors read and approved the final manuscript.

Funding

The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

The Harvard Aging Brain Study is funded by the National Institute on Aging (P01AG036694; Principal Investigators R.A.S., K.A.J.) with additional support from several philanthropic organizations. K.V.P. (1K23AG053422-01) is supported by a K23 award from NIA and an award from the Alzheimer's Association.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health, Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Bio- gen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Develop- ment, LLC; Johnson & Johnson Pharmaceutical Research & Development

LLC; Lumosity; Lundbeck; Merck&Co, Inc.; Meso ScaleDiagnostics, LLC; NeuroRx Re- search; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging;Servier;Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (http://www. fnih.org/). The grantee organization is the Northern California Institute for Re- search and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuroimaging at the University of Southern California.

The NACC database is funded by NIA/NIH Grant U01 AG016976. NACC data are contributed by the NIA-funded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P30 AG062428-01 (PI James Leverenz, MD), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI An- drew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P30 AG062421-01 (PI Bradley Hyman, MD, PhD), P30 AG062422-01 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wisniewski, MD), P30 AG013854 (PI Robert Vassar, PhD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD, MS), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P30 AG062429-01(PI James Brewer, MD, PhD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG053760 (PI Henry Paulson, MD, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P30 AG049638 (PI Suzanne Craft, PhD), P50 AG005136 (PI Thomas Grabowski, MD), P30 AG062715-01 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), and P50 AG047270 (PI Stephen Strittmatter, MD, PhD).

The Alzheimer Center Amsterdam is supported by Alzheimer Nederland and Stichting VUmc Fonds. The clinical database structure of the Amsterdam Dementia Cohort (ADC) was developed with funding from Stichting Dioraphte.

The work for the EMIF cohorts has received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking (EMIF grant no. 115372). It also received in kind sponsoring of the CSF assay from ADx NeuroSciences and Euroimmun, and the PET-tracer 18F-flutemetamol from GE Healthcare.

Availability of data and materials

The data and materials used in this study can be made available by the corresponding author upon request.

Ethics approval and consent to participate

The studies of which we used data in this study have been approved by their institutions' respective ethical review boards. All subjects provided written informed consent for the use of their data for research purposes, in accordance with the Declaration of Helsinki.

Competing interests

M.A.D., R.J.J., S.E.T.F., R.E.A., R.F.B., P.J.V., D.M.R., K.A.J., M.J.P., A.S., N.D., J.R.G., B.N.M.B. and R.A.S. report no disclosures relevant to this manuscript. 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. G.A.M. has received research salary support from Eisai Inc., Eli Lilly and Company, Janssen Alzheimer Immunotherapy, Novartis, and Genentech. Addition- ally, GAM has served as a consultant for Grifols Shared Services North America, Inc., Eisai Inc., and Pfizer. W.M.F. holds the Pasman chair. C.E.T. received grants from the European Commission, the Dutch Research Council (ZonMW), Association of Frontotemporal Dementia/Alzheimer's Drug Discovery Foundation, The Weston Brain Institute, Alzheimer Nederland. C.E.T. has functioned in advisory boards of Roche, received non-financial support in the form of research consumables from ADxNeurosciences and Euroimmun, performed contract research, or received grants from Probiodrug, Biogen, Esai, Toyama, Janssen prevention center, Boehringer, AxonNeurosciences, EIP farma, PeopleBio, and Roche. K.V.P. has served as a paid consultant for Biogen.

References

1. Jack CR, Jr., Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement. Apr 2018;14(4):535-562. doi:10.1016/j.jalz.2018.02.018
2. Jessen F, Amariglio RE, Van Boxtel M, et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimers Dement. Nov 2014;10(6):844-52. doi:10.1016/j.jalz.2014.01.001
3. Slot RER, Sikkes SAM, Berkhof J, et al. Subjective cognitive decline and rates of incident Alzheimer's disease and non-Alzheimer's disease dementia. Alzheimers Dement. Mar 2019;15(3):465-476. doi:10.1016/j.jalz.2018.10.003
4. Albert MS, DeKosky ST, Dickson D, 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. May 2011;7(3):270-9. doi:10.1016/j.jalz.2011.03.008
5. Marshall GA, Rentz DM, Frey MT, et al. Executive function and instrumental activities of daily living in mild cognitive impairment and Alzheimer's disease. Alzheimers Dement. May 2011;7(3):300-8. doi:10.1016/j.jalz.2010.04.005
6. Peres K, Helmer C, Amieva H, et al. Natural history of decline in instrumental activities of daily living performance over the 10 years preceding the clinical diagnosis of dementia: a prospective population-based study. J Am Geriatr Soc. Jan 2008;56(1):37-44. doi:10.1111/j.1532-5415.2007.01499.x
7. Amieva H, Le Goff M, Millet X, et al. Prodromal Alzheimer's disease: successive emergence of the clinical symptoms. Ann Neurol. Nov 2008;64(5):492-8. doi:10.1002/ana.21509
8. Triebel KL, Martin R, Griffith HR, et al. Declining financial capacity in mild cognitive impairment: A 1-year longitudinal study. Neurology. Sep 22 2009;73(12):928-34. doi:10.1212/WNL.0b013e3181b87971
9. Goldberg TE, Koppel J, Keehlisen L, et al. Performance-based measures of everyday function in mild cognitive impairment. Am J Psychiatry. Jul 2010;167(7):845-53. doi:10.1176/appi.ajp.2010.09050692
10. Morris JC. Revised Criteria for Mild Cognitive Impairment May Compromise the Diagnosis of Alzheimer Disease Dementia. Archives of Neurology. Jun 2012;69(6):700-708. doi:10.1001/archneurol.2011.3152
11. Marshall GA, Dekhtyar M, Bruno JM, et al. The Harvard Automated Phone Task: new performance-based activities of daily living tests for early Alzheimer's disease. J Prev Alzheimers Dis. Dec 2015;2(4):242-253. doi:10.14283/jpad.2015.72
12. Czaja SJ, Loewenstein DA, Sabbag SA, Curiel RE, Crocco E, Harvey PD. A novel method for direct assessment of everyday competence among older adults. Journal of Alzheimer's Disease. 2017;57(4):1229-1238. doi:10.3233/Jad-161183
13. Marshall GA, Aghjayan SL, Dekhtyar M, et al. Activities of daily living measured by the Harvard Automated Phone Task track with cognitive decline over time in non-demented elderly. J Prev Alzheimers Dis. 2017;4(2):81-86. doi:10.14283/jpad.2017.10
14. Zoller AS, Gaal IM, Royer CA, et al. SIST-M-IR activities of daily living items that best discriminate clinically normal elderly from those with mild cognitive impairment. Curr Alzheimer Res. 2014;11(8):785-791. doi:Doi 10.2174/1567205011081409101222220

15. Marshall GA, Zoller AS, Kelly KE, et al. Everyday Cognition scale items that best discriminate between and predict progression from clinically normal to mild cognitive impairment. Curr Alzheimer Res. 2014;11(9):853-861.
16. Marshall GA, Zoller AS, Lorius N, et al. Functional Activities Questionnaire items that best discriminate and predict progression from clinically normal to mild cognitive impairment. Curr Alzheimer Res. 2015;12(5):493-502.
17. Verlinden VJA, Van der Geest JN, De Bruijn RFAG, Hofman A, Koudstaal PJ, Arfan Ikram M. Trajectories of decline in cognition and daily functioning in preclinical dementia. Alzheimers Dement. Feb 2016;12(2):144-153. doi:10.1016/j.jalz.2015.08.001
18. Edgar CJ, Vradenburg G, Hassenstab J. The 2018 Revised FDA Guidance for Early Alzheimer's Disease: Establishing the Meaningfulness of Treatment Effects. J Prev Alz Dis. 2019;6doi:10.14283/jpad.2019.30
19. Springate BA, Tremont G. Dimensions of caregiver burden in dementia: impact of demographic, mood, and care recipient variables. Am J Geriatr Psychiatry. Mar 2014;22(3):294-300. doi:10.1016/j.jagp.2012.09.006
20. Jutten RJ, Harrison J, de Jong FJ, et al. A composite measure of cognitive and functional progression in Alzheimer's disease: Design of the Capturing Changes in Cognition study. Alzheimers Dement (N Y). Jan 2017;3(1):130-138. doi:10.1016/j.trci.2017.01.004
21. van der Flier WM, Scheltens P. Amsterdam Dementia Cohort: Performing Research to Optimize Care. J Alzheimers Dis. 2018;62(3):1091-1111. doi:10.3233/JAD-170850
22. Beekly DL, Ramos EM, Lee WW, et al. The National Alzheimer's Coordinating Center (NACC) database: the Uniform Data Set. Alzheimer Dis Assoc Dis. Jul-Sep 2007;21(3):249-58. doi:10.1097/WAD.0b013e318142774e
23. Aisen PS, Petersen RC, Donohue MC, et al. Clinical Core of the Alzheimer's Disease Neuroimaging Initiative: progress and plans. Alzheimers Dement. May 2010;6(3):239-46. doi:10.1016/j.jalz.2010.03.006
24. Dagley A, LaPoint M, Huijbers W, et al. Harvard Aging Brain Study: Dataset and accessibility. Neuroimage. Jan 2017;144(Pt B):255-258. doi:10.1016/j.neuroimage.2015.03.069
25. Konijnenberg E, Carter SF, Ten Kate M, et al. The EMIF-AD PreclinAD study: study design and baseline cohort overview. Alzheimers Res Ther. Aug 4 2018;10(1):75. doi:10.1186/s13195-018-0406-7
26. Legdeur N, Badissi M, Carter SF, et al. Resilience to cognitive impairment in the oldest-old: design of the EMIF-AD 90+ study. BMC Geriatr. Nov 26 2018;18(1):289. doi:10.1186/s12877-018-0984-z
27. Clark CM, Schneider JA, Bedell BJ, et al. Use of florbetapir-PET for imaging beta-amyloid pathology. JAMA. Jan 19 2011;305(3):275-83. doi:10.1001/jama.2010.2008
28. Mormino EC, Betensky RA, Hedden T, et al. Amyloid and APOE epsilon4 interact to influence short-term decline in preclinical Alzheimer disease. Neurology. May 20 2014;82(20):1760-7. doi:10.1212/WNL.0000000000000431
29. Besser L, Kukull W, Knopman DS, et al. Version 3 of the National Alzheimer's Coordinating Center's Uniform Data Set. Alzheimer Dis Assoc Dis. Oct-Dec 2018;32(4):351-358. doi:10.1097/WAD.0000000000000279
30. Tijms BM, Willemse EAJ, Zwan MD, et al. Unbiased approach to counteract upward drift in cerebrospinal fluid amyloid-β 1–42 analysis results. Clin Chem. Mar 2018;64(3):576-585. doi:10.1373/clinchem.2017.281055
31. Jutten RJ, Sikkes SAM, Amariglio RE, et al. Identifying sensitive measures of cognitive decline at different clinical stages of Alzheimer's disease. JINS. 2020:1-13. doi:10.1017/S1355617720000934

32. Pfeffer RI, Kurosaki TT, Harrah CH, Jr., Chance JM, Filos S. Measurement of functional activities in older adults in the community. Article. J Gerontol. May 1982;37(3):323-9.
33. Farias ST, Mungas D, Harvey DJ, Simmons A, Reed BR, Decarli C. The measurement of Everyday Cognition (ECog): Development and validation of a short form. Alzheimers Dement. Nov 2011;7(6):593-601. doi:10.1016/j.jalz.2011.02.007
34. Farias ST, Park LQ, Harvey DJ, et al. Everyday cognition in older adults: associations with neuropsychological performance and structural brain imaging. J Int Neuropsychol Soc. Apr 2013;19(4):430-41. doi:10.1017/S1355617712001609
35. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. Alzheimers Dement. Nov 2012;8(6):536-43. doi:10.1016/j.jalz.2011.08.006
36. Dubbelman MA, Jutten RJ, Tomaszewski Farias SE, et al. Decline in cognitively complex everyday activities accelerates along the Alzheimer's disease continuum. Alzheimers Res Ther. 2020;doi:10.1186/s13195-020-00706-2
37. Jutten RJ, Peeters CFW, Leijdesdorff SMJ, et al. Detecting functional decline from normal aging to dementia: Development and validation of a short version of the Amsterdam IADL Questionnaire. Alzheimers Dement (Amst). 2017;8:26-35. doi:10.1016/j.dadm.2017.03.002
38. Koster N, Knol DL, Uitdehaag BM, Scheltens P, Sikkes SAM. The sensitivity to change over time of the Amsterdam IADL Questionnaire((c)). Alzheimers Dement. Oct 2015;11(10):1231-40. doi:10.1016/j.jalz.2014.10.006
39. Sikkes SAM, Knol DL, Pijnenburg YAL, 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(1):35-41. doi:10.1159/000346277
40. Sikkes SAM, Pijnenburg YAL, 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. Dec 2013;26(4):244-50. doi:10.1177/0891988713509139
41. Bates D, Machler M, Bolker BM, Walker SC. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. Oct 2015;67(1):1-48. doi:10.18637/jss.v067.i01
42. R: A language and environment for statistical computing. Version 4.0.3. 2020.
43. Stata Statistical Software: Release 14. Version 14. StataCorp LP.; 2015.
44. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. 5th ed. American Psychiatric Association; 2013.
45. Scheltens P, Blennow K, Breteler MMB, et al. Alzheimer's disease. The Lancet. 2016;388(10043):505-517. doi:10.1016/S0140-6736(15)01124-1
46. 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. Oct 1 2012;2(5):483-491. doi:10.2217/nmt.12.55
47. Sikkes SAM, Visser PJ, Knol DL, et al. Do instrumental activities of daily living predict dementia at 1- and 2-year follow-up? Findings from the Development of Screening guidelines and diagnostic Criteria for Predementia Alzheimer's disease study. J Am Geriatr Soc. Dec 2011;59(12):2273-81. doi:10.1111/j.1532-5415.2011.03732.x
48. Farias ST, Chou E, Harvey DJ, et al. Longitudinal trajectories of everyday function by diagnostic status. Psychol Aging. Dec 2013;28(4):1070-5. doi:10.1037/a0034069
49. Weintraub S, Carrillo MC, Farias ST, et al. Measuring cognition and function in the preclinical stage of Alzheimer's disease. Alzheimers Dement (N Y). 2018;4:64-75. doi:10.1016/j.trci.2018.01.003

50. Snyder PJ, Kahle-Wrobleski K, Brannan S, et al. Assessing cognition and function in Alzheimer's disease clinical trials: do we have the right tools? Alzheimers Dement. Nov 2014;10(6):853-60. doi:10.1016/j.jalz.2014.07.158

51. Mortamais M, Ash JA, Harrison J, et al. Detecting cognitive changes in preclinical Alzheimer's disease: A review of its feasibility. Alzheimers Dement. Apr 2017;13(4):468-492. doi:10.1016/j.jalz.2016.06.2365

52. Sikkes SAM, De Rotrou J. A qualitative review of instrumental activities of daily living in dementia: what's cooking? Neurodegener Dis Manag. 2014;4(5):393-400. doi:10.2217/nmt.14.24

53. Mura T, Proust-Lima C, Jacqmin-Gadda H, et al. Measuring cognitive change in subjects with prodromal Alzheimer's disease. J Neurol Neurosurg Psychiatry. Apr 2014;85(4):363-370. doi:10.1136/jnnp-2013-305078

54. Vermunt L, Sikkes SAM, van den Hout A, et al. Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease in relation to age, sex, and APOE genotype. Alzheimers Dement. Jul 2019;15(7):888-898. doi:10.1016/j.jalz.2019.04.001

55. Jack CR, Jr., Therneau TM, Weigand SD, et al. Prevalence of Biologically vs Clinically Defined Alzheimer Spectrum Entities Using the National Institute on Aging-Alzheimer's Association Research Framework. JAMA Neurol. Jul 15 2019;76(10):1174-1183. doi:10.1001/jamaneurol.2019.1971

6.5 Supplementary Material

Cohort descriptions

In addition to the sample characteristic table included in the main text, we provide descriptive statistics of each of the included cohorts in Supplemental Table 1 below.

Supplemental Table 1 | Cohort sample descriptives.

	HABS	ADNI	NACC	ADC	EMIF-AD	EMIF-90+
N	259	829	201	178	73	15
Age	73.64 ± 6.1	74.20 ± 7.1	69.69 ± 7.9	66.12 ± 6.8	67.57 ± 6.0	92.08 ± 1.6
Female, n (%)	156 (60)	390 (47)	90 (45)	73 (41)	54 (74)	—
Education years	15.93 ± 3.0	16.21 ± 2.7	16.35 ± 2.8	11.22 ± 2.9	11.70 ± 2.6	13.60 ± 4.9
MMSE	29.03 ± 1.1	27.57 ± 2.8	15.57 ± 3.6	22.09 ± 4.8	29.05 ± 1.2	28.73 ± 1.4
White race, n (%)	215 (82)	760 (92)	186 (93)	—	—	—

Data are presented as mean ± standard deviation, unless otherwise indicated. "—" denotes that data were unavailable. MMSE scores were taken from the baseline assessment.
Abbreviations: ADC, Amsterdam Dementia Cohort; ADNI, Alzheimer's Disease Neuroimaging Initiative; EMIF, European Medical Information Framework; HABS, Harvard Aging Brain Study; MMSE, Mini-Mental State Examination; NACC, National Alzheimer's Coordinating Center.

Sensitivity analyses

We ran sensitivity analyses to investigate whether our findings were independent of study cohort and type of recruitment setting. Models in which we excluded one of the study cohorts are shown in Supplemental Table 2. In general, the interpretation of the effects remained the same across all models. The estimates without ADNI are somewhat different, likely because ADNI was the largest sample in our study and comprised more than half the sample size. Without ADNI, the change in IADL functioning in stage 2 was not different from the change in IADL functioning in amyloid negative individuals.

Supplemental Table 2 | Sensitivity analyses excluding each cohort once.

Slopes	Without HABS	Without ADNI	Without NACC	Without ADC/EMIF
Amyloid negative	-0.05 [-0.16, 0.07]	-0.11 [-0.20, -0.01]	-0.05 [-0.12, 0.01]	-0.05 [-0.12, 0.02]
Amyloid positive	-1.09 [-1.18, -1.00]	-0.87 [-1.00, -0.75]	-0.91 [-0.98, -0.84]	-1.00 [-1.08, -0.92]
Stage 1	-0.16 [-0.40, 0.09]	-0.14 [-0.35, 0.08]	-0.12 [-0.27, 0.03]	-0.14 [-0.30, 0.02]
Stage 2	-0.35 [-0.54, -0.16]	-0.21 [-0.43, 0.00]	-0.34 [-0.47, -0.21]	-0.31 [-0.50, -0.16]
Stage 3	-1.06 [-1.16, -0.95]	-0.83 [-1.04, -0.61]	-1.01 [-1.09, -0.93]	-1.08 [-1.17, -0.98]
Stage 4+	-1.95 [-2.15, -1.75]	-2.01 [-2.29, -1.74]	-1.65 [-1.84, -1.47]	-2.04 [-2.24, -1.83]

Shown here are unstandardized betas [95% confidence intervals], adjusted for baseline age, gender, and years of education. The betas represent yearly change in the combined Z-score of the three IADL instruments.

Stratified models for community-based and memory clinic cohorts are in Supplemental Table 3. In the community-based cohorts, the rate of decline in IADL functioning is very similar between stages 1 and 2, whereas the rate of decline is faster in stage 2 than in stage 1 in the memory clinic cohorts. The effects in the other stages are similar between community-based and memory clinic studies.

Supplemental Table 3 | Sensitivity analyses stratified for community-based vs. memory clinic study cohorts.

Slopes	Community-based	Memory clinic
Amyloid negative	-0.07 [-0.19, 0.04]	-0.01 [-0.11, 0.09]
Amyloid positive		
Stage 1	-0.14 [-0.38, 0.10]	-0.14 [-0.37, 0.09]
Stage 2	-0.13 [-0.41, 0.15]	-0.40* [-0.58, -0.21]
Stage 3	-1.21* [-1.60, -0.81]	-1.02* [-1.12, -0.93]
Stage 4+	-2.55* [-3.01, -2.08]	-1.67* [-1.87, -1.47]

Shown here are unstandardized betas [95% confidence intervals], adjusted for baseline age, gender, and years of education. The betas represent yearly change in the combined Z-score of the three IADL instruments.

* different from amyloid negative controls

Finally, we ran sensitivity models using each of the three instruments separately, which can be found in Supplemental Table 4. The FAQ and ECog show similar effects in all stages, with increasing rates of decline in each subsequent stage, and a significant difference in rate of decline between stage 2 and amyloid negatives. The A-IADL-Q shows similar effects as well, however, these are not significant, possibly due to the smaller sample sizes. The estimate of the slope in stage 1 is based on just 4 individuals and should hence be interpreted with caution.

Supplemental Table 4 | Sensitivity analyses for each separate IADL instrument.

Slopes	FAQ	ECog	A-IADL-Q
Amyloid negative	-0.01 [-0.16, 0.14]	-0.04 [-0.08, 0.01]	-0.30 [-0.57, -0.04]
Amyloid positive			
Stage 1	-0.16 [-0.47, 0.15]	-0.11 [-0.21, -0.02]	1.04[a] [-0.48, 2.57]
Stage 2	-0.43* [-0.70, -0.16]	-0.25* [-0.34, -0.16]	-0.41 [-0.80, -0.02]
Stage 3	-1.46* [-1.60, -1.31]	-0.53* [-0.59, -0.48]	-0.67 [-0.94, -0.40]
Stage 4+	-2.62* [-2.92, -2.31]	-0.83* [-1.00, -0.67]	-1.56* [-1.89, -1.22]

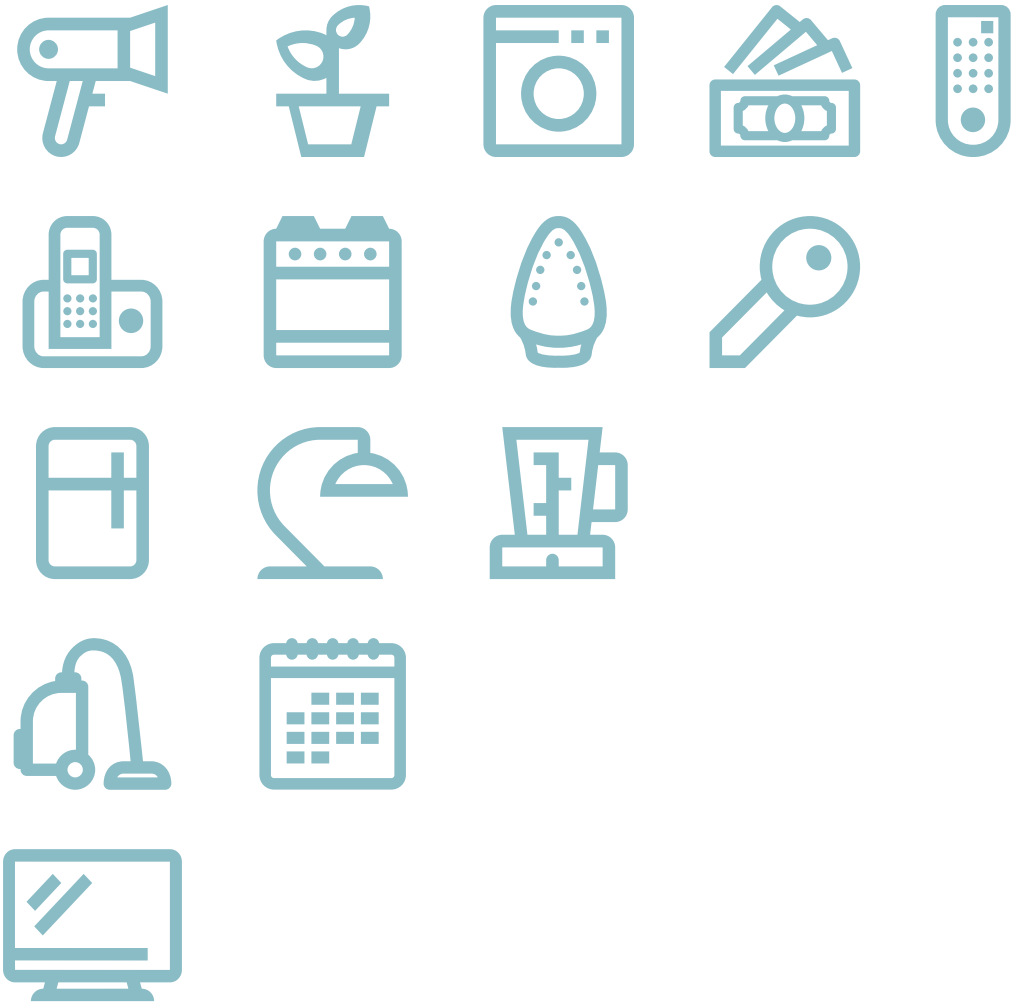
Shown here are unstandardized betas [95% confidence intervals], adjusted for baseline age, gender, and years of education. The betas represent yearly change in the combined Z-score of each IADL instrument.

* different from amyloid negative controls, [a] based on $n = 4$

Abbreviations: A-IADL-Q, Amsterdam Instrumental Activities of Daily Living Questionnaire; ECog, Everyday Cognition; FAQ, Functional Activities Questionnaire.

Amyloid classification procedures

PET data were available for ADNI, HABS and both EMIF-AD studies, while both CSF and PET were used in NACC and ADC. Amyloid binding was measured in HABS using ^{11}C-Pittsburgh compound-B (^{11}C-PiB), ^{18}F-florbetapir in ADNI, and one of ^{11}C-PiB, ^{18}F-flutemetamol, ^{18}F-florbetapir, or ^{18}F-florbetaben in the ADC and EMIF-AD studies. Amyloid positivity was determined in HABS on distribution volume ratio of mean uptake in frontal, lateral parietal and temporal, and retrosplenial regions (cutoff ≥ 1.20). In ADNI, amyloid positivity was based on standard uptake value ratios of mean uptake in four cortical regions, normalized to the whole cerebellum uptake (cutoff value ≥ 1.10). In both EMIF-AD studies, amyloid positivity was determined by consensus on visual read of PET scans by multiple independent physicians. In the ADC, PET scans were visually rated by an experienced nuclear medicine physician who was blinded to clinical information. In the NACC cohort, amyloid positivity in PET or CSF was determined using each center's local cutoffs. In the ADC, CSF amyloid positivity was determined as being below the cutoff of 813 pg/mL. Where CSF and PET were available for the same individual, PET results were favored in case of disagreement.



CHAPTER 7

IDENTIFYING AND CHARACTERIZING HETEROGENEITY IN EVERYDAY FUNCTIONING PROGRESSION AMONG MEMORY CLINIC PATIENTS

Mark A. Dubbelman, Michael Caelers, Roos J. Jutten, Afina W. Lemstra, Charlotte E. Teunissen, Bart N.M. van Berckel, Frederik Barkhof, Wiesje M. van der Flier, Philip Scheltens, Sietske A.M. Sikkes

In preparation

Abstract

Introduction: The rate of decline in the performance of cognitively complex 'instrumental activities of daily living' (IADL) along the clinical spectrum of Alzheimer's disease and related disorders differs greatly between individuals. To optimize the accuracy of prognosis for decline in everyday functioning, we aimed to characterize the heterogeneity in trajectories of functional change over time among memory clinic patients.

Methods: We included 948 memory clinic patients (64.4 ± 7.5 years; 40% female) from the Amsterdam Dementia Cohort who were followed over time (mean follow-up 2.3 ± 1.7 years). Everyday functioning was measured using the Amsterdam IADL Questionnaire (A-IADL-Q). Longitudinal latent class analysis identified clusters of patients showing similar rates of change in A-IADL-Q scores over time. We then investigated group differences in clinical information (including demographics, family history, psychosocial factors, cognitive performance, brain imaging and Alzheimer's disease biomarkers) from the baseline visit using multinomial logistic least absolute shrinkage models capable of handling a large number of predictors.

Results: Seven clusters were identified that could be grouped and classified as stable (*n* = 283; 30%), slow progression (*n* = 226; 24%) and fast progression (*n* = 439; 46%). Patients who showed fast progression were older (odds ratio (OR)=1.01), were more often left-handed (OR=1.12), had more medial temporal (OR=1.04) and global cortical atrophy (OR=1.29), and were more often diagnosed with Alzheimer's disease dementia (OR=1.52) or dementia with Lewy bodies (OR=1.93) than patients who remained stable or showed slow progression.

Discussion: Our findings may help in forming a more accurate prognosis of future functional decline by signaling characteristics of patients presenting at a memory clinic who are more likely to experience a rapid decline in daily functioning.

7.1 Introduction

Alzheimer's disease and related disorders are characterized by cognitive and functional decline culminating in impairment of everyday functioning and dependence on others.[1] With improvements in the early detection of disease biomarkers, especially for Alzheimer's disease, the focus in both clinical and research settings has shifted to earlier disease stages, when clinical symptoms are still absent or minimal. This shift implicates a substantial extension of the disease duration, thus necessitating more information about disease progression for adequate prognosis.

Performance of higher-order cognitive everyday activities, so-called 'instrumental activities of daily living' (IADL),[2] appears to worsen over the years preceding dementia. This has been well-established in Alzheimer's disease,[3-7] and the performance of IADL has also been investigated in related disorders such as frontotemporal dementia,[8,9] vascular dementia,[10,11] and dementia with Lewy bodies.[11,12] The individual rates of change within diagnoses and disease stages, however, are highly variable.[3,13] This suggests that the heterogeneity of changes in everyday functioning is related to other characteristics besides diagnosis. Consequently, there is still a challenge in improving prognostic accuracy by identifying other clinical characteristics that may be associated with an accelerated progression in functional impairment. This is relevant because daily functioning plays an important role in the ability to function independently in society.

Thus, we aimed to investigate whether it is possible to detect groups, and if so, how many, of patients showing similar changes over time in everyday functioning. We then aimed to identify baseline characteristics that were associated with those groups in a cohort of unselected memory clinic patients. Everyday functioning was assessed longitudinally using the Amsterdam Instrumental Activities of Daily Living Questionnaire,[14] which has previously been shown to be sensitive to changes over time.[15,16]

7.2 Methods

Study sample

Patients who visited the Alzheimer Center Amsterdam between May 2013 and August 2021 and who had at least a baseline and one follow-up assessment of the

Amsterdam IADL Questionnaire were included in this longitudinal study. We did not select for diagnosis or other patient characteristics. All patients were included in the Amsterdam Dementia Cohort.[17]

At their initial visit, all patients underwent extensive dementia screening, including neurological and neuropsychological assessment, magnetic resonance imaging (MRI) and electroencephalography (EEG), and a lumbar puncture.[17] Diagnoses were subsequently made in a multidisciplinary consensus meeting, based on diagnostic criteria for Alzheimer's disease,[18] frontotemporal dementia,[19] dementia with Lewy bodies,[20,21] vascular dementia,[22] mild cognitive impairment,[23] and subjective cognitive decline.[24]

Measures

Everyday functioning: the Amsterdam Instrumental Activities of Daily Living Questionnaire

The Amsterdam Instrumental Activities of Daily Living Questionnaire (A-IADL-Q) measures impairment in the performance of 30 cognitively complex everyday activities due to cognitive impairment.[14,25] The questionnaire is completed by an informal caregiver of the patient and has been extensively validated.[16,26,27] Total scores are calculated using item response theory, and represent the construct of everyday functioning. Scores are normally distributed in the memory clinic around a mean of 50 and with a standard deviation of 10, thus ranging from approximately 20 to 80. Higher total scores represent better functioning.

Neuropsychological assessment

All patients received standardized neuropsychological assessment.[17] Tasks focusing on the assessment of episodic memory included the Dutch auditory verbal learning test (AVLT),[28] the Visual Association Test (VAT), in which series of unusual combinations of drawings (e.g., a gorilla holding an umbrella) need to be memorized,[29] and the delayed recall of the Rey Complex Figure Test (RCFT).[30] For executive functioning, we used part B of the Trail Making Test (TMT),[31] the backward condition of the digit span,[32] card III the Stroop Color-Word Test,[33] and letter fluency.[34] To measure attention and processing speed, we used the forward condition of the digit span, part A of the TMT, card I of the Stroop Color-Word Test, and the Letter Digit Substitution Test (LDST).[35] Animal fluency and the drawing

naming portion of the VAT were used to assess language, and the copy of the RCFT as well as the number location and fragmented letters of the Visual Object and Space Perception (VOSP) battery[36] were used to measure visuospatial functions.

We calculated Z-scores per cognitive domain using the baseline mean and standard deviations in the group of patients without objective cognitive impairment. We confirmed whether tests included in each domain score loaded on the same factor and found adequate (language, visuospatial functioning) to good fit (episodic memory, executive functioning, attention/speed).

Brain imaging

MRI was performed on different scanners (1.5 or 3 Tesla, Philips Medical Systems, Best, The Netherlands). Experienced neuroradiologists visually assessed atrophy and cerebrovascular abnormalities.[17] They rated medial temporal lobe atrophy using coronal T1-weighted images on a five-point scale (0–4),[37] global and posterior cortical atrophy on fluid attenuated inversion recovery (FLAIR) images using four-point scales (0–3),[38,39] and white matter hyperintensities on a four-point scale using FLAIR images (0–3).[40] Finally, they counted the number of lacunar infarctions.

AD biomarkers

As part of the standard clinical work-up, patients underwent lumbar puncture. Cerebrospinal fluid (CSF) amyloid-beta$_{1-42}$ (Aβ_{1-42}), tau, and p-tau were then analyzed using Innotest (enzyme-linked immunosorbent assay (ELISA); Innotest, Fujirebio, Ghent, Belgium) or Elecsys (Roche Diagnostics, GmbH, Mannheim, Germany). Elecsys values were transformed to Innotest results using previously established formulas.[41] CSF Aβ_{1-42} positivity was determined as being below the cutoff of 813 pg./mL.[42] A subset of patients (*n* = 276; 29%) also had an amyloid positron emission tomography (PET) scan, using one of ^{11}C-Pittsburgh compound-B (N), ^{18}F-flutemetamol (N), ^{18}F-florbetapir (N), or ^{18}F-florbetaben (N) as a tracer. PET scans were visually rated by an experienced nuclear medicine physician (BNMvB) who was blinded to clinical information. CSF and PET amyloid status were treated separately.

Other measures

Apolipoprotein (APOE)-ε4 genotyping was performed using the LightCycler ApoE mutation Detection Kit (Roche Diagnostics, GmbH, Mannheim, Germany) after

DNA isolation from 10 mL ethylenediamine tetraacetic acid (EDTA) vacutainer tubes. Patients were classified as APOE-ε4 carrier (heterozygous or homozygous) or noncarrier.

Statistical analyses

Analyses were performed in R version 4.1.3.[43] We first estimated latent class linear mixed models (LCLMMs) using the 'lcmm' package version 1.9.3,[44] with A-IADL-Q total scores as the dependent and time in years as the independent variable. Additionally, to investigate non-linear patterns of change, we included time in years squared. No other clinical data was used in the LCLMMs: the latent classes thus represent groups showing similar change in IADL over time. We selected the model with the optimal model fit, as determined by the lowest Akaike information criterion from models fitted with one through ten latent classes. Group differences were subsequently tested using *t*-tests or chi squared-tests, as appropriate, with post-hoc correction for multiple comparisons using Tukey's Honest Significant Difference test.

Change in A-IADL-Q scores over time was subsequently modeled in linear mixed models with a random intercept and random effect for time. A time by cluster interaction was used as the main independent variable of interest. Estimated marginal means were used to test group differences. We used the 'lme4' package version 1.1-28[45] and 'emmeans' package version 1.7.3 for the linear mixed models and estimated marginal means, respectively.

Subsequently, we identified factors that were associated with cluster membership. These factors included all sociodemographic, AD biomarker, brain imaging, neuropsychological and genetic data obtained at the baseline visit. To compensate for missing clinical data, we applied multiple imputations using the 'mice' package version 3.14.0[46] and created five imputed datasets. We then performed multinomial logistic least absolute shrinkage and selection operator ('lasso') regressions using the 'glmnet' package version 4.1-3.[47] First, we performed a tenfold cross-validation in each of the five imputed datasets to obtain the optimal lambda, a value that determines when predictors should be removed from the model. Then, using that lambda, we ran the lasso regressions. Lasso regressions will identify the variables that contribute to the prediction and will set the estimate of all non-influential variables to zero. We used the results from the lasso regression to predict to what

cluster each individual belonged and computed the area under the receiver operating curve. Last, we averaged the odds ratios of the factors that were not null from the five lasso regressions, provided the factor was not null in at least three out of the five lasso regressions.

7.3 Results

A total of 948 patients (mean age 64.4 ± 7.5 years, 39.6% female, follow-up duration 2.3 ± 1.7 years) were included.

Identifying the clusters

The best LCLMM identified seven latent classes ('clusters'), with a quadratic time term, modeling non-linear change over time. The clusters can be characterized as follows, based on the group starting point of IADL performance and trajectory over time on the A-IADL-Q: (1) high starting point/improvement over time (*n* = 213, 22.5% of sample), (2) average starting point/no change over time (*n* = 70, 7.4%), (3) average starting point/decline over time (*n* = 226, 23.8%), (4) average starting point/steep decline over time (*n* = 96, 10.1%), (5) below average starting point/steep decline over time (*n* = 244, 25.7%), (6) low starting point/decline over time (*n* = 80, 8.4%), and (7) low starting point/very steep decline over time (*n* = 19, 2.0%). Table 1 lists the main characteristics of the overall sample and all seven clusters. Clusters differed in age, education, baseline A-IADL-Q and MMSE scores and diagnosis distributions (all $p < .001$), but not in sex distribution. Post-hoc analysis with Tukey's honest significant difference corrections showed that patients in clusters 1, 2 and 7 were younger than those in the other clusters. Patients in clusters 1, 2 and 3 were most highly educated, followed by patients in clusters 4, 5 and 6. Patients in cluster 7 had the fewest years of education. Patients in cluster 1 had the highest baseline A-IADL-Q scores, followed by those in cluster 2, clusters 3 and 4, cluster 5, cluster 6, and cluster 7. Similarly, patients in clusters 1 and 2 had the highest MMSE scores, followed by patients in clusters 3 and 4, cluster 5, and clusters 6 and 7.

Table 1 | Baseline characteristics

	Whole Sample	**Cluster 1** 'High/ Improve'	**Cluster 2** 'Average/ No change'	**Cluster 3** 'Average/ Slow decline'	**Cluster 4** 'Average/ Decline'	**Cluster 5** 'Below Average/ Decline'	**Cluster 6** 'Low/ Decline'	**Cluster 7** 'Low/Steep decline'
n (%)[a]	948 (100.0)	213 (22.5)	70 (7.4)	226 (23.8)	96 (10.1)	244 (25.7)	80 (8.4)	19 (2.0)
Age	64.4 ± 7.5	62.4 ± 7.3	62.6 ± 7.1	64.9 ± 7.0	65.9 ± 7.1	65.1 ± 8.0	65.9 ± 8.2	63.3 ± 7.5
Female, n (%)	375 (39.6)	93 (43.7)	20 (28.6)	79 (35.0)	35 (36.5)	103 (42.2)	35 (43.8)	10 (52.6)
Education in years, m (iqr)	9 (8–10)	10 (9–10)	9 (9–10)	9 (9–10)	9 (8.75–10)	9 (8–9)	8 (8–9)	8 (6–8)
Baseline A-IADL-Q	51.9 ± 10.1	62.9 ± 6.4	56.4 ± 5.9	52.7 ± 6.1	51.9 ± 7.2	46.5 ± 5.9	35.4 ± 6.4	40.9 ± 10.5
Baseline MMSE	24.5 ± 4.6	27.6 ± 2.6	27.3 ± 2.0	25.7 ± 3.1	25.0 ± 2.8	22.2 ± 4.5	19.2 ± 5.6	16.8 ± 6.5
Diagnosis, n (%)								
SCD	238 (25.1)	136 (63.8)	34 (48.6)	49 (21.7)	11 (11.5)	8 (3.3)	0 (0.0)	0 (0.0)
MCI	162 (17.1)	33 (15.5)	22 (31.4)	70 (31.0)	15 (15.6)	19 (7.8)	3 (3.8)	0 (0.0)
AD	333 (35.1)	12 (5.6)	8 (11.4)	62 (27.4)	43 (44.8)	151 (61.9)	48 (60.0)	9 (47.4)
FTD	26 (2.7)	2 (0.9)	0 (0.0)	4 (1.8)	4 (4.2)	9 (3.7)	4 (5.0)	3 (15.8)
DLB	72 (7.6)	3 (1.4)	1 (1.4)	8 (3.5)	12 (12.5)	31 (12.7)	14 (17.5)	3 (15.8)
Other dementia	47 (5.0)	9 (4.2)	1 (1.4)	15 (6.6)	5 (5.2)	7 (2.9)	7 (8.8)	3 (15.8)
Other diagnosis	70 (7.4)	18 (8.5)	4 (5.7)	18 (8.0)	6 (6.2)	19 (7.8)	4 (5.0)	1 (5.3)

Data displayed as mean ± standard deviation, except as noted otherwise. [a] Percentage represents row percentage.

Abbreviations: AD, Alzheimer's disease dementia; A-IADL-Q, Amsterdam Instrumental Activities of Daily Living Questionnaire; DLB, dementia with Lewy bodies; FTD, frontotemporal dementia; IQR, interquartile range; M, median; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; SCD, subjective cognitive decline.

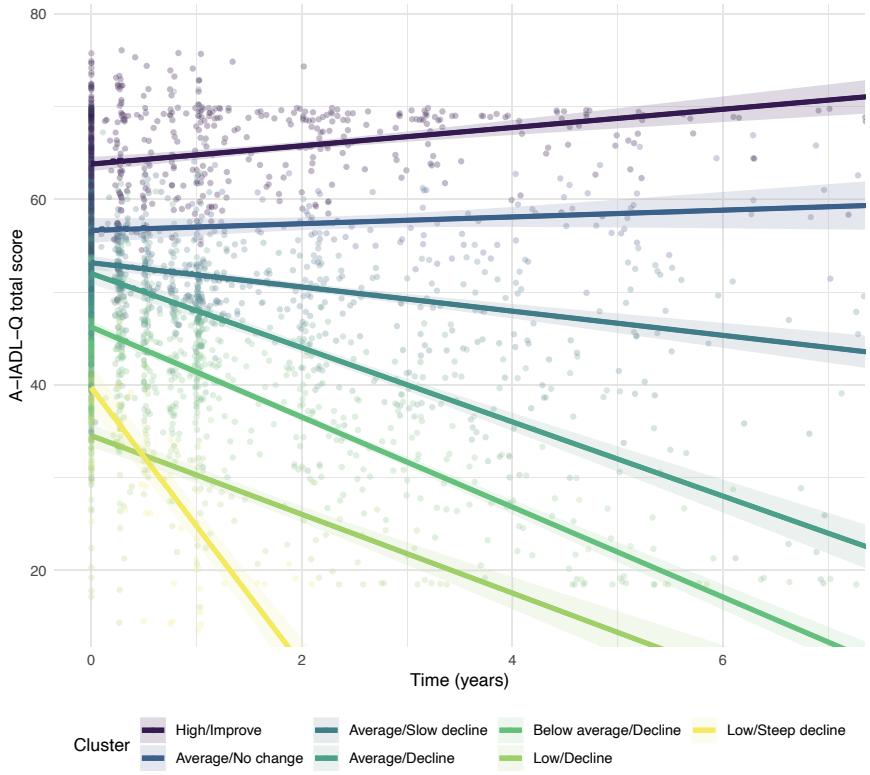


Figure 1 | Change over time in everyday functioning, per cluster.
Figure 1 shows the individual change over time of all patients, including the average slopes per cluster. All clusters except cluster 2 ('average/no change') showed a significant change over time in everyday functioning (cluster 2 p = .153, all other p < .001). There were differences in the rates of change between most clusters, except clusters 1 and 2, and cluster 4, cluster 5 and cluster 6. Table 2 lists the time trends per cluster, as well as between-cluster differences, as tested using estimated marginal means.

Table 2 | Time trends showing yearly change in A-IADL-Q scores per cluster.

Cluster	Time trend [95%CI]	P-value	Different from
1 'High/Improve'	0.99 [0.66, 1.31]	< .001	Clusters 3, 4, 5, 6, 7
2 'Average/No change'	0.37 [-0.13, 0.86]	.153	Clusters 3, 4, 5, 6, 7
3 'Average/Slow decline'	-1.30 [-1.61, -0.99]	< .001	All other clusters
4 'Average/Decline'	-4.00 [-4.44, -3.56]	< .001	Clusters 1, 2, 3, 7
5 'Below average/Decline'	-4.85 [-5.16, -4.54]	< .001	Clusters 1, 2, 3, 7
6 'Low/Decline'	-4.24 [-4.85, -3.63]	< .001	Clusters 1, 2, 3, 7
7 'Low/Steep decline'	-14.94 [-16.89, -13.00]	< .001	All other clusters

Abbreviations: A-IADL-Q, Amsterdam Instrumental Activities of Daily Living Questionnaire; 95%CI, 95% confidence interval.

Characterizing the clusters

Due to a lack of clinical data, 12 patients (1.3%) were excluded from these analyses. Based on the differences in rates of change in A-IADL-Q scores, we combined the identified clusters into a 'Stable' group (clusters 1 and 2), a 'Slow progression' group (cluster 3) and a 'Fast progression' group (clusters 4, 5, 6 and 7). Using baseline clinical data, patients could be accurately classified into these groups with an average area under the curve (AUC) of 0.826.

Patients diagnosed with subjective cognitive decline (odds ratio (OR) = 1.91) and who had a higher MMSE (OR = 1.10), better episodic memory (OR = 1.06) and visuospatial functioning (OR = 1.16), a higher BMI (OR = 1.02) and who consumed more weekly units of alcohol (OR = 1.07) were more likely to be in the Stable group, compared to the two other groups. Conversely, patients who were older, were diagnosed with AD dementia, had more medial temporal lobe atrophy and more lacunar infarctions and who had positive biomarkers for AD as established in CSF or by PET, were less likely to be stable. Patients with higher GDS scores, but also with better executive functioning and FAB scores, were more likely to be in the Slow progression group, compared to the other groups. Those who were left-handed were less likely to be slow progressors, as were those with a family history of psychiatric disorders. Finally, patients who were diagnosed with DLB or AD dementia, were older, had higher CDR and ZBI scores, and more medial temporal and global cortical atrophy were more likely to be in the Fast progression group. Those who performed worse on neuropsychological tests assessing executive functioning, attention and speed, language and visuospatial functioning were also more likely to be fast progressors. Other patient characteristics assessed at baseline (patient sex, education level, marital status, whether they had children, whether they lived at home or in an assisted living facility, whether they smoked and APOE ε4 carriership) were not associated with group membership. Descriptive information per group of all investigated characteristics is included in the Supplementary Material.

Table 3 | Odds ratios for three groups from lasso regressions

	Stable	Slow progression	Fast progression
N clinical data (%)	282 (99.6)	221 (97.8)	443 (98.6)
Age in years	0.98	—	1.01
Alcohol use, weekly units	1.07	—	—
Years since onset of complaints	1.01	0.98	—
Family history			
Dementia	1.02	—	—
Psychiatry	—	0.97	—
Cardiovascular disease	—	—	—
Left-handed	—	0.75	1.12
BMi	1.02	—	—
Diagnosis			
SCD	1.92	—	0.80
MCI	—	—	0.65
AD dementia	0.87	—	1.52
FTD	—	—	—
DLB	—	—	1.93
Other dementia	—	—	—
Psychiatry	—	—	—
Screeners			1
MoCA	0.97	—	.02
MMSE	1.10	—	0.90
FAB	—	1.02	—
GDS	—	1.02	1.00
CDR	0.58	—	1.78
Zarit Caregiver Burden	0.95	—	1.01
Neuropsychological assessment			
Episodic memory	1.06	—	—
Executive functioning	—	1.04	0.98
Attention/speed	—	—	0.83
Language	—	—	0.99
Visuospatial functions	1.16	—	0.96
MR			
Medial temporal lobe atrophy	0.85	—	1.04
Parietal atrophy	—	—	—
Global cortical atrophy	—	—	1.29
White matter lesions	—	—	—
Lacunar infarctions	0.91	—	—
Amyloid			
PET positive	0.98	—	—
CSF positive	0.84	—	—

Odds ratios displayed only for predictors that were not null in the lasso regressions.

Abbreviations: AD, Alzheimer's dementia; BMI, body mass index; CDR, Clinical Dementia Rating; CSF, cerebrospinal fluid; DLB, dementia with Lewy bodies; FAB, Frontal Assessment Battery; FTD, frontotemporal dementia; GDS, Geriatric Depression Scale; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; MR, magnetic resonance; PET, positron emission tomography; SCD, subjective cognitive decline.

7.4 Discussion

In this study, we analyzed the longitudinal data on everyday functioning of a cohort of consecutive memory clinic patients with the goal of identifying groups of patients who cluster together in their rates of change over time in everyday functioning. A longitudinal latent class analysis identified seven clusters of patients, from which we created three groups with unique patterns of functional change over time (stable, slow progression and fast progression). We subsequently found that several baseline characteristics were associated with group membership, including diagnosis, cognitive test performance and measures of brain atrophy.

We aimed to characterize the heterogeneity in the progression of difficulties in everyday functioning among memory clinic patients. It has been established that disease progression of Alzheimer's disease is vastly different between individuals.[13,48-50] In the present study, we discerned seven clusters of patients who showed similar changes in everyday functioning over time, based on latent class analysis that included only the A-IADL-Q scores over time, not only in Alzheimer's disease, but in all patients presenting at a memory clinic. The clustering was not dictated by other clinical information, nor by a priori expectations about the number of clusters. This approach allowed us to uncover groups of patients who specifically share similar trajectories of functional decline, regardless of other clinical information. Intuitively, patients in an earlier disease stage should show less severe progression. However, there seems to be more to change in everyday functioning than just diagnosis.

Some of the clusters uniquely identified by the latent class analysis appeared to have similar trajectories of functional change, as the rates of change did not differ significantly from each other. We grouped those clusters together, so that two clusters that had an above average baseline A-IADL-Q score and that showed improvement or no change over time were grouped together in a 'Stable' group. One cluster differed significantly from all others and formed a 'Slow progression' group. Four clusters that had a low baseline level of everyday functioning and showed faster decline over time were grouped together in a 'Fast progression' group. While one cluster showed a markedly steeper decline over time compared to all others, it was still included in the 'Fast progression' group rather than forming a separate group, due to the small size of the cluster. Previous studies also identified three clusters of patients showing comparable rates of clinical progression, similarly characterized, in different samples of patients with Alzheimer's disease.[49,51]

Patients from our study who were in the Fast progression group were older, more often left-handed and were more often diagnosed with Alzheimer's disease dementia or dementia with Lewy bodies. Previous work has shown that the rate of decline in more advanced disease stages is accelerated, compared to earlier stages.[3] Patients who showed a steep decline in everyday functioning also had lower scores on cognitive and functional screeners, including the MMSE and CDR, as Haaksma and colleagues showed earlier.[49] They also indicated more depressive symptoms and performed worse on neuropsychological tests assessing executive functioning, attention and speed, language and visuospatial functioning. Marshall and colleagues[52] previously showed that worse executive functioning was associated with worse everyday functioning. Patients' caregivers indicated that the burden of taking care of these patients was high. We previously showed that meaningful improvement in everyday functioning was less likely when caregiver burden was high.[53] Finally, patients had more global cortical as well as more medial temporal atrophy, the latter of which we have also shown to be associated with worse everyday functioning and meaningful decline in everyday functioning.[53,54]

Patients who showed a slow progression of impairment in everyday functioning had a shorter interval between the onset of complaints and their initial memory clinic visit, less often had a family history of psychiatric disorders and were less often left-handed. They had a better performance on a screener for frontal symptoms and indicated more depressive feelings. Mood disturbances have been linked to worse clinical progression in Alzheimer's disease,[55] although it is possible that people presenting at a memory clinic with mood disturbances may not have an underlying neurodegenerative disease, thus resulting in a slower clinical progression. Finally, they had better performance on neuropsychological tests assessing executive functioning, mirroring the group of patients who showed a fast progression of functional impairment.

Patients who remained stable in their everyday functioning were younger than those who showed a slow or fast progression. Stable patients more often had a family history of dementia, possibly because people with a positive family for dementia may be hypervigilant for changes in cognitive and daily functioning, leading them to visit a memory clinic sooner than someone who does not have a genetic burden and might cast off worries about cognitive decline as being part of normal aging. Stable patients were more often diagnosed with subjective cognitive decline and less often with Alzheimer's disease dementia. While subjective cognitive decline

forms an at-risk stage for future cognitive decline,[56] not everyone shows clinical progression,[57] because not everyone has an underlying neurodegenerative disease. More years had passed since the onset of complaints before patients who showed no progression first visited the memory clinic. This may reflect a group of patients who had less severe complaints, for which they did not initially feel the need to visit a memory clinic. Patients who did not decline also had better performance on most screening instruments and performed better on tests assessing episodic memory and visuospatial functioning. Their caregivers indicated a low burden of care. They had less medial temporal atrophy and fewer lacunar infarctions, and less often had abnormal amyloid deposition. These last findings may all hint at an absence of underlying Alzheimer's disease, which might explain why there is no clinical decline.

This study had a few limitations. First, many clinical characteristics are interdependent and influence each other and as such, we cannot consider them entirely separately. Still, by using an analysis that is designed to identify the most important characteristics and consider all others as non-contributors, we were able to single out specific factors that contributed most to predicting who belongs in which group of functional decline. Second, the diagnostic workup used at our memory clinic was quite extensive, including standardized neuropsychological assessment, magnetic resonance imaging and lumbar puncture, occasionally supplemented with positron emission tomography. While this allowed us to define our patients rather well and find ample characteristics that were associated with change over time in daily functioning, other clinics may not have access to all these data, thus limiting the repeatability of our procedures. Finally, although the grouping of patients reduced heterogeneity in trajectories of functional change, it is possible that there are still individual differences within groups. Future studies could model individual change in everyday functioning and relate it to baseline characteristics. An important strength of our study was the fact that the clusters were based solely on change in everyday functioning, which allowed us to group together patients who showed similar rates of change across diagnoses and regardless of other clinical characteristics. Furthermore, we had a large sample of well-defined memory clinic patients who were followed up to seven years and used a functional outcome measure that has been extensively validated.

In conclusion, we identified groups of patients who showed a similar change over time in everyday functioning and characterized these groups based on information

available at the initial memory clinic visit. Our findings may improve prognostic accuracy by identifying those who present at a memory clinic who are at a greater risk of experiencing progression of impairments in everyday functioning.

Declarations

Acknowledgments

The Alzheimer Center Amsterdam is supported by Alzheimer Nederland and Stichting VUMC funds. Research of the Alzheimer Center Amsterdam is part of the neurodegeneration research program of Amsterdam Neuroscience. The clinical database structure was developed with funding from Stichting Dioraphte. The SCIENCe project is supported by a research grant from Gieskes-Strijbis fonds. W.M.F. holds the Pasman chair. W.M.F. received funding in the context of Deltaplan Dementie from ZonMW Memorabel (Netherlands Consortium of Dementia Cohorts [NCDC] project 73305095005) and Alzheimer Nederland. PET scans were funded by research grants from AVID and Piramal Neuroimaging.

M.A.D., M.C., and R.J.J. report no relevant disclosures. C.E.T. is supported by the European Commission (Marie Curie International Training Network, grant agreement No 860197 (MIRIADE), Innovative Medicines Initiatives 3TR (Horizon 2020, grant no 831434) and JPND (bPRIDE), National MS Society (Progressive MS alliance) and Health Holland, the Dutch Research Council (ZonMW), Alzheimer Drug Discovery Foundation, The Selfridges Group Foundation, Alzheimer Netherlands, Alzheimer Association. CT is recipient of ABOARD, which is a public-private partnership receiving funding from ZonMW (#73305095007) and Health~Holland, Topsector Life Sciences & Health (PPP-allowance; #LSHM20106). ABOARD also receives funding from Edwin Bouw Fonds and Gieskes-Strijbisfonds. C.E.T. has a collaboration contract with ADx Neurosciences, Quanterix and Eli Lilly, performed contract research or received grants from AC-Immune, Axon Neurosciences, Bioconnect, Bioorchestra, Brainstorm Therapeutics, Celgene, EIP Pharma, Eisai, Novo Nordisk, PeopleBio, Roche, Toyama, Vivoryon, serves on editorial boards of Medidact Neurologie/Springer, Alzheimer Research and Therapy, Neurology: Neuroimmunology & Neuroinflammation, and is editor of a Neuromethods book Springer. B.N.M.B. has received research support from EU-FP7, CTMM, ZonMw, NWO and Alzheimer Nederland, has performed contract research for Rodin, IONIS, AVID, Eli Lilly, UCB, DIAN-TUI and Janssen, was a speaker at a symposium organized by Springer Healthcare, has a consultancy agreement with IXICO for the reading of

PET scans, and is a trainer for GE. All funds are paid to his institution. W.M.F.'s research programs have been funded by ZonMW, the Netherlands Organization of Scientific Research, Alzheimer Nederland, Cardiovascular Onderzoek Nederland, Stichting Dioraphte, Gieskes-Strijbis fonds, Pasman stichting, Boehringer Ingelheim, Life-MI, AVID, Biogen MA, and Combinostics (all funding paid to her institution). P.S. has received consultancy fees (paid to the institution) from Alzheon, Brainstorm Cell, ImmunoBrain Checkpoint, Green Valley. Within his university affiliation he is PI of studies with AC Immune, FUJI-film/Toyama, IONIS, UCB, Novo Nordisk and Vivoryon. He is also an employee of EQT Life Sciences (formerly LSP). S.A.M.S. provided consultancy services for Boehringer Ingelheim, Takeda, and Nutrica; her work is supported by grants from the Dutch Research Council (ZonMW) and the EU Joint Programme—Neurodegenerative Disease Research (all funds paid to her institution).

References

1. Scheltens P, De Strooper B, Kivipelto M, et al. Alzheimer's disease. The Lancet. 2021;doi:10.1016/s0140-6736(20)32205-4
2. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. Autumn 1969;9(3):179-86.
3. Dubbelman MA, Jutten RJ, Tomaszewski Farias SE, et al. Decline in cognitively complex everyday activities accelerates along the Alzheimer's disease continuum. Alzheimers Res Ther. 2020;doi:10.1186/s13195-020-00706-2
4. Facal D, Carabias MAR, Pereiro AX, et al. Assessing everyday activities across the dementia spectrum with the Amsterdam IADL Questionnaire. Curr Alzheimer Res. 2018;15(13):1261-1266. doi:10.2174/1567205015666180925113411
5. Lilamand M, Cesari M, Cantet C, et al. Relationship Between Brain Amyloid Deposition and Instrumental Activities of Daily Living in Older Adults: A Longitudinal Study from the Multidomain Alzheimer Prevention Trial. J Am Geriatr Soc. Oct 2018;66(10):1940-1947. doi:10.1111/jgs.15497
6. 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. Oct 1 2012;2(5):483-491. doi:10.2217/nmt.12.55
7. Marshall GA, Gatchel JR, Donovan NJ, et al. Regional Tau Correlates of Instrumental Activities of Daily Living and Apathy in Mild Cognitive Impairment and Alzheimer's Disease Dementia. J Alzheimers Dis. 2019;67(2):757-768. doi:10.3233/JAD-170578
8. Amanzio M, D'Agata F, Palermo S, et al. Neural correlates of reduced awareness in instrumental activities of daily living in frontotemporal dementia. Exp Gerontol. Oct 2016;83:158-64. doi:10.1016/j.exger.2016.08.008
9. Moheb N, Mendez MF, Kremen SA, Teng E. Executive Dysfunction and Behavioral Symptoms Are Associated with Deficits in Instrumental Activities of Daily Living in Frontotemporal Dementia. Dement Geriatr Cogn Disord. 2017;43(1-2):89-99. doi:10.1159/000455119
10. Boyle PA, Cahn-Weiner D. Assessment and prediction of functional impairment in vascular dementia. Expert Rev Neurother. Jan 2004;4(1):109-14. doi:10.1586/14737175.4.1.109
11. Gill DP, Hubbard RA, Koepsell TD, et al. Differences in rate of functional decline across three dementia types. Alzheimers Dement. Oct 2013;9(5 Suppl):S63-71. doi:10.1016/j.jalz.2012.10.010
12. Hamilton JM, Salmon DP, Raman R, et al. Accounting for functional loss in Alzheimer's disease and dementia with Lewy bodies: beyond cognition. Alzheimers Dement. Mar 2014;10(2):171-8. doi:10.1016/j.jalz.2013.04.003
13. Haaksma ML, Rizzuto D, Leoutsakos JS, et al. Predicting Cognitive and Functional Trajectories in People With Late-Onset Dementia: 2 Population-Based Studies. J Am Med Dir Assoc. Nov 2019;20(11):1444-1450. doi:10.1016/j.jamda.2019.03.025
14. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. Alzheimers Dement. Nov 2012;8(6):536-43. doi:10.1016/j.jalz.2011.08.006
15. Jutten RJ, Harrison JE, Brunner AJ, et al. The Cognitive-Functional Composite is sensitive to clinical progression in early dementia: Longitudinal findings from the Catch-Cog study cohort. Alzheimers Dement (N Y). 2020;6(1):e12020. doi:10.1002/trc2.12020
16. Koster N, Knol DL, Uitdehaag BM, Scheltens P, Sikkes SAM. The sensitivity to change over time of the Amsterdam IADL Questionnaire((c)). Alzheimers Dement. Oct 2015;11(10):1231-40. doi:10.1016/j.jalz.2014.10.006

17. van der Flier WM, Scheltens P. Amsterdam Dementia Cohort: Performing Research to Optimize Care. J Alzheimers Dis. 2018;62(3):1091-1111. doi:10.3233/JAD-170850

18. McKhann GM, Knopman DS, Chertkow H, 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 & Dementia. May 2011;7(3):263-269. doi:10.1016/j.jalz.2011.03.005

19. Rascovsky K, Hodges JR, Knopman DS. Sensitivity of revised diagnostic criteria for the behavioral variant of frontotemporal dementia. Brain. 2011;134:2456-77.

20. McKeith IG, Dickson DW, Lowe J, et al. Diagnosis and management of dementia with Lewy bodies - Third report of the DLB consortium. Review. Neurology. Dec 2005;65(12):1863-1872. doi:10.1212/01.wnl.0000187889.17253.b1

21. McKeith IG, Boeve B, Dickson DW, et al. Diagnosis and Management of Dementia with Lewy Bodies: Fourth Consensus Report of the DLB Consortium. Neurology. 2017;89:1-13. doi:10.1212/WNL.0000000000004058

22. Roman GC, Tatemichi TK, Erkinjuntti T, et al. Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. Neurology. Feb 1993;43(2):250-60.

23. Petersen RC, Lopez O, Armstrong MJ, et al. Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. Neurology. Jan 16 2018;90(3):126-135. doi:10.1212/WNL.0000000000004826

24. Jessen F, Amariglio RE, Van Boxtel M, et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimers Dement. Nov 2014;10(6):844-52. doi:10.1016/j.jalz.2014.01.001

25. Jutten RJ, Peeters CFW, Leijdesdorff SMJ, et al. Detecting functional decline from normal aging to dementia: Development and validation of a short version of the Amsterdam IADL Questionnaire. Alzheimers Dement (Amst). 2017;8:26-35. doi:10.1016/j.dadm.2017.03.002

26. Sikkes SAM, Knol DL, Pijnenburg YAL, 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(1):35-41. doi:10.1159/000346277

27. Sikkes SAM, Pijnenburg YAL, 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. Dec 2013;26(4):244-50. doi:10.1177/0891988713509139

28. Saan R, Deelman B. De 15-Woordentest A en B (Een Voorlopige Handleiding). Afdeling Neuropsychologie; 1986.

29. Lindeboom J, Schmand B, Tulner L, Walstra G, Jonker C. Visual association test to detect early dementia of the Alzheimer type. J Neurol Neurosurg Psychiatry. Aug 2002;73(2):126-33. doi:10.1136/jnnp.73.2.126

30. Meyers JE, Meyers KR. Rey complex figure test under four different administration procedures. The Clinical Neuropsychologist. 1995;9(1):63-67. doi:10.1080/13854049508402059

31. Tombaugh TN. Trail Making Test A and B: Normative Data Stratified by Age and Education. Archives of Clinical Neuropsychology. 2004;19:203-214. doi:10.1016/S0887-6177(03)00039-8

32. Lindeboom J, Matto D. Digit series and Knox cubes as concentration tests for elderly subjects. Tijdschr Gerontol Geriatr. 1994;25:63-68.

33. Stroop JR. Studies of interference in serial verbal reactions. J Exp Psychol Gen. Mar 1992;121(1):15.

34. Van der Elst W, Van Boxtel MP, Van Breukelen GJ, Jolles J. Normative data for the Animal, Profession and Letter M Naming verbal fluency tests for Dutch speaking participants and the effects of age, education, and sex. J Int Neuropsychol Soc. Jan 2006;12(1):80-9. doi:10.1017/S1355617706060115

35. van der Elst W, van Boxtel MP, van Breukelen GJ, Jolles J. The Letter Digit Substitution Test: normative data for 1,858 healthy participants aged 24-81 from the Maastricht Aging Study (MAAS): influence of age, education, and sex. J Clin Exp Neuropsychol. Aug 2006;28(6):998-1009. doi:10.1080/13803390591004428

36. Warrington EK, James M. The Visual Object and Space Perception Battery. Thames Valley Test Company; 1991.

37. Scheltens P, Leys D, Barkhof F, et al. Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. J Neurol Neurosurg Psychiatry. Oct 1992;55(10):967-72. doi:10.1136/jnnp.55.10.967

38. Koedam EL, Lehmann M, van der Flier WM, et al. Visual assessment of posterior atrophy development of a MRI rating scale. Eur Radiol. Dec 2011;21(12):2618-25. doi:10.1007/s00330-011-2205-4

39. Scheltens P, Pasquier F, Weerts JG, Barkhof F, Leys D. Qualitative assessment of cerebral atrophy on MRI: inter- and intra-observer reproducibility in dementia and normal aging. Eur Neurol. 1997;37(2):95-9. doi:10.1159/000117417

40. Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. AJR Am J Roentgenol. Aug 1987;149(2):351-6. doi:10.2214/ajr.149.2.351

41. Willemse EAJ, van Maurik IS, Tijms BM, et al. Diagnostic performance of Elecsys immunoassays for cerebrospinal fluid Alzheimer's disease biomarkers in a nonacademic, multicenter memory clinic cohort: The ABIDE project. Alzheimers Dement (Amst). 2018;10:563-572. doi:10.1016/j.dadm.2018.08.006

42. Tijms BM, Willemse EAJ, Zwan MD, et al. Unbiased approach to counteract upward drift in cerebrospinal fluid amyloid-β 1–42 analysis results. Clin Chem. Mar 2018;64(3):576-585. doi:10.1373/clinchem.2017.281055

43. R: A language and environment for statistical computing. Version 4.1.3. 2022. https://www.R-project.org/

44. Proust-Lima C, Philipps V, Liquet B. Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm. Journal of Statistical Software. 2017;78(2)doi:10.18637/jss.v078.i02

45. Bates D, Machler M, Bolker BM, Walker SC. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. Oct 2015;67(1):1-48. doi:10.18637/jss.v067.i01

46. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations inR. Journal of Statistical Software. 2011;45(3)doi:10.18637/jss.v045.i03

47. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33(1):1-22.

48. Cortes F, Nourhashemi F, Guerin O, et al. Prognosis of Alzheimer's disease today: a two-year prospective study in 686 patients from the REAL-FR Study. Alzheimers Dement. Jan 2008;4(1):22-9. doi:10.1016/j.jalz.2007.10.018

49. Haaksma ML, Calderon-Larranaga A, Olde Rikkert MGM, Melis RJF, Leoutsakos JS. Cognitive and functional progression in Alzheimer disease: A prediction model of latent classes. Int J Geriatr Psychiatry. Aug 2018;33(8):1057-1064. doi:10.1002/gps.4893

50. Wang Y, Haaksma ML, Ramakers IHGB, et al. Cognitive and functional progression of dementia in two longitudinal studies. Int J Geriatr Psychiatry. Jul 18 2019;doi:10.1002/gps.5715

51. Leoutsakos JM, Forrester SN, Corcoran CD, et al. Latent classes of course in Alzheimer's disease and predictors: the Cache County Dementia Progression Study. Int J Geriatr Psychiatry. Aug 2015;30(8):824-32. doi:10.1002/gps.4221

52. Marshall GA, Rentz DM, Frey MT, et al. Executive function and instrumental activities of daily living in mild cognitive impairment and Alzheimer's disease. Alzheimers Dement. May 2011;7(3):300-8. doi:10.1016/j.jalz.2010.04.005

53. Dubbelman MA, Verrijp M, Terwee CB, et al. Pursuing clinical meaningfulness: Determining the minimal important change of everyday functioning in dementia. Neurology. accepted;

54. Jutten RJ, Dicks E, Vermaat L, et al. Impairment in complex activities of daily living is related to neurodegeneration in Alzheimer's disease-specific regions. Neurobiol Aging. Nov 26 2018;75:109-116. doi:10.1016/j.neurobiolaging.2018.11.018

55. Zahodne LB, Ornstein K, Cosentino S, Devanand DP, Stern Y. Longitudinal relationships between Alzheimer disease progression and psychosis, depressed mood, and agitation/aggression. Am J Geriatr Psychiatry. Feb 2015;23(2):130-40. doi:10.1016/j.jagp.2013.03.014

56. Pike KE, Cavuoto MG, Li L, Wright BJ, Kinsella GJ. Subjective Cognitive Decline: Level of Risk for Future Dementia and Mild Cognitive Impairment, a Meta-Analysis of Longitudinal Studies. Neuropsychol Rev. Nov 8 2021;doi:10.1007/s11065-021-09522-3

57. Slot RER, Sikkes SAM, Berkhof J, et al. Subjective cognitive decline and rates of incident Alzheimer's disease and non-Alzheimer's disease dementia. Alzheimers Dement. Mar 2019;15(3):465-476. doi:10.1016/j.jalz.2018.10.003

7.5 Supplementary Material

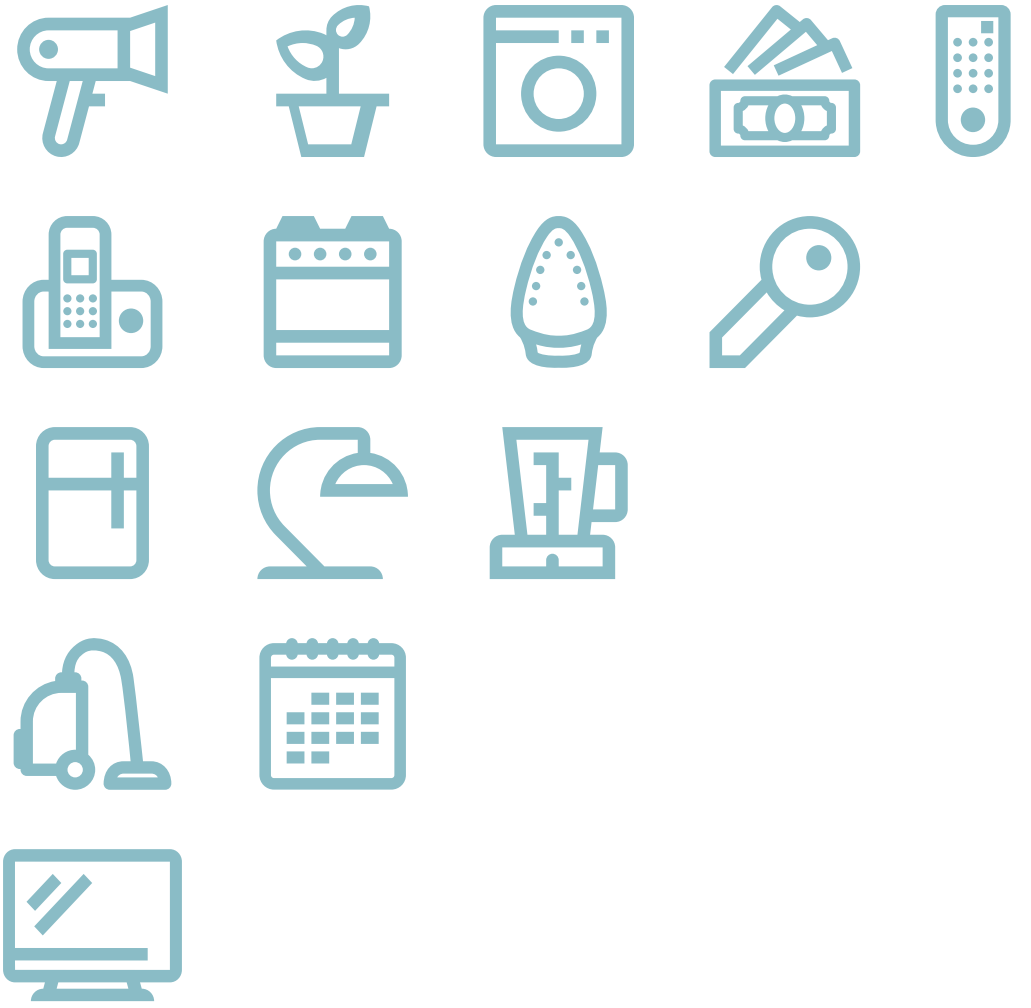
Supplemental Table 1 | Group characteristics.

Characteristic	All	Group 1 'Stable'	Group 2 'Slow Progression'	Group 3 'Fast Progression'	P
n (%)	936	282 (30.1)	221 (23.6)	433 (46.3)	
Age	64.6 ± 7.5	63.0 ± 7.4	64.8 ± 6.8	65.5 ± 7.8	< .001
Female, N (%)	369 (39.4)	112 (39.7)	77 (34.8)	180 (41.6)	.248
Education in years, m (IQR)	9 (8–10)	9.5 (9–10)	9 (8–10)	9 (8–10)	< .001
Diagnosis, N (%)					
Subjective cognitive decline	238 (25.4)	170 (60.3)	49 (22.2)	19 (4.4)	< .001
Mild cognitive impairment	158 (16.9)	54 (19.1)	68 (30.8)	36 (8.3)	< .001
Alzheimer's disease dementia	326 (34.8)	20 (7.1)	59 (26.7)	247 (57.0)	< .001
Frontotemporal dementia	26 (2.8)	2 (0.7)	4 (1.8)	20 (4.6)	.005
Dementia with Lewy bodies	71 (7.6)	4 (1.4)	8 (3.6)	59 (13.6)	< .001
Other dementia	47 (5.0)	10 (3.5)	15 (6.8)	22 (5.1)	.255
Psychiatric	22 (2.4)	8 (2.8)	7 (3.2)	7 (1.6)	.377
Married, N (%)	782 (85.7)	233 (84.1)	186 (88.6)	363 (85.4)	.366
Children, N (%)	822 (90.1)	241 (86.7)	191 (90.5)	390 (92.2)	.056
Lives in an assisted living facility, N (%)	7 (0.7)	0 (0.0)	2 (1.0)	5 (1.2)	.202
Smokes/smoked, N (%)	450 (49.2)	123 (44.2)	114 (52.3)	213 (50.8)	.134
Alcohol, weekly units	1.0 ± 1.2	1.1 ± 1.3	1.0 ± 1.1	1.0 ± 1.2	.482
Years since symptom onset	3.3 ± 2.9	3.2 ± 3.2	3.1 ± 2.2	3.4 ± 3.0	.339
Family history, N (%)					
Dementia	431 (48.0)	156 (56.9)	101 (48.1)	174 (42.1)	.001
Psychiatric disorders	190 (21.5)	61 (22.5)	38 (18.3)	91 (22.4)	.438
Cardiovascular disease	475 (53.4)	153 (56.7)	113 (53.8)	209 (51.0)	.343
Left-handed, N (%)	87 (9.9)	28 (10.4)	12 (5.9)	47 (11.7)	.074
BMI	25.6 ± 4.1	26.0 ± 4.2	25.7 ± 3.9	25.3 ± 4.0	.074
MMSE, total score	24.5 ± 4.6	27.5 ± 2.5	25.7 ± 3.1	22.1 ± 5.0	< .001
MoCA, total score	21.5 ± 5.3	24.2 ± 3.6	22.7 ± 4.6	19.3 ± 5.6	< .001
FAB, total score	14.7 ± 3.4	16.4 ± 2.1	15.6 ± 2.5	13.1 ± 3.9	< .001
GDS, total score	3.2 ± 2.8	2.8 ± 2.3	3.3 ± 3.0	3.4 ± 2.9	.028
CDR global, M (IQR)	0.5 (0.5–1)	0.5 (0–0.5)	0.5 (0.5–0.5)	1 (0.5–1)	< .001
ZBI, total score	19.5 ± 13.8	10.2 ± 9.1	20.1 ± 13.9	24.7 ± 13.2	< .001
A-IADL-Q T-score	52.0 ± 10.1	61.3 ± 6.9	52.7 ± 6.1	45.6 ± 8.4	< .001
Neuropsychological testing, cognitive domain Z-scores					
Episodic memory	-1.3 ± 1.6	-0.5 ± 1.2	-1.2 ± 1.5	-1.8 ± 1.7	< .001
Executive functioning	-0.8 ± 1.2	-0.2 ± 0.8	-0.5 ± 1.0	-1.3 ± 1.3	< .001
Attention/speed	-1.2 ± 2.0	-0.2 ± 1.0	-0.7 ± 1.1	-2.1 ± 2.5	< .001
Language	-1.1 ± 2.3	-0.3 ± 1.1	-0.8 ± 2.0	-1.8 ± 2.7	< .001
Visuospatial functioning	-0.5 ± 1.2	0.0 ± 0.7	-0.1 ± 0.7	-1.3 ± 1.6	< .001

Characteristic	All	Group 1 'Stable'	Group 2 'Slow Progression'	Group 3 'Fast Progression'	P
MR					
MTA, average, M (IQR)	1 (0–1.5)	0.5 (0–1)	1 (0.5–1.5)	1 (0.5–2)	< .001
PA, average, M (IQR)	1 (0–1)	1 (0–1)	1 (1–1.5)	1 (1–2)	< .001
GA, M (IQR)	1 (0.5–2)	0 (0–1)	1 (0–1)	1 (1–1)	< .001
Fazekas, M (IQR)	1 (0–1)	1 (0–1)	1 (0–1)	1 (0–1)	.005
Lacunar infarctions	0.2 ± 1.0	0.1 ± 0.3	0.2 ± 0.6	0.4 ± 1.4	.002
APOE ε4 carrier, N (%)					
Non-carrier	444 (49.2)	156 (57.1)	106 (49.1)	182 (44.1)	.003
Heterozygote	333 (36.9)	92 (33.7)	83 (38.4)	158 (38.3)	
Homozygote	125 (13.9)	25 (9.2)	27 (12.5)	73 (17.7)	
AD biomarkers					
Amyloid positive by CSF, N (%)	450 (64.6)	80 (39.0)	106 (62.7)	264 (81.7)	< .001
Amyloid positive by PET, N (%)	139 (50.4)	30 (29.4)	32 (57.1)	77 (65.3)	< .001

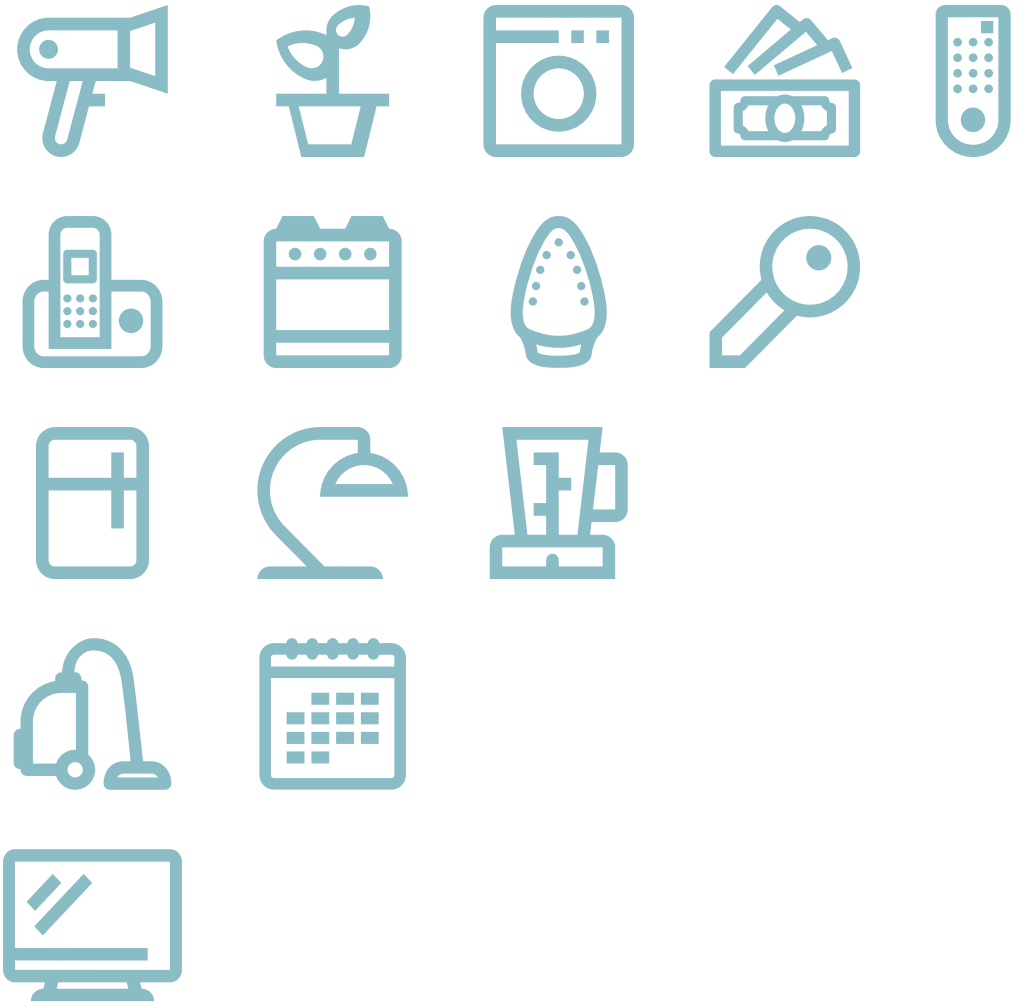
All values shown are mean ± standard deviation, except as stated otherwise. Percentages are true percentages, i.e., excluding missing observations.

Abbreviations: A-IADL-Q, Amsterdam Instrumental Activities of Daily Living Questionnaire; APOE, apolipoprotein; BMI, body mass index; CDR, Clinical Dementia Rating; CSF, cerebrospinal fluid; FAB, Frontal Assessment Battery; GA, global atrophy; GDS, Geriatric Depression Scale; IQR, interquartile range; M, median; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; MTA, medial temporal atrophy; MR, magnetic resonance; PA, parietal atrophy; PET, positron emission tomography; ZBI, Zarit Burden Interview.



CHAPTER 8

THE CLINICAL MEANINGFULNESS OF CHANGES IN EVERYDAY FUNCTIONING



CHAPTER 8.1

CLINICAL MEANINGFULNESS OF FUNCTIONAL IMPAIRMENT: GIVING MEANING TO THE SCORES OF THE AMSTERDAM INSTRUMENTAL ACTIVITIES OF DAILY LIVING QUESTIONNAIRE

Mark A. Dubbelman, Caroline B. Terwee, Merike Verrijp, Leonie N.C. Visser, Philip Scheltens, Sietske A.M. Sikkes

Published: March 24, 2022

In: Health and Quality of Life Outcomes

Abstract

Background: Everyday functioning is a clinically relevant concept in dementia, yet little is known about the clinical meaningfulness of scores on functional outcome measures. We aimed to establish clinically meaningful scoring categories for the Amsterdam Instrumental Activities of Daily Living Questionnaire (A-IADL-Q), representing no, mild, moderate, and severe problems in daily functioning.

Methods: We included informal caregivers (*n* = 6) of memory clinic patients and clinicians (*n* = 13), including neurologists and nurse specialists, working at various memory clinics in The Netherlands. In focus groups, participants individually ranked nine summaries of fictional patients from least to most impairment in daily functioning. Then they placed bookmarks to demarcate thresholds for mild, moderate, and severe problems. Individual bookmark placements were then discussed to reach consensus. Clinicians completed a survey in which they placed bookmarks, individually.

Results: While individual categorizations varied somewhat, caregivers and clinicians generally agreed on the thresholds, particularly about the distinction between 'no' and 'mild problems'. Score categories were no problems (*T*-score ≥ 60), mild problems (*T*-score 50–59), moderate problems (*T*-score 40–49), and severe problems (*T*-score <40), on a scale ranging 20–80.

Conclusion: Our findings provide categories for determining the level of functional impairment, which can facilitate interpretation of A-IADL-Q scores. These categories can subsequently be used by clinicians to improve communication with patients and caregivers.

8.1.1 Introduction

Impairment in daily functioning due to cognitive decline is a core characteristic of dementia.[1] Recent studies have shown that changes in daily functioning, in particular in 'instrumental activities of daily living' (IADL),[2] may occur well before dementia and even as early as the preclinical stage of Alzheimer's disease.[3-6] IADL comprise cognitively complex activities such as doing grocery shopping, cooking, and using a computer and, as such, reflect cognitive functions in everyday life. IADL assessments can be helpful for monitoring disease progression and evaluating treatment effects.[7,8]

Impairment in IADL is fundamentally clinically important, as it reflects a person's inability to live independently. IADL impairment is considered a key element in measuring clinically meaningful treatment effects, because it is related to reduced quality of life, caregiver burden, and apathy.[9,10] However, a given score on an IADL instrument does not directly indicate whether the level of impairment requires clinical attention.[11] Also, to patients and caregivers, the score itself does not translate to a meaningful concept of problems in daily functioning.

In this study, we set out to investigate the clinical meaningfulness of Amsterdam IADL Questionnaire (A-IADL-Q) scores by establishing clinically meaningful score cutoffs, representing no, mild, moderate, and severe problems in daily functioning. Establishing these cutoffs could aid in the meaningful interpretation of A-IADL-Q scores, which could in turn improve communication between clinicians, patients, and caregivers.

8.1.2 Methods

Participants

We asked informal caregivers of patients who visited our outpatient memory clinic between May and August 2019 to participate in a one-time, three-hour focus group. Additionally, we recruited caregivers through our center's social media accounts. We approached neurologists, geriatricians, nurse specialists and neuropsychologists from various memory clinics in the Netherlands through contacts of the authors and by using a mailing list for members of the Dutch memory clinics network ('Nederlands Geheugenpoli Netwerk').

The study was approved by the ethical review board of the VU University Medical Center, and all participants provided written informed consent.

Measures

The Amsterdam Instrumental Activities of Daily Living Questionnaire (A-IADL-Q) is an outcome measure that is self-completed by a caregiver and was designed to capture early impairment in daily functioning due to cognitive decline.[12] For the current study, we used the short version of the instrument,[13] which consists of a selection of 30 activities from the original 70-item version. Items were selected based on cross-cultural applicability, frequency of endorsement, and clinical relevance, as judged by clinicians, caregivers and patients.[13] Items are rated on a five-point scale ranging from 'no difficulty performing the activity' to 'unable to perform the activity'. The A-IADL-Q is scored using item response theory (IRT), which accounts for varying 'difficulty' of items such that impairment in a more complex activity (e.g., managing the household budget) contributes differently to the total score than impairment in a relatively simple activity (e.g., using the TV remote control). This information is contained in the scoring parameters, as described in detail elsewhere.[13,14] The total score, or *T*-score, represents the latent trait of 'daily functioning' and is normally distributed with a mean of 50 and a standard deviation (SD) of 10 in a memory clinic population. Scores thus range from approximately 20 to 80, with higher scores representing better daily functioning.

We created nine short clinical summaries ('vignettes') of fictional patients who had some degree of functional impairment, using combinations of five items of the A-IADL-Q for each vignette. We selected a subset of fifteen items to reduce the number of different activities presented in each vignette and increase comparability between them. The selection was made based on the IRT parameters to have items distributed across the latent trait, so that both more and less impaired ends of the daily functioning spectrum were covered. We then determined what item response category would be most likely to be endorsed given a certain *T*-score, based on the methods and using an R script adapted from Morgan and colleagues.[15] An overview of the most likely item responses of the fifteen items is included in the Supplementary Material. The vignettes were created by combining the most likely responses of five items at different *T*-scores (i.e., different degrees of impairment), and were placed five points (0.5 SD units) apart, ranging from 20 (all most likely item responses were 'unable to perform') to 60 (all most likely item responses were 'no difficulty'). We randomly assigned each vignette a gender, common Dutch surname, random age in the range of 60 to 70 years, and a stock photo. The vignettes can be found in the Supplementary Material.

Procedures

In the focus groups, we asked each panelist to describe what they considered 'mild', 'moderate' and 'severe problems' in daily functioning, to understand how the panelists defined these categories and create a framework for the subsequent categorization and discussion. Subsequently, panelists individually ordered the vignettes from the one representing the least functional impairment to the one representing the most. Panelists then discussed the order of the vignettes and reached a consensus ordering. Then, panelists individually placed bookmarks between the vignettes to create categories representing no, mild, moderate, and severe problems in daily functioning. This 'bookmarking' method was previously developed by Cook and colleagues.[16] Finally, a second group discussion resulted in a consensus categorization. Group discussions were based on the nominal group theory.[17]

Clinicians individually completed an online survey that was modeled after the focus group procedures, and in which they first asked to describe what they considered 'mild', 'moderate' and 'severe problems'. Next, the nine vignettes were presented in order from least to most impaired, and the clinicians were instructed to categorize them into no, mild, moderate and severe problems.

Statistical analyses

As the clinicians completed the survey independently, consensus between them was determined by taking the mode of the categorization for each vignette (1 = no problems, 2 = mild problems, 3 = moderate problems, 4 = severe problems). The overall consensus categorization was the mode of the three separate consensus categorizations: two from the focus groups with informal caregivers, and the consensus between clinicians. Analyses were performed in R version 4.1.0.[18]

8.1.3 Results

Forty patient caregivers were invited through the Alzheimer Center Amsterdam to participate in the focus groups. Six individuals (age 68 ± 10 years old, 4 women) agreed to participate, and they were spread across two focus groups. Four panelists were partners and two were adult children of a person with dementia. Clinicians were approached through contacts of the authors, as well as through a mailing list for clinicians working in memory clinics in the Netherlands. Thirteen clinicians (five

neurologists, five nurse specialists, two neuropsychologists and a geriatrician; age 46±13 years old, 8 women) completed the survey.

Caregivers and clinicians had differing definitions of what they considered 'problems in daily functioning'. One caregiver defined 'problems' as having any amount of difficulty with performing some activity, whereas another stated that they considered 'problems' to be the complete inability to perform an activity. Clinicians wrote that 'mild problems' cause minimal impairment, predominantly in the most complex activities, whereas 'severe problems' imply that a person can no longer function independently. As a result of the various personal definitions, individual categorizations differed slightly, with some panelists categorizing more strictly, where fewer problems were classified as more severe, while others were more lenient, classifying more problems as less severe. Consensus between the focus groups was largely similar, except that in one group, two more vignettes were classified as representing 'severe problems', creating a 10-point difference between the cutoffs for 'severe problems' in the two groups (see Figure 1). The vignettes at the extremes, i.e., 'no problems' and 'severe problems' were classified the same across clinicians and caregivers. The classifications of 'moderate' and 'severe' problems differed among clinicians, similar to the caregivers.

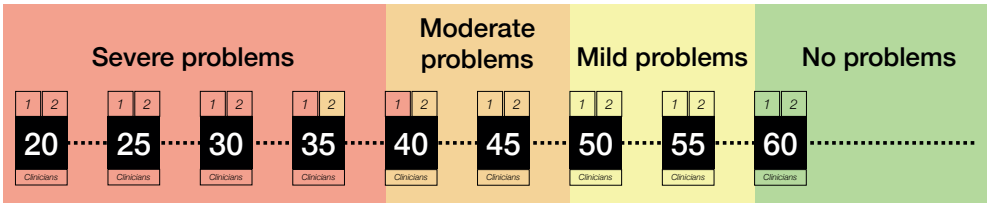


Figure 1 | Vignettes and classifications.
Each vignette is represented by a black square showing the corresponding *T*-score. The final classifications as determined in consensus are shown in the background and are color-coded: red for 'severe problems', orange for 'moderate problems', yellow for 'mild problems' and green for 'no problems'. The consensus classifications per focus group are shown directly above the vignettes (1 = focus group 1, 2 = focus group 2); the consensus classifications for clinicians are shown below.

The final average categorization was as follows: *T*-scores ≥60 were classified as showing 'no problems', *T*-scores 50–59 were classified as 'mild problems', *T*-scores 40–49 as 'moderate problems' and *T*-scores <40 as 'severe problems' (Figure 1).

8.1.4 Discussion

In this study, we involved stakeholders to determine clinically meaningful scoring categories for the measurement of functional impairment using the Amsterdam IADL Questionnaire. Informal caregivers and clinicians established categories representing no (*T*-score ≥60), mild (*T*-score 50–59), moderate (*T*-scores 40–49), and severe problems (*T*-scores <40) in IADL.

Clinical meaningfulness in the context of Alzheimer's disease (AD) and related disorders has been gaining attention over recent years.[19,20] Clinicians have a good understanding of the disease and its effects on patients and caregivers. Still, when conclusions are based solely on judgments by clinicians, these only comprise part of the picture. Especially, caregivers could add a unique perspective since they observe and can therefore reflect on functioning in AD patients in everyday life. This is a major advantage of our study.

The cutoffs between no and mild, and mild and moderate problems were unanimously agreed upon by caregivers and clinicians. This is especially important, as it seems that clear, clinically meaningful distinctions can be made in subtle degrees of IADL impairment. There was, however, some disagreement among caregivers on the precise placement of a cutoff to make the distinction between moderate and severe problems. It is arguable that the difference between these categories is of less importance, as there is already considerable impairment. Our findings show that the clinical interpretation may depend on individual definitions and opinions, which has likely contributed to the slight differences we found in categorizations. The categories we present may not reflect everyone's personal interpretation of different degrees of functional impairment.

Nevertheless, the proposed categories can help clarify the meaning of a given score, and thus provide concrete guidance for communicating test results with patients and their caregivers. This is important as many patients and caregivers report unmet information needs, especially about what test results mean.[21,22] When discussing test results, communication may benefit from the use of clear language and interpretable categories, rather than raw scores. Our study provides such ready-to-use scoring categories for the Amsterdam IADL Questionnaire.

An important strength of this work is that we used a qualitative approach involving stakeholders (both caregivers and clinicians) to determine clinically meaningful

categories in the scoring of a functional outcome measure. Limitations of this study include its small sample size, predominance of women, and recruitment in the Netherlands only, which limit the generalizability of our results. A future study should expand on our work by including a larger sample size representing a more diverse group of caregivers. Future work should also focus on the meaningfulness of changes in daily functioning, as changes may be meaningful, even when they fall entirely within the scoring categories we established here.

Conclusion

In conclusion, we used caregiver and clinician input to place thresholds and thus create meaningful categories for assessing the severity of impairment in everyday functioning in the context of AD. Specifically, these categories may be useful for distinguishing absence of any problems from the existence of mild problems, which is relevant in early disease stages. Our findings give meaning to total scores, which in and of their own are usually rather unintuitive. By providing clear language about the level of impairment, the categories could support clinicians in explaining the meaning of test results to patients and their caregivers.

Declarations

Acknowledgments

The authors would like to thank the caregivers and clinicians who provided their useful input for this study.

Authors' contributions

MAD, CBT, MV, LCNV and SAMS designed and conceptualized the study. MAD prepared all materials and performed all analyses. MAD, MV and SAMS led the focus groups. MAD wrote the first draft of the manuscript. LCNV, CBT, MV, PS and SAMS revised the manuscript. All authors read and approved the final manuscript.

Funding

The work for this study was supported by public–private funding of which SAMS was the recipient. She received funding from Health~Holland, Topsector Life Sciences & Health (PPP-allowance; LSHM20084-SGF, project DEFEAT-AD), and the National Institutes of Health, as well as license fees from Green Valley, VtV Therapeutics,

Alzheon, Vivoryon, and Roche, and honoraria from Boehringer and Toyama. All funding is paid to her institution. LNCV is supported by a fellowship grant received from Alzheimer Nederland (WE.15-2019-05) and recipient of ABOARD, which is a public–private partnership receiving funding from ZonMW (#73305095007) and Health~Holland, Topsector Life Sciences & Health (PPP-allowance; LHSM20106). All funding is paid to the institution.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

This study was approved by the ethical review board of the VU University Medical Center. All participants provided written informed consent prior to participation.

Consent for publication

Not applicable.

Competing interests

SAMS and PS co-developed the A-IADL-Q, which is freely available for use by academic users and not-for-profit organizations. License fees from for-profit organizations are paid to their institution. The other authors declare that they have no competing interests.

References

1. Scheltens P, De Strooper B, Kivipelto M, et al. Alzheimer's disease. The Lancet. 2021;doi:10.1016/s0140-6736(20)32205-4
2. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. Autumn 1969;9(3):179-86.
3. Sikkes SAM, De Rotrou J. A qualitative review of instrumental activities of daily living in dementia: what's cooking? Neurodegener Dis Manag. 2014;4(5):393-400. doi:10.2217/nmt.14.24
4. Marshall GA, Aghjayan SL, Dekhtyar M, et al. Activities of daily living measured by the Harvard Automated Phone Task track with cognitive decline over time in non-demented elderly. J Prev Alzheimers Dis. 2017;4(2):81-86. doi:10.14283/jpad.2017.10
5. Dubbelman MA, Jutten RJ, Tomaszewski Farias SE, et al. Decline in cognitively complex everyday activities accelerates along the Alzheimer's disease continuum. Alzheimers Res Ther. 2020;doi:10.1186/s13195-020-00706-2
6. Peres K, Helmer C, Amieva H, et al. Natural history of decline in instrumental activities of daily living performance over the 10 years preceding the clinical diagnosis of dementia: a prospective population-based study. J Am Geriatr Soc. Jan 2008;56(1):37-44. doi:10.1111/j.1532-5415.2007.01499.x
7. Luck T, Riedel-Heller SG, Luppa M, et al. A hierarchy of predictors for dementia-free survival in old-age: results of the AgeCoDe study. Acta Psychiatr Scand. Jan 2014;129(1):63-72. doi:10.1111/acps.12129
8. Giebel CM, Sutcliffe C, Stolt M, et al. Deterioration of basic activities of daily living and their impact on quality of life across different cognitive stages of dementia: a European study. Int Psychogeriatr. Aug 2014;26(8):1283-93. doi:10.1017/S1041610214000775
9. Zanetti O, Geroldi C, Frisoni GB, Bianchetti A, Trabucchi M. Contrasting results between caregiver's report and direct assessment of activities of daily living in patients affected by mild and very mild dementia: the contribution of the caregiver's personal characteristics. J Am Geriatr Soc. Feb 1999;47(2):196-202. doi:10.1111/j.1532-5415.1999.tb04578.x
10. Martyr A, Nelis SM, Quinn C, et al. The relationship between perceived functional difficulties and the ability to live well with mild-to-moderate dementia: Findings from the IDEAL programme. Int J Geriatr Psychiatry. Aug 2019;34(8):1251-1261. doi:10.1002/gps.5128
11. Cella D, Choi S, Garcia S, et al. Setting standards for severity of common symptoms in oncology using the PROMIS item banks and expert judgment. Qual Life Res. Dec 2014;23(10):2651-2661. doi:10.1007/s11136-014-0732-6
12. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. Alzheimers Dement. Nov 2012;8(6):536-43. doi:10.1016/j.jalz.2011.08.006
13. Jutten RJ, Peeters CFW, Leijdesdorff SMJ, et al. Detecting functional decline from normal aging to dementia: Development and validation of a short version of the Amsterdam IADL Questionnaire. Alzheimers Dement (Amst). 2017;8:26-35. doi:10.1016/j.dadm.2017.03.002
14. Sikkes SAM, Knol DL, Pijnenburg YAL, 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(1):35-41. doi:10.1159/000346277

15. Morgan EM, Mara CA, Huang B, et al. Establishing clinical meaning and defining important differences for Patient-Reported Outcomes Measurement Information System (PROMIS((R))) measures in juvenile idiopathic arthritis using standard setting with patients, parents, and providers. Qual Life Res. Mar 2017;26(3):565-586. doi:10.1007/s11136-016-1468-2

16. Cook KF, Cella D, Reeve BB. PRO-Bookmarking to Estimate Clinical Thresholds for Patient-reported Symptoms and Function. Med Care. May 2019;57 Suppl 5 Suppl 1:S13-S17. doi:10.1097/MLR.0000000000001087

17. Jones J, Hunter D. Consensus methods for medical and health services research. BMJ. Aug 5 1995;311(7001):376-80. doi:10.1136/bmj.311.7001.376

18. R: A language and environment for statistical computing. Version 4.1.3. 2022. https://www.R-project.org/

19. Siemers E, Holdridge KC, Sundell KL, Liu-Seifert H. Function and clinical meaningfulness of treatments for mild Alzheimer's disease. Alzheimers Dement (Amst). 2016;2:105-12. doi:10.1016/j.dadm.2016.02.006

20. Rentz DM, Wessels AM, Annapragada AV, et al. Building clinically relevant outcomes across the Alzheimer's disease spectrum. Alzheimers Dement (N Y). 2021;7(1):e12181. doi:10.1002/trc2.12181

21. Visser LNC, Kunneman M, Murugesu L, et al. Clinician-patient communication during the diagnostic workup: The ABIDE project. Alzheimers Dement (Amst). Dec 2019;11:520-528. doi:10.1016/j.dadm.2019.06.001

22. Fruijtier AD, Visser LNC, Bouwman FH, et al. What patients want to know, and what we actually tell them: The ABIDE project. Alzheimers Dement (N Y). 2020;6(1):e12113. doi:10.1002/trc2.12113

8.1.5 Supplementary Material

Vignette creation

We adapted an R script from Morgan and colleagues[1] to obtain the most likely responses from the IRT scoring parameters and selected fifteen optimal items to be included in the vignettes based on: (a) the distribution across the latent trait, so as to include both relatively easy and relatively difficult activities, and (b) endorsement levels, to ensure that all activities described in the vignettes were widely relevant.

The table below shows the response categories most likely to be selected at each *T*-score.

Table 1 | Most likely response categories for the items used in the vignettes.

	T = 20	T = 25	T = 30	T = 35	T= 40	T = 45	T = 50	T = 55	T = 60
Cooking	UN	UN	UN	UN	MD	MD	SMD	SMD	ND
Preparing sandwiches	UN	UN	MD	SMD	SMD	ND	ND	ND	ND
Using household appliances	UN	UN	MMD	MD	MD	SMD	SMD	ND	ND
Using the microwave	UN	UN	UN	MMD	MD	SMD	ND	ND	ND
Paying bills	UN	UN	UN	UN	UN	MD	SMD	ND	ND
Managing the paperwork	UN	UN	UN	UN	UN	MMD	SMD	SMD	ND
Withdrawing cash from an ATM	UN	UN	UN	UN	MD	SMD	ND	ND	ND
Paying with cash	UN	UN	MMD	MD	SMD	SMD	ND	ND	ND
Making and keeping appointments	UN	UN	UN	UN	MMD	MD	SMD	SMD	ND
Filling in forms	UN	UN	UN	UN	MMD	MD	SMD	SMD	ND
Working	UN	UN	UN	UN	UN	MMD	MD	SMD	ND
Using the TV remote control	UN	MMD	MMD	MD	SMD	SMD	ND	ND	ND
Driving a car	UN	UN	UN	UN	MD	SMD	SMD	ND	ND

Higher *T*-scores represent better overall IADL functioning. Item responses are coded as follows: UN, unable to perform the activity; MMD, much more difficulty; MD, more difficulty; SMD, somewhat more difficulty; ND, no difficulty.

Focus group vignettes (translated from Dutch)

1. Ms. Jonker (65 years old) [T= 60]

In the past four weeks, Ms. Jonker did not have any difficulty cooking. She also did not experience any difficulty using household appliances. She had no difficulty making and keeping appointments. She also did not have any difficulty using the TV remote control. She did not have any difficulty being responsible for her own medication.

In summary, Ms. Jonker:

- Did not have any difficulty cooking,
- Did not have any difficulty using household appliances,
- Did not have any difficulty making and keeping appointments,
- Did not have any difficulty using the TV remote control, and
- Did not have any difficulty being responsible for her own medication.

2. Ms. Smit (63 years old) [T = 55]

Ms. Smit did not have any difficulty preparing sandwiches in the past four weeks. She also did not have any difficulty obtaining the correct amount of cash from an ATM. She did not have any difficulty using public transportation. She did have somewhat more difficulty managing the paperwork than she had in the past. She also did find it more difficult to fill in forms than she had in the past.

In summary, Ms. Smit:

- Did not have any difficulty preparing sandwiches,
- Did not have any difficulty obtaining the correct amount of cash from an ATM,
- Did not have any difficulty using public transportation, but
- Had somewhat more difficulty managing the paperwork than she had in the past, and
- Had somewhat more difficulty filling in forms than she had in the past.

3. Ms. Prins (68 years old) [T = 50]

Ms. Prins did not have any difficulty paying with cash in the past four weeks. She did have somewhat more difficulty cooking, as well as somewhat more difficulty using household appliances than she had in the past. She also had somewhat more difficulty paying bills than she had in the past. She had more difficulty working than she had in the past.

In summary, Ms. Prins:

- Did not have any difficulty paying with cash, but
- Had somewhat more difficulty cooking than she had in the past,
- Had somewhat more difficulty using household appliances than she had in the past,
- Had somewhat more difficulty paying the bills than she had in the past, and
- Had more difficulty working than she had in the past.

4. Mr. Molenaar (62 years old) [T = 45]

In the past four weeks, Mr. Molenaar had no difficulty using the TV remote control. However, he did have somewhat more difficulty more difficulty using the microwave than he had in the past. He also had somewhat more difficulty driving a car than he had in the past. He had more difficulty making and keeping appointments. He had much more difficulty managing the paperwork than he had in the past.

In summary, Mr. Molenaar:

- Did not have any difficulty using the TV remote control, but
- Had somewhat more difficulty using the microwave than he had in the past,
- Had somewhat more difficulty driving a car than he had in the past,
- Had more difficulty making and keeping appointments than he had in the past, and
- Had much more difficulty managing the paperwork than he had in the past.

5. Mr. Blom (61 years old) [T = 40]

In the past four weeks, Mr. Blom had somewhat more difficulty preparing sandwiches than he had in the past. He also had somewhat more difficulty being responsible for his own medication than he had in the past. He had more difficulty cooking and had much more difficulty using public transportation than he had in the past. He was no longer able to work.

In summary, Mr. Blom:

- Had somewhat more difficulty preparing sandwiches than he had in the past,
- Had somewhat more difficulty being responsible for his own medication than he had in the past,
- Had more difficulty cooking than he had in the past,
- Had much more difficulty using public transportation than he had in the past, and
- He was no longer able to work.

6. Mr. Dekker (63 years old) [T = 35]

Mr. Dekker had more difficulty using household appliances in the past four weeks than he had in the past. He also had more difficulty paying with cash than he had in the past. He also had more difficulty using the TV remote control than he had in the past. He had much more difficulty using the microwave than he had in the past. He was no longer able to make and keep appointments.

In summary, Mr. Dekker:

- Had more difficulty using household appliances than he had in the past,
- Had more difficulty paying with cash than he had in the past,
- Had more difficulty using the TV remote control than he had in the past,
- Had much more difficulty using the microwave than he had in the past, and
- He was no longer able to make and keep appointments.

7. Ms. Vermeulen (60 years old) [T = 30]

In the past four weeks, Ms. Vermeulen had more difficulty preparing sandwiches than she had in the past. She had much more difficulty being responsible for her own medication than she had in the past. She was no longer able to pay the bills. She was also no longer able to manage the paperwork and was no longer able to work.

In summary, Ms. Vermeulen:

- Had more difficulty preparing sandwiches than she had in the past,
- Had much more difficulty being responsible for her own medication than she had in the past,
- Was no longer able to pay the bills,
- Was no longer able to manage the paperwork, and
- Was no longer able to work.

8. Mr. De Vries (66 years old) [T = 25]

In the past four weeks, Mr. De Vries had much more difficulty using the TV remote control than he had in the past. He was no longer able to use the microwave. He was also no longer able to make or keep appointments. He was no longer able to fill in forms. He was no longer able to use public transportation.

In summary, Mr. De Vries:

- Had much more difficulty using the TV remote control than he had in the past,
- Was no longer able to use the microwave,

- Was no longer able to make or keep appointments,
- Was no longer able to fill in forms, and
- Was no longer able to use public transportation.

9. Ms. De Ruiter (66 years old) [T = 20]

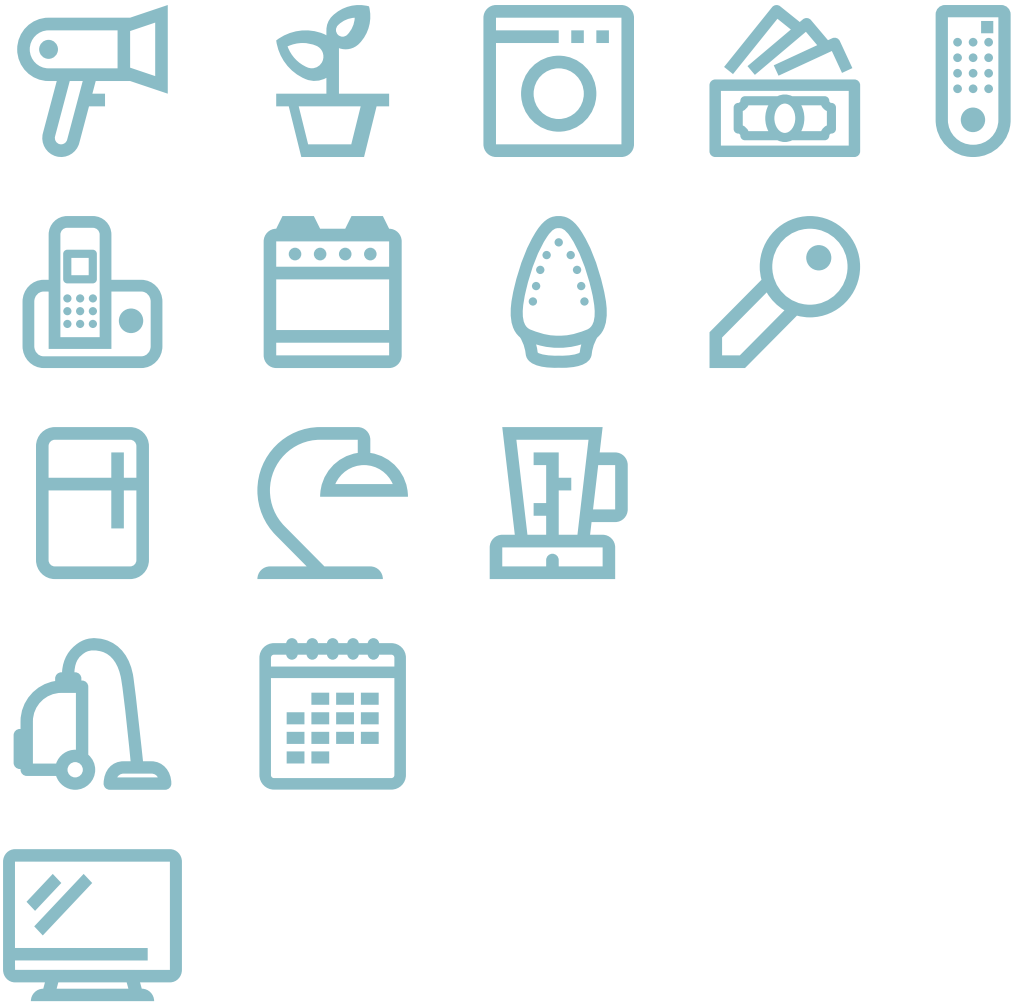
In the past four weeks, Ms. De Ruiter was no longer able to prepare sandwiches. She was also no longer able to use household appliances, or to manage the paperwork. She was no longer able to obtain the correct amount of cash from an ATM. She was no longer able to drive a car.

In summary, Ms. De Ruiter:

- Was no longer able to prepare sandwiches,
- Was no longer able to use household appliances,
- Was no longer to manage the paperwork,
- Was no longer able to obtain the correct amount of cash from an ATM, and
- Was no longer able to drive a car.

References

1. Morgan EM, Mara CA, Huang B, et al. Establishing clinical meaning and defining important differences for Patient-Reported Outcomes Measurement Information System (PROMIS((R))) measures in juvenile idiopathic arthritis using standard setting with patients, parents, and providers. Qual Life Res. Mar 2017;26(3):565-586. doi:10.1007/s11136-016-1468-2



CHAPTER 8.2

PURSUING CLINICAL MEANINGFULNESS: DETERMINING THE MINIMAL IMPORTANT CHANGE OF EVERYDAY FUNCTIONING IN DEMENTIA

Mark A. Dubbelman, Merike Verrijp, Caroline B. Terwee, Roos J. Jutten, Merel C. Postema, Frederik Barkhof, Bart N.M. van Berckel, Freek Gillissen, Vivianne Teeuwen, Charlotte E. Teunissen, Wiesje M. van der Flier, Philip Scheltens, Sietske A.M. Sikkes

Published: May 31, 2022

In: Neurology

Abstract

Background and objectives: Decline in everyday functioning is a key clinical change in Alzheimer's disease and related disorders (ADRD). An important challenge remains the determination of what constitutes a clinically meaningful change in everyday functioning. We aimed to investigate this by establishing the minimal important change (MIC): the smallest amount of change that has a meaningful impact on patients' lives. We retrospectively investigated meaningful change in a memory clinic cohort.

Methods: In the first, qualitative part of the study, community-recruited informal caregivers of ADRD patients and memory clinic clinicians completed a survey in which they judged various situations representing changes in everyday functioning. Their judgments of meaningful change were used to determine thresholds for MIC, both for decline and improvement, on the Amsterdam Instrumental Activities of Daily Living Questionnaire (A-IADL-Q). In the second, qualitative part, we applied these values in an independent longitudinal cohort study of unselected memory clinic patients.

Results: MIC thresholds were established at the average threshold of caregivers (*N* = 1,629; 62.4 ± 9.5 years; 77% female) and clinicians (*N* = 13): -2.2 points for clinically meaningful decline and +5.0 points for clinically meaningful improvement. Memory clinic patients (*N* = 230; 64.3 ± 7.7 years; 39% female; 60% dementia diagnosis) were followed for one year. One-hundred and two (45%) showed a decline larger than the MIC, after a mean of 6.7 ± 3.5 months. Patients with a dementia diagnosis and more atrophy of the medial temporal lobe had larger odds (odds ratio (OR) = 3.4, 95% confidence interval (95%CI) = [1.5, 7.8] and OR = 5.0, 95%CI = [1.2, 20.0], respectively) for passing the MIC threshold for decline than those with subjective cognitive complaints and no atrophy.

Discussion: We were able to operationalize clinically meaningful decline in IADL by determining the MIC. The usefulness of the MIC was supported by our findings from the clinical sample that nearly half of a sample of unselected memory clinic patients showed a meaningful decline in less than a year. Disease stage and medial temporal atrophy were predictors of functional decline greater than the MIC. Our findings provide guidance in interpreting changes in IADL and may help evaluate treatment effects as well as monitor disease progression.

8.2.1 Introduction

Alzheimer's disease and related disorders (ADRD) are characterized by a gradual decline in cognitive and daily functioning, eventually leading to dementia.[1] Although changes in cognitively complex 'instrumental activities of daily living' (IADLs) may occur in preclinical and prodromal disease stages,[2,3] little is known about the clinical meaningfulness of these initial changes. Determining clinical meaningfulness has become especially important, as treatment and preventions studies are increasingly targeting early populations.[4,5] Regulatory agencies emphasize that the clinical efficacy of newly developed drugs should be predicated on a meaningful effect on relevant outcome measures.[6]

The clinical meaningfulness of changes addresses a fundamental issue: what amount of change on a clinical outcome measure constitutes a change that is meaningful, or important, to the patient? This question has only been sparsely investigated, and definitions are inconsistent. Some have argued that the mere presence of any change in performance on questionnaires addressing everyday functioning is clinically meaningful.[7,8] Others have reasoned that clinical meaningfulness comprises prediction of future conversion from normal cognition to mild cognitive impairment (MCI) or dementia.[9] The first definition may overgeneralize and include changes due to noise, while the second may miss more subtle changes that can still have an impact on a patient's life. In the present work, we use the term 'minimal important change' (MIC), which has been defined as the smallest within-person change that is important to the patient.[10,11]

The MIC can be determined using anchors,[12] in which an external appraisal of the change, such as a single question on global perceived change, is used as an 'anchor' to determine a MIC on an instrument (e.g., "On a scale of 0–10, how would you describe the patient now, compared to one year ago? (0: no change, 10: much worse)"). A downside of this method is that the MIC then depends on the anchor and the anchor's quality. It has been shown that the anchor can be more strongly influenced by the patient's final status, rather than reflecting the actual change.[13] An alternative can be found in a new systematic, qualitative approach in which stakeholders (i.e., patients, caregivers, clinicians) are asked to compare fictional patient summaries with different levels of impairment in the area that is being measured.[14] Thresholds are then placed at the first point where the stakeholders indicate that a difference is meaningful.[14] The thresholds thus represent the MIC, and any change beyond it is deemed clinically meaningful.

We set out to establish the thresholds for MIC on the Amsterdam IADL Questionnaire (A-IADL-Q), an extensively validated measure of everyday functioning.[15,16] Subsequently, we applied the MIC thresholds to data from a cohort of memory clinic patients and registered how many passed the MIC threshold and which demographic, biological and neuropsychological factors were associated with surpassing the MIC threshold.

8.2.2 Methods

Our study comprised two parts: a qualitative part to establish the MIC thresholds and a quantitative part in which we applied the MIC to a cohort of memory clinic patients, to investigate the frequency of passing the MIC threshold within one year and which factors were associated with surpassing the MIC threshold.

Standard protocol approvals, registrations, and patient consent

The study was approved by the ethical review board of the VU University Medical Center. All included participants provided informed consent for the use of their data, in accordance with the Declaration of Helsinki.

Establishing MIC thresholds

Participants

We recruited participants for an online survey to establish MIC thresholds on the A-IADL-Q through the Dutch Brain Research Registry (hersenonderzoek.nl).[17] We selected people who indicated that they were direct relatives and/or informal caregivers of people diagnosed with a dementia-related diagnosis. Potential participants were excluded if they reported to have received such diagnosis themselves. Recruitment ran from February to April 2020. We also invited clinicians (neurologists, geriatricians, nurse specialists and neuropsychologists) working in memory clinics in the Netherlands, to complete the same survey.

Materials: Amsterdam IADL Questionnaire

The A-IADL-Q is an adaptive questionnaire aimed at measuring functional impairment in early dementia.[16] The questionnaire is self-administered and completed by a caregiver. Previous studies have shown robust psychometric properties, including sensitivity to change and good construct validity.[18,19] The

questionnaire consists of 70 items assessing cognitively complex everyday activities. Total scores ('*T*-scores') are computed using item response theory (IRT), which uses mathematical models to calculate probabilities for item endorsement given a person's ability. This scoring method is described in more detail elsewhere.[15,19] The *T*-scores have a mean of 50 and a standard deviation of 10 in the memory clinic. Lower scores indicate more impairment.

Materials: Vignettes

We created eighteen vignettes using IRT item parameters that showed the most likely item responses at different total scores, i.e., at different levels of functional impairment. To find the most likely responses at various *T*-scores, we used a script created by Morgan and colleagues.[21] To obtain the optimal balance between distinguishable levels of functional impairment and small distances between the vignettes, they were placed 0.2 standard deviations apart. We created six reference vignettes, spread across the total score distribution, and representing different base levels of everyday functioning. Cases were given a random sex and common last name and placed at the following *T*-scores: (1) 'Ms. Smith', *T* = 54; (2) 'Mr. Jones', *T* = 50; (3) 'Mr. Williams', *T* = 46; (4) 'Ms. Brown', *T* = 42; (5) 'Ms. Johnson', *T* = 38; and (6) 'Mr. Garcia', *T* = 34. More details about the vignette creation can be found in the Supplementary Material.

Procedures

Survey respondents (both caregivers and clinicians) were randomly branched into one of six groups, each of which received a different 'case' with a unique reference vignette. They were then shown seven 'comparison vignettes', which ranged from -8 to +6 points from the reference vignette.

Following the procedures outlined by Cook et al.,[14] we presented vignettes in pairs, with the reference vignette representing the patient's functioning "one year ago", and each comparison vignette representing a new situation "now". Respondents judged whether the functioning "now" was better, worse, or the same as "one year ago" (see Figure 1). If the respondent considered there to be a decline or an improvement, they were then asked to state whether the decline or improvement in functioning would make a meaningful difference in everyday life. This was the core question of the survey. If the respondent judged both vignettes to represent the same level of daily functioning, the next situation was shown.

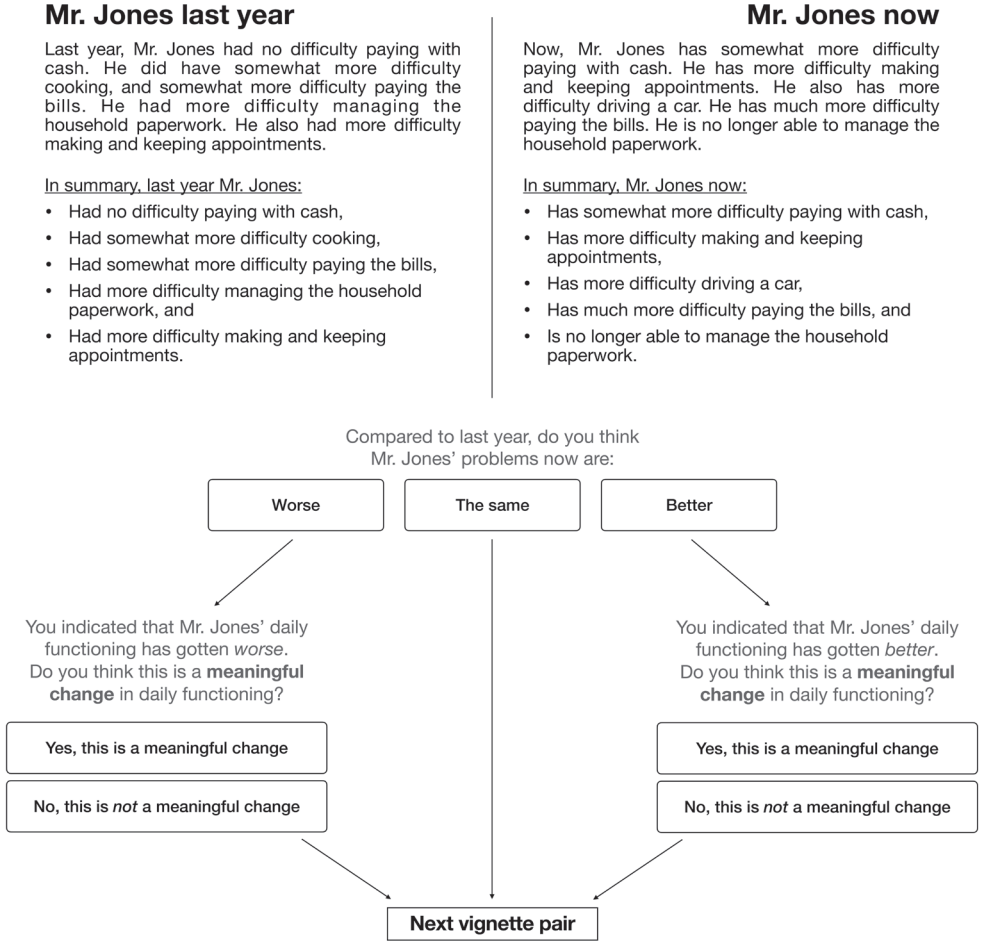


Figure 1 | Example question from the MIC survey.
First, two vignettes are shown side-by-side, with one representing functioning of a fictional patient one year ago (the 'reference vignette' on the top left) and one representing functioning now (the 'comparison vignette' on the top right). The respondent is asked to indicate whether they think the problems have worsened, remained the same, or improved from one year ago to now. Depending on the answer, they will be asked a follow-up question to determine whether the change (if any) was meaningful. This process was repeated a total of 7 times.

Individual MIC thresholds resulting from the survey responses represent the smallest change indicated as being meaningful. Thus, the score difference for the first situation that the respondent rated as a meaningful change in daily functioning was considered the threshold for MIC. Thresholds were determined separately for decline and improvement and could range from -8 to -2, and +2 to +6, respectively. When a respondent did not rate any of the presented comparison

vignettes as a clinically meaningful change, their threshold was considered missing. We also investigated two types of misjudgment. First, when a respondent judged a comparison vignette anchored on a score representing more severe functional impairment than the reference vignette as an improvement (or vice versa), this judgment was considered out-of-range and treated as a judgment of no change. Second, we examined paradoxical judgments. When a smaller distance between reference and comparison vignettes was rated as a meaningful change and a larger distance was not (e.g., a four-point decrease is judged as meaningful, whereas a six-point decrease is not), the latter judgment is considered paradoxical.

MIC in clinical practice

Participants and procedures

Next, we applied the MIC thresholds retrospectively to a cohort of consecutive memory clinic patients and their caregivers from the Amsterdam Dementia Cohort,[20] who visited Alzheimer Center Amsterdam for dementia screening between July 2013 and May 2015. Eligibility criteria were: (1) a completed baseline A-IADL-Q from the screening visit; (2) the presence of a caregiver; (3) the availability to complete the follow-up A-IADL-Q online at home; and (4) adequate knowledge of the Dutch language. We did not select for diagnosis.

At the baseline visit, caregivers completed the A-IADL-Q while the patients underwent a standard neuropsychological test battery. The screening visit also included a neurological exam, brain magnetic resonance imaging (MRI), and a lumbar puncture.[20] Diagnoses were made in a multidisciplinary consensus meeting in which the results from the screening visit were discussed.[20] Clinical diagnoses were made according to the criteria for subjective cognitive decline, mild cognitive impairment, dementia, Alzheimer's disease, frontotemporal dementia, dementia with Lewy bodies, and vascular dementia.[20] Non-Alzheimer's disease types of dementia were grouped to avoid small group sizes.

Caregivers were then invited to complete the A-IADL-Q from home at four follow-up waves: three, six, nine, and twelve months after baseline. At each follow-up wave, caregivers were also asked to rate on a visual analogue scale ranging from 0 (no decline/no burden) to 100 (very large decline/very large burden) (1) how much they thought the patient declined from baseline and (2) how much burden they experienced from taking care of the patient. These two questions served as

anchors. They could opt out at any point during the study. Invitations to participate were sent through e-mail at each wave, even when a previous wave was missed, unless the caregiver explicitly opted out of the study.

Measures

A standardized neuropsychological assessment was performed at baseline and included the Dutch version of the Auditory Verbal Learning Task[21] and the Visual Association Test,[22] to measure episodic memory. The Trail Making Test part B,[23] Wechsler Adult Intelligence Scale (WAIS) Digit Span backwards,[24] letter fluency,[25] and Stroop Color-Word Task card III[26] were used to measure executive functioning. Attention and speed were measured using the Trail Making Test part A,[23] Stroop Color-Word Task card I,[26] the Letter Digit Substitution Test[27] and the WAIS Digit Span forward.[24] Language tasks included the naming portion of the Visual Association Test[22] and the category fluency (animal naming) task.[25]

We calculated Z-scores for the neuropsychological domains: episodic memory, executive functioning, attention/speed of processing, and language. Prior to Z-scoring, tests were reverse scored as necessary so that higher Z-scores represent better cognitive functioning. The Z-scores were computed using the means and standard deviations of the measures in the entire sample.

The 15-item version of the Geriatric Depression Scale (GDS) was used as an indicator for depressive symptoms.[28] with higher scores representing more severe depressive complaints. The Zarit Burden Interview (ZBI) was used to determine the level of burden the caregiver experienced from caring for the patient, with scores ranging from 0 to 88 and higher scores indicating a larger caregiver burden.[29]

At baseline, patients underwent a standard MRI protocol on a 1.5 or 3 Tesla scanner.[20] All scans were visually rated by a radiologist who was blind to other clinical information. Visual rating scales were used on T1-weighted and fluid-attenuated inversion recovery images to provide measures of atrophy and other neurodegenerative structural changes and included the medial temporal atrophy (MTA) scale,[30] the posterior atrophy (PA) scale,[31] the global cortical atrophy (GCA) scale,[32] and the Fazekas scale[238] for white matter hyperintensities. Cerebral microbleeds were counted.

Amyloid beta$_{1\text{-}42}$ (Aβ) levels in cerebrospinal fluid were measured using ELISA (Innogenetics-Fujirebio, Ghent, Belgium) at the Neurochemistry laboratory.[33] We

dichotomized amyloid status into negative or positive for AD, based on our center's cutoff of <813 pg/mL.[34] We also computed the ratio between phosphorylated tau and Aβ. A subset of participants underwent amyloid positron emission tomography (PET) scans, using either ^{11}C-Pittsburgh compound-B, ^{18}F-flutemetamol, ^{18}F-florbetapir or ^{18}F-florbetaben. The result of the PET scan was dichotomized as either negative or positive for AD, based on visual read by an independent nuclear radiologist.

APOE genotyping was performed after automated genomic DNA isolation from 2–4 mL EDTA blood. It was subjected to PCR testing, checked for size and quantity using a QIAxcel DNA Fast Analysis kit (Qiagen), and sequenced using Sanger sequencing on an AB130XL. Patients with either one or two ε4 alleles were classified as APOE ε4 carriers.

Statistical analyses

To obtain MIC thresholds, we averaged individual thresholds separately for each of the six cases, as well as all informal caregivers, clinicians, and the entire survey sample. Taking the average thresholds of all caregivers and the average thresholds of the clinicians, we established the final MIC thresholds as the average of the two.

In the clinical cohort, patients were divided into three groups at each follow-up visit, based on whether they surpassed the thresholds for MIC: (1) patients showing no meaningful change, (2) patients showing a meaningful decline, and (3) patients showing a meaningful improvement. In addition, patients were also classified in the same groups as based on their last visit (i.e., final status). The time in months from baseline to the first visit at which the MIC thresholds were surpassed, was also recorded.

Group differences were tested using linear or logistic regressions, as appropriate. Tukey's range test was used to correct for multiple comparisons. Possible attrition bias was investigated by comparing baseline characteristics of patients who completed the last follow-up wave to those who dropped out.

Finally, we ran multinomial logistic regression models to identify baseline characteristics that were associated with the MIC groups (decline or improvement greater than the MIC, with no change beyond the MIC as the reference group), including screening instruments (MMSE, GDS, ZBI, diagnostic group),

neuropsychological assessments (episodic memory, executive functioning, attention/processing speed, and language domain Z-scores), Alzheimer's disease genetic risk factors and amyloid biomarkers, and MRI. All factors were investigated individually, with adjustments for gender, education, baseline age and syndrome diagnosis (SCD, MCI or dementia).

Analyses were run in R version 4.1.1,[35] using the 'nnet' package version 7.3-16 for the multinomial logistic regressions.[36]

Data availability

Data not provided in the article because of space limitations may be shared (anonymized) at the request of any qualified investigator for purposes of replicating procedures and results.

8.2.3 Results

Establishing the MIC

A total of 1,629 caregivers (mean age 62.4 ± 9.5 years, 77% female) completed the survey to establish the MIC thresholds. Most caregivers (75%) were adult children of people diagnosed with dementia, others were partners (6%), friends (3%), or other relatives (16%). Thirteen clinicians (five neurologists, five nurse specialists, two neuropsychologists and a geriatrician) completed the survey.

Almost all caregivers (n = 1,599; 98%) rated at least one of the situations as showing an important decline. An overview of how many caregivers reached the MIC threshold in each situation can be found in the Supplementary Material. We observed a difference in the proportion of caregivers who reached the threshold between those who saw the case with the lowest reference *T*-score ('Mr. Garcia', T = 34) and all other cases ($p < .001$). The average MIC threshold for decline was 2.4 ± 1.0 points for all caregivers (see Table 1). The average threshold varied by the reference vignette: caregivers who judged the 'Mr. Garcia' case with the lowest *T*-score had the highest average threshold. The average threshold was also significantly higher in the group of caregivers who judged the case with a *T*-score of 50, compared to the other groups. Most participants (n = 1,216; 75%) made no paradoxical judgments for decline. Clinicians unanimously rated the smallest decline in scores as an important decline, placing the clinicians' MIC for decline at -2.0.

Table 1 | Minimal important change thresholds.

Group	N	Decline		Improvement	
		Reached threshold	Average threshold	Reached threshold	Average threshold
Caregivers	1,629	1,599 (98.2)	-2.4 ± 1.0	362 (22.2)	+4.7 ± 1.3
'Ms. Smith' (T = 54)	268	265 (98.9)	-2.3 ± 1.0	101 (37.7)	+4.9 ± 1.2
'Mr. Jones' (T = 50)	260	257 (98.8)	-2.6 ± 1.0	141 (54.2)	+4.9 ± 1.3
'Mr. Williams' (T = 46)	284	283 (99.6)	-2.1 ± 0.4	10 (3.5)	+5.4 ± 1.4
'Ms. Brown' (T = 42)	265	263 (99.2)	-2.2 ± 0.8	65 (24.5)	+4.0 ± 1.0
'Ms. Johnson' (T = 38)	282	278 (98.6)	-2.3 ± 1.0	29 (10.3)	+4.2 ± 1.4
'Mr. Garcia' (T = 34)	270	253 (93.7)	-2.9 ± 1.6	16 (5.9)	+4.4 ± 1.1
Clinicians	13	13 (100.0)	-2.0 ± 0.0	5 (38.5)	+5.2 ± 1.1

Note: Thresholds are displayed as mean ± standard deviation.

Most participants (*n* = 1,078; 66%) made no paradoxical judgments for improvement. Only 362 caregivers (22%) rated any of the improvements as important. In the groups where the reference vignette had a higher level of functioning (*T* = 54 and *T* = 50), more caregivers reached the MIC threshold for improvement. The average MIC threshold for improvement was 4.7 ± 1.3 points (see Table 1). Five clinicians detected a meaningful improvement, with an average threshold of 5.2 ± 1.1.

Taken together, the MIC threshold for decline was established at -2.2 (i.e., the average of -2.4 for caregivers and -2.0 for clinicians), with a decline of 2.2 points or more indicating a meaningful decline. The MIC threshold for improvement was established at +5.0 (i.e., the average of +4.7 for caregivers and +5.2 for clinicians), meaning that an increase in the *T*-score of 5.0 points or more shows a meaningful improvement in everyday functioning.

The MIC in clinical practice

We included 230 patients (64.3 ± 7.7 years, 39% female) in the clinical cohort. They had diagnoses of subjective cognitive decline (*n* = 37), mild cognitive impairment (*n* = 22), AD dementia (*n* = 81), non-AD dementia (*n* = 58), or a different diagnosis (*n* = 36). Mean follow-up duration was 8.8 ± 3.4 months.

The number of patients showing a meaningful decline from baseline increased with each follow-up wave, whereas the number of patients showing meaningful improvement or no meaningful change declined. In subsequent analyses, we used the groups as defined at the patient's last completed visit. At the last visit, 104 patients (45%) showed a meaningful decline, while 36 (16%) showed a meaningful

improvement. The remaining 90 patients (39%) did not show a meaningful change during their follow-up. The anchors indicated that there was a stronger decline from baseline in the patients who surpassed the MIC for decline (mean 39.0 ± 30.0) than patients who showed no meaningful change (19.3 ± 21.5; mean difference $p < .001$) or meaningful improvement (12.1 ± 17.2; mean difference $p < .001$). Similarly, caregivers experienced a greater burden from taking care of patients who surpassed the MIC for decline (38.2 ± 28.5) than patients who did not change meaningfully (29.2 ± 26.0; mean difference $p < .001$) and patients who surpassed the MIC for improvement (15.7 ± 23.2; mean difference $p < .001$).

Table 2 | MIC per follow-up wave

Follow-up wave	n (%)	Meaningful decline from baseline	Meaningful improvement from baseline	No meaningful change from baseline
3 months	159 (69.1)	56 (35.2)	25 (15.7)	78 (49.1)
6 months	123 (53.5)	55 (44.7)	22 (17.9)	46 (37.4)
9 months	102 (44.3)	50 (49.0)	13 (12.7)	39 (38.2)
12 months	88 (38.3)	48 (54.5)	9 (10.2)	31 (35.2)
Last completed visit	230 (100.0)	104 (45.2)	36 (15.7)	90 (39.1)

Note: An overview of the number of patients surpassing the MIC broken down by diagnostic group can be found in the Supplementary Material.

Table 2 shows the number of patients who surpassed the MIC thresholds for decline and improvement. Most patients passed the MIC thresholds consistently across all visits: only 34 patients (15%) inconsistently passed the MIC thresholds, 12 of whom (35%) surpassed the MIC for decline initially but ended up not showing a meaningful change, and 10 of whom (29%) surpassed the MIC for improvement initially but ended up showing no meaningful change. Table 3 shows the number of patients who reached the MIC thresholds for decline and improvement and the average time in months it took to reach them, for the entire sample, as well as for each diagnostic group separately. There were no significant differences between any of the diagnostic groups in time to reach the MIC threshold for either decline or improvement, after correction for multiple comparisons.

Table 3 | Time in months until the MIC threshold was reached.

	Decline		Improvement	
Group	**N (%)**	**Time**	**N (%)**	**Time**
All	104 (45.2)	6.72 ± 3.50	90 (39.1)	6.81 ± 3.52
SCD	7 (18.9)	9.70 ± 4.38	21 (56.8)	7.16 ± 3.48
MCI	9 (40.9)	5.72 ± 3.76	8 (36.4)	3.65 ± 2.77
AD dementia	43 (53.1)	6.65 ± 3.21	29 (35.8)	8.45 ± 3.43
Non-AD dementia	34 (58.6)	6.92 ± 3.60	20 (34.5)	5.73 ± 1.53
Other	15 (41.7)	5.69 ± 3.07	12 (33.3)	7.52 ± 4.01

Figure 2 shows a visual representation of the number of patients at each follow-up wave that surpassed the threshold for meaningful decline and improvement.

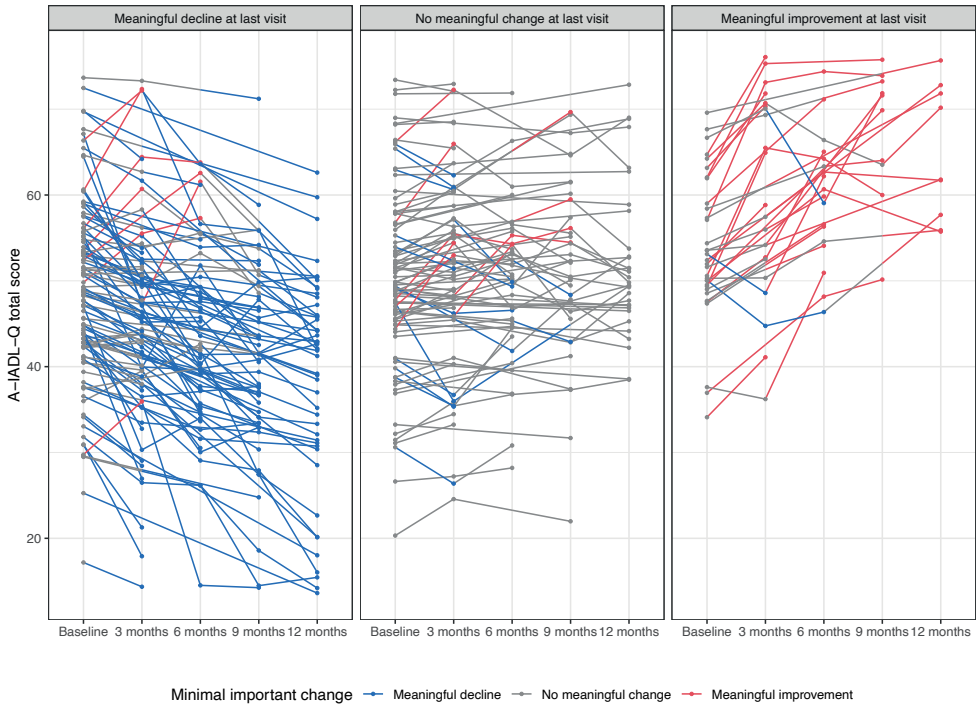


Figure 2 | Change in IADL functioning over one year.
The individual lines are colored based on whether the patient had a meaningful decline (blue), no meaningful change (gray) or a meaningful improvement (red) in IADL functioning compared to their baseline level of functioning.

Multinomial logistic regressions showed that those with a dementia diagnosis were more likely to surpass the MIC threshold for decline (odds ratio (OR) = 2.53, 95%

confidence interval (95%CI) = [1.05, 6.12], p = .039) as well as less likely to surpass the MIC threshold for improvement (OR = 0.35, 95%CI = [0.13, 0.94], p = .037), compared to patients with SCD. Patient with an MTA score of 1.5 were more likely to pass the MIC threshold for decline, compared to patients with an MTA of 0 or 0.5 (OR = 4.97, 95%CI = [1.23, 19.99], p = .024). When the caregiver experienced a higher burden, i.e., had a higher ZBI, the odds of the patient surpassing the MIC threshold for improvement was lower (OR = 0.89, 95%CI = [0.82, 0.97], p = .009). No associations were found between the MIC groups and the other determinants we investigated (including age, sex, education, AD biomarkers, objective cognitive performance, and other MRI variables; see Supplementary Material).

8.2.4 Discussion

In this study, we involved informal caregivers and clinicians of patients with Alzheimer's disease and related disorders to determine what amount of change in functional impairments constitutes a clinically meaningful change. We established thresholds for the minimal important change, both for evaluating meaningful decline and meaningful improvement on the A-IADL-Q. We found that patients with dementia and more severe atrophy of the medial temporal lobe were more likely to show a meaningful decline in daily functioning than patients with SCD and with no atrophy.

The clinical meaningfulness of changes in cognitive and functional measures is of vital importance to track disease progression in clinical practice. It is also important for evaluating potential treatment effects. Full approval by the United States Food and Drug Administration of disease-modifying treatments is contingent on the evidence of a meaningful benefit,[6] yet the interpretation of outcome measures remains difficult[37] and there is considerable variability in how clinical meaningfulness is defined and investigated. Consensus is yet to be reached.[38] Some methods have methodological and conceptual limitations, including inadequate reliability and validity.[14,39,40] Distribution-based methods rely on statistics and are neither informed by clinical information nor do they translate to what is clinically meaningful. External anchors can give an indication of the perceived magnitude or importance of a change, but they may also be affected by current status,[13,40] which renders them less reliable for investigating the clinical meaningfulness of changes. More importantly, neither method considers input from the target population, even though only the individuals themselves, and those who are close

to them, can indicate whether a change is impactful or not. Still, these methods are commonly used in dementia research,[8,41-44] possibly because more elaborate qualitative approaches require extensive work. Our study is unique in the field of ADRD research in that it employs a systematic qualitative method involving the most important stakeholders.

Overall, we found that most caregivers considered the smallest amount of decline clinically meaningful. This suggests that even subtle decline in IADL functioning has a meaningful impact on the daily life of a patient. Depending on the base level of functioning, slightly differing amounts of change were considered meaningful. When someone's level of functioning is more impaired, a stronger decline may be necessary before it is considered meaningful. When functioning is relatively good, a small decline in functioning appears to have a meaningful impact.

When looking at changes in the opposite direction, we found that only when impairments were initially relatively limited, more than half the respondents identified important improvements. Interestingly, however, the thresholds for minimal important improvement were higher when the level of functioning was better, compared to when there were more impairments at baseline. This finding seems to suggest that meaningful improvement from a more impaired status may require a somewhat smaller change, whereas meaningful improvement from a less impaired baseline may only occur when the change is relatively large.

This last finding links to another important point of discussion in the context of disease-modifying treatments and prevention studies: does the absence of a meaningful decline constitute a clinical benefit, or should a meaningful improvement be achieved? We found that determining the threshold for meaningful improvement was much more difficult than for decline. Less than a quarter of caregivers considered any of the situations to represent a meaningful improvement, which seems to implicate that improvements in functioning need to be larger before they have an impact on daily life. However, it is also possible that imagining an improvement in daily functioning in the context of dementia is difficult, as this is currently not a reality. With the rapid developments in drug development,[24] the exercise of establishing MIC thresholds on outcome measures may need to be repeated, as our understandings of what is possible change.

The second part of our study was to apply the MIC thresholds in a real-life dataset. Just under half of a non-selected group of memory clinic patients passed the MIC

threshold for decline within one year and thus showed a meaningful decline, on average within approximately seven months. Patients who were diagnosed with dementia were more likely to show a meaningful decline than those diagnosed with subjective cognitive decline. Those with more medial temporal atrophy were more likely to show a meaningful decline than those with no atrophy. When the caregiver experienced a large burden, the patient was less likely to surpass the MIC threshold for improvement. These findings provide further evidence that biological and cognitive factors underlie changes in IADL functioning: we previously found that any decline in IADL functioning was associated with disease severity, i.e., that patients with dementia declined faster than patients with subjective cognitive complaints,[18] and that worse IADL performance was associated with atrophy in the medial temporal lobe.[45] Studies with other IADL measures related changes in IADL to disease stage,[3] amyloid burden,[46] and executive functioning,[47] irrespective of the clinical meaningfulness of changes. In the present work, we show that disease stage, atrophy and caregiver burden are associated with clinically meaningful changes in everyday functioning. It is therefore recommended that these factors be included in research of disease progression.

This study has some limitations. The qualitative method we used in the first part of the study is relatively new, which means that methodological guidelines are yet to be established. We followed earlier work and presented changes that ranged from one fifth to four fifths of a standard deviation in the total score. Had we presented a smaller amount of change (e.g., a tenth of a standard deviation), it is possible that the MIC threshold would still be lower. However, such small changes may have been too subtle to distinguish and may also fall within the measurement error of the instrument. Similarly, if we had included larger amounts of change, more respondents may have reached the MIC threshold for improvement, which would then be more reliable. Future studies could replicate our findings in new samples, including outside of The Netherlands and representing individuals with different backgrounds and older ages. In the second part of our study, non-adherence was quite high. Dropouts and missed visits may have affected our estimates of the number of patients who passed the MIC thresholds. It is possible that patients who decline more severely discontinued their participation in the study, which may have led to an underestimation of actual decline. We did not find that patients who dropped out differed from those who completed the last visit, making this a less likely explanation. A further limitation was that we applied the MIC thresholds

retrospectively and therefore did not ask the participants in the clinical sample whether they agreed with the MIC category that their loved one fell into. However, we did find that, on the anchor questions, participants indicated that their loved ones declined more strongly, and that the caregiver burden was larger, when the patient passed the MIC for meaningful decline.

A particular strength of this study was our qualitative approach to establish thresholds for meaningful changes, involving different stakeholders (informal caregivers and clinicians). The frequent measurements with short intervals allowed us to pinpoint after how much time each patient first passed the threshold for meaningful decline. Finally, all patients underwent an elaborate diagnostic workup which provided a clear clinical diagnosis and allowed us to investigate a range of baseline characteristics to relate to IADL changes.

In conclusion, we performed a crucial investigation of the clinical meaningfulness of changes in IADL functioning. We applied a qualitative method involving stakeholders to determine the smallest amount of change in everyday functioning that has a meaningful impact on the patient's life and applied the thresholds we established to a cohort of memory clinic patients. Our findings have implications for evaluating possible treatment effects in clinical trials, as well as for monitoring disease progression in clinical practice.

References

1. Scheltens P, De Strooper B, Kivipelto M, et al. Alzheimer's disease. The Lancet. 2021;doi:10.1016/s0140-6736(20)32205-4
2. Marshall GA, Aghjayan SL, Dekhtyar M, et al. Activities of daily living measured by the Harvard Automated Phone Task track with cognitive decline over time in non-demented elderly. J Prev Alzheimers Dis. 2017;4(2):81-86. doi:10.14283/jpad.2017.10
3. Dubbelman MA, Jutten RJ, Tomaszewski Farias SE, et al. Decline in cognitively complex everyday activities accelerates along the Alzheimer's disease continuum. Alzheimers Res Ther. 2020;doi:10.1186/s13195-020-00706-2
4. Edgar CJ, Vradenburg G, Hassenstab J. The 2018 Revised FDA Guidance for Early Alzheimer's Disease: Establishing the Meaningfulness of Treatment Effects. J Prev Alz Dis. 2019;6doi:10.14283/jpad.2019.30
5. Liu KY, Schneider LS, Howard R. The need to show minimum clinically important differences in Alzheimer's disease trials. The Lancet Psychiatry. 2021;doi:10.1016/s2215-0366(21)00197-8
6. Early Alzheimer's disease: developing drugs for treatment - guidance for industry. 2018.
7. Siemers E, Holdridge KC, Sundell KL, Liu-Seifert H. Function and clinical meaningfulness of treatments for mild Alzheimer's disease. Alzheimers Dement (Amst). 2016;2:105-12. doi:10.1016/j.dadm.2016.02.006
8. Rockwood K, Howlett SE, Hoffman D, Schindler R, Mitnitski A. Clinical meaningfulness of Alzheimer's Disease Assessment Scale-Cognitive subscale change in relation to goal attainment in patients on cholinesterase inhibitors. Alzheimers Dement. Oct 2017;13(10):1098-1106. doi:10.1016/j.jalz.2017.02.005
9. Papp KV, Buckley R, Mormino E, et al. Clinical meaningfulness of subtle cognitive decline on longitudinal testing in preclinical AD. Alzheimers Dement. Mar 2020;16(3):552-560. doi:10.1016/j.jalz.2019.09.074
10. De Vet HCW, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes. Aug 22 2006;4:54. doi:10.1186/1477-7525-4-54
11. Terwee CB, Peipert JD, Chapman R, et al. Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res. Oct 2021;30(10):2729-2754. doi:10.1007/s11136-021-02925-y
12. Devji T, Carrasco-Labra A, Qasim A, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. BMJ. Jun 4 2020;369:m1714. doi:10.1136/bmj.m1714
13. de Vet HC, Foumani M, Scholten MA, et al. Minimally important change values of a measurement instrument depend more on baseline values than on the type of intervention. J Clin Epidemiol. May 2015;68(5):518-24. doi:10.1016/j.jclinepi.2014.07.008
14. Cook KF, Kallen MA, Coon CD, Victorson D, Miller DM. Idio Scale Judgment: evaluation of a new method for estimating responder thresholds. Qual Life Res. Nov 2017;26(11):2961-2971. doi:10.1007/s11136-017-1625-2
15. Jutten RJ, Peeters CFW, Leijdesdorff SMJ, et al. Detecting functional decline from normal aging to dementia: Development and validation of a short version of the Amsterdam IADL Questionnaire. Alzheimers Dement (Amst). 2017;8:26-35. doi:10.1016/j.dadm.2017.03.002

16. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. Alzheimers Dement. Nov 2012;8(6):536-43. doi:10.1016/j.jalz.2011.08.006
17. Zwan MD, van der Flier WM, Cleutjens S, et al. Dutch Brain Research Registry for study participant recruitment: Design and first results. Alzheimers Dement (N Y). 2021;7(1):e12132. doi:10.1002/trc2.12132
18. Koster N, Knol DL, Uitdehaag BM, Scheltens P, Sikkes SAM. The sensitivity to change over time of the Amsterdam IADL Questionnaire((c)). Alzheimers Dement. Oct 2015;11(10):1231-40. doi:10.1016/j.jalz.2014.10.006
19. Sikkes SAM, Knol DL, Pijnenburg YAL, 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(1):35-41. doi:10.1159/000346277
20. van der Flier WM, Scheltens P. Amsterdam Dementia Cohort: Performing Research to Optimize Care. J Alzheimers Dis. 2018;62(3):1091-1111. doi:10.3233/JAD-170850
21. Saan R, Deelman B. De 15-Woordentest A en B (Een Voorlopige Handleiding). Afdeling Neuropsychologie; 1986.
22. Lindeboom J, Schmand B, Tulner L, Walstra G, Jonker C. Visual association test to detect early dementia of the Alzheimer type. J Neurol Neurosurg Psychiatry. Aug 2002;73(2):126-33. doi:10.1136/jnnp.73.2.126
23. Tombaugh TN. Trail Making Test A and B: Normative Data Stratified by Age and Education. Archives of Clinical Neuropsychology. 2004;19:203-214. doi:10.1016/S0887-6177(03)00039-8
24. Wechsler D. WAIS-III Administration and Scoring Manual. The Psychological Corporation; 1997.
25. Van der Elst W, Van Boxtel MP, Van Breukelen GJ, Jolles J. Normative data for the Animal, Profession and Letter M Naming verbal fluency tests for Dutch speaking participants and the effects of age, education, and sex. J Int Neuropsychol Soc. Jan 2006;12(1):80-9. doi:10.1017/S1355617706060115
26. Stroop JR. Studies of interference in serial verbal reactions. J Exp Psychol Gen. Mar 1992;121(1):15.
27. van der Elst W, van Boxtel MP, van Breukelen GJ, Jolles J. The Letter Digit Substitution Test: normative data for 1,858 healthy participants aged 24-81 from the Maastricht Aging Study (MAAS): influence of age, education, and sex. J Clin Exp Neuropsychol. Aug 2006;28(6):998-1009. doi:10.1080/13803390591004428
28. Yesavage JA, Sheikh JI. Geriatric Depression Scale (GDS) - Recent evidence and development of a shorter version. Clinical Gerontologist. 1986;5(1-2):165-173. doi:10.1300/J018v05n01_09
29. Zarit SH, Reever KE, Bach-Peterson J. Relatives of the impaired elderly: correlates of feelings of burden. Gerontologist. Dec 1980;20(6):649-55. doi:10.1093/geront/20.6.649
30. Scheltens P, Launer LJ, Barkhof F, Weinstein HC, van Gool WA. Visual assessment of medial temporal lobe atrophy on magnetic resonance imaging: interobserver reliability. J Neurol. Sep 1995;242(9):557-60.
31. Koedam EL, Lehmann M, van der Flier WM, et al. Visual assessment of posterior atrophy development of a MRI rating scale. Eur Radiol. Dec 2011;21(12):2618-25. doi:10.1007/s00330-011-2205-4
32. Pasquier F, Leys D, Weerts JG, Mounier-Vehier F, Barkhof F, Scheltens P. Inter- and intraobserver reproducibility of cerebral atrophy assessment on MRI scans with hemispheric infarcts. Eur Neurol. 1996;36(5):268-72. doi:10.1159/000117270

33. Mulder C, Verwey NA, Van der Flier WM, et al. Amyloid-beta(1-42), total tau, and phosphorylated tau as cerebrospinal fluid biomarkers for the diagnosis of Alzheimer disease. Clin Chem. Feb 2010;56(2):248-53. doi:10.1373/clinchem.2009.130518

34. Tijms BM, Willemse EAJ, Zwan MD, et al. Unbiased approach to counteract upward drift in cerebrospinal fluid amyloid-β 1–42 analysis results. Clin Chem. Mar 2018;64(3):576-585. doi:10.1373/clinchem.2017.281055

35. R: A language and environment for statistical computing. Version 4.1.3. 2022. https://www.R-project.org/

36. Venables WN, Ripley BD. Modern Applied Statistics with S. Fourth ed. Springer; 2002.

37. Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. BMJ. Feb 28 1998;316(7132):690-3. doi:10.1136/bmj.316.7132.690

38. Rentz DM, Wessels AM, Annapragada AV, et al. Building clinically relevant outcomes across the Alzheimer's disease spectrum. Alzheimers Dement (N Y). 2021;7(1):e12181. doi:10.1002/trc2.12181

39. Wyrwich KW, Bullinger M, Aaronson N, et al. Estimating clinically significant differences in quality of life outcomes. Qual Life Res. Mar 2005;14(2):285-95. doi:10.1007/s11136-004-0705-2

40. Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. J Clin Epidemiol. Jul 2010;63(7):760-766 e1. doi:10.1016/j.jclinepi.2009.09.009

41. Andrew MK, Rockwood K. A five-point change in Modified Mini-Mental State Examination was clinically meaningful in community-dwelling elderly people. J Clin Epidemiol. Aug 2008;61(8):827-31. doi:10.1016/j.jclinepi.2007.10.022

42. Schrag A, Schott JM, Alzheimer's Disease Neuroimaging I. What is the clinically relevant change on the ADAS-Cog? J Neurol Neurosurg Psychiatry. Feb 2012;83(2):171-3. doi:10.1136/jnnp-2011-300881

43. Andrews JS, Desai U, Kirson NY, Zichlin ML, Ball DE, Matthews BR. Disease severity and minimal clinically important differences in clinical outcome assessments for Alzheimer's disease clinical trials. Alzheimers Dement (N Y). 2019;5:354-363. doi:10.1016/j.trci.2019.06.005

44. Insel PS, Weiner M, Mackin RS, et al. Determining clinically meaningful decline in preclinical Alzheimer disease. Neurology. Jul 23 2019;93(4):e322-e333. doi:10.1212/WNL.0000000000007831

45. Jutten RJ, Dicks E, Vermaat L, et al. Impairment in complex activities of daily living is related to neurodegeneration in Alzheimer's disease-specific regions. Neurobiol Aging. Nov 26 2018;75:109-116. doi:10.1016/j.neurobiolaging.2018.11.018

46. Lilamand M, Cesari M, Cantet C, et al. Relationship Between Brain Amyloid Deposition and Instrumental Activities of Daily Living in Older Adults: A Longitudinal Study from the Multidomain Alzheimer Prevention Trial. J Am Geriatr Soc. Oct 2018;66(10):1940-1947. doi:10.1111/jgs.15497

47. Marshall GA, Rentz DM, Frey MT, et al. Executive function and instrumental activities of daily living in mild cognitive impairment and Alzheimer's disease. Alzheimers Dement. May 2011;7(3):300-8. doi:10.1016/j.jalz.2010.04.005

8.2.5 Supplementary Material

Vignette creation

The Amsterdam Instrumental Activities of Daily Living Questionnaire (A-IADL-Q)[1,2] is scored using item response theory (IRT). In IRT, it is assumed that the instrument measures a single construct that cannot be directly observed.[3] Each of the items of the instrument have their own level of information and this information is contained in so-called item parameters. A mathematical formula is used to compute the total score from the item responses, and it is considered more precise than classical test theory, where item scores are summed or averaged to obtain a total score. The total score is represented by theta, which is normally distributed around a mean of 0 with a standard deviation of 1 in the population of interest. For the A-IADL-Q, the theta is linearly transformed into a *T*-score, which has a mean of 50 and a standard deviation of 10. Scores range approximately ±3 standard deviations from the mean, i.e., from approximately 20 to 80. The construct that is measured with the A-IADL-Q is 'IADL functioning', with lower scores representing more impaired IADL functioning.

Using the item parameters, it is possible to obtain the most likely responses to each of the items given a certain total score. Using an R script adapted from Morgan et al.[4] and the item parameters estimated for the A-IADL-Q (see Jutten et al.[1]), we obtained the most likely responses to all items of the A-IADL-Q for total scores along the entire range of the scale, between 24 (all most likely item responses are the most impaired category) and 60 (all most likely item responses are the least impaired category), in increments of 1 point.

Subsequently, we selected 15 out of the 30 items from the A-IADL-Q that had item information covering the entire spectrum of IADL functioning. We then assessed how many unique vignettes could be made from the selected items while maintaining a distinguishability and equal distance between the vignettes. As such, we arrived at a total of 18 unique vignettes in the range of 24 to 60, with each vignette placed 2 points (1/5 standard deviation) apart.

We created six groups and randomly assigned respondents to one of the groups. Each group had its own reference vignette, and the reference vignettes were spread across the range of the scale. Respondents viewed seven situations, three showing improvement (the *T*-score of the comparison vignette was higher) and

four showing decline (the *T*-score of the comparison vignette was lower). The vignettes shown to each respondent were proximal to the reference vignette. Supplemental Table 1 shows the T-scores associated with all vignettes that were presented to the respondents in each group.

Supplemental Table 2 shows how many respondents reached the MIC threshold for each of the seven presented situations. Three of the seven situations showed an increase in score, four showed a decrease of the score. Displayed in Supplemental Table 2 are the within-range judgments, that is, the first situation each respondent indicated they felt represented an important change in the right direction.

Supplemental Table 1 | *T*-scores of vignettes, per group.

	Reference	Improve 3	Improve 2	Improve 1	Decline 1	Decline 2	Decline 3	Decline 4
'Ms. Smith'	54	60	58	56	52	50	48	46
'Mr. Jones'	50	56	54	52	48	46	44	42
'Mr. Williams'	46	52	50	48	44	42	40	38
'Ms. Brown'	42	48	46	44	40	38	36	34
'Ms. Johnson'	38	44	42	40	36	34	32[a]	28
'Mr. Garcia'	34	40	38	36	32[a]	28	26	24

[a] The most likely responses for T-scores of 32 and 30 were exactly the same for the items selected to be included in the vignette creation.

Supplemental Table 2 | Number of respondents who reached the threshold at each vignette.

	Total n	Situation 1	Situation 2	Situation 3	Situation 4 *(decline only)*	Threshold not reached
Improvement		+2	+4	+6		
'Ms. Smith'	268	6 (2.2%)	45 (16.8%)	50 (18.7%)	—	167 (37.8%)
'Mr. Jones'	260	10 (3.8%)	57 (21.9%)	74 (28.5%)	—	119 (45.8%)
'Mr. Williams'	284	1 (0.4%)	1 (0.4%)	8 (2.8%)	—	274 (96.5%)
'Ms. Brown'	265	8 (3.0%)	48 (18.1%)	9 (3.4%)	—	200 (75.5%)
'Ms. Johnson'	282	5 (1.8%)	16 (5.7%)	8 (2.8%)	—	253 (89.7%)
'Mr. Garcia'	270	1 (0.4%)	11 (4.1%)	4 (1.5%)	—	254 (94.1%)
Decline		-2	-4	-6	-8	
'Ms. Smith'	268	283 (88.8%)	19 (7.1%)	4 (1.5%)	4 (1.5%)	3 (1.1%)
'Mr. Jones'	260	187 (71.9%)	66 (25.4%)	3 (1.2%)	1 (0.4%)	3 (1.2%)
'Mr. Williams'	284	274 (96.5%)	9 (3.2%)	0 (0.0%)	0 (0.0%)	1 (0.4%)
'Ms. Brown'	265	239 (90.2%)	19 (7.2%)	4 (1.5%)	1 (0.4%)	2 (0.8%)
'Ms. Johnson'	282	249 (88.3%)	20 (7.1%)	6 (2.1%)	3 (1.1%)	4 (1.4%)
'Mr. Garcia'	270	182 (67.4%)	47 (17.4%)	12 (4.4%)	12 (4.4%)	17 (6.3%)

MIC in clinical practice

Supplemental Table 3 shows the number of patients who reached the MIC thresholds for decline and for improvement at each visit, stratified by diagnostic group.

Supplemental Table 3 | Breakdown of patients who reach the MIC thresholds at each visit, stratified by diagnostic group.

Follow-up wave	N (%)	Meaningful decline	Meaningful improvement	No meaningful change
3 months	**157 (68.3)**	**55 (35.0)**	**23 (14.6)**	**79 (50.3)**
SCD	18 (48.6)	3 (16.7)	4 (22.2)	11 (61.1)
MCI	18 (81.8)	8 (44.4)	5 (27.8)	5 (27.8)
AD dementia	56 (69.1)	17 (30.4)	4 (7.1)	35 (62.5)
Non-AD dementia	43 (74.1)	17 (39.5)	7 (16.3)	19 (44.2)
Other	26 (72.2)	11 (42.3)	3 (11.5)	12 (46.2)
6 months	**121 (52.6)**	**55 (45.5)**	**23 (19.0)**	**43 (35.5)**
SCD	13 (35.1)	3 (23.1)	6 (46.2)	4 (30.8)
MCI	9 (40.9)	2 (22.2)	5 (55.6)	2 (22.2)
AD dementia	51 (63.0)	25 (49.0)	6 (11.8)	20 (39.2)
Non-AD dementia	34 (58.6)	17 (50.0)	5 (14.7)	12 (35.3)
Other	14 (38.9)	8 (57.1)	1 (7.1)	5 (35.7)
9 months	**102 (44.3)**	**50 (49.0)**	**15 (14.7)**	**37 (36.3)**
SCD	10 (27.0)	0 (0.0)	4 (40.0)	6 (60.0)
MCI	9 (40.9)	2 (22.2)	1 (11.1)	6 (66.7)
AD dementia	42 (51.9)	24 (57.1)	6 (14.3)	12 (28.6)
Non-AD dementia	28 (48.3)	16 (57.1)	1 (3.6)	11 (39.3)
Other	13 (36.1)	8 (61.5)	3 (23.1)	2 (15.4)

Follow-up wave	N (%)	Meaningful decline	Meaningful improvement	No meaningful change
12 months	**88 (38.3)**	**48 (54.5)**	**9 (10.2)**	**31 (35.2)**
SCD	16 (43.2)	6 (37.5)	3 (18.8)	7 (43.8)
MCI	7 (31.8)	3 (42.9)	1 (14.3)	3 (42.9)
AD dementia	32 (39.5)	23 (71.9)	2 (6.3)	7 (21.9)
Non-AD dementia	22 (37.9)	13 (59.1)	1 (4.5)	8 (36.4)
Other	11 (30.6)	3 (27.3)	2 (18.2)	6 (54.5)
Last completed visit	**230 (100.0)**	**104 (45.2)**	**36 (15.7)**	**90 (39.1)**
SCD	37 (100.0)	7 (18.9)	9 (24.3)	21 (56.8)
MCI	22 (100.0)	9 (40.9)	5 (22.7)	8 (36.4)
AD dementia	81 (100.0)	43 (53.1)	9 (11.1)	29 (35.8)
Non-AD dementia	58 (100.0)	34 (58.6)	4 (6.9)	20 (34.5)
Other	36 (100.0)	15 (41.7)	9 (25.0)	12 (33.3)

Abbreviations: AD, Alzheimer's disease; MCI, mild cognitive impairment; SCD, subjective cognitive decline

Factors associated with surpassing the MIC threshold

Supplemental Table 4 shows the odds ratios (OR), 95% confidence intervals (95%CI) and p-values for the simple multinomial logistic regressions associating various factors with the MIC thresholds.

Supplemental Table 4 | Odds ratios, 95% confidence intervals and p-values for simple multinomial logistic regressions.

Factor	MIC for decline		MIC for improvement	
	OR [95%CI]	P	OR [95%CI]	P
Age in years[1]	1.01 [0.98, 1.05]	0.528	0.95 [0.91, 1.01]	0.076
Female sex[1]	1.78 [0.99, 3.18]	0.053	1.05 [0.45, 2.46]	0.907
Education in years[1]	0.98 [0.89, 1.09]	0.743	0.97 [0.84, 1.12]	0.666
Syndrome diagnosis[1,2]	2.48 [0.74, 8.31]	0.142	0.98 [0.27, 3.61]	0.978
MCI	3.41 [1.49, 7.82]	0.004	0.35 [0.14, 0.90]	0.029
Dementia				
MMSE[3]	0.95 [0.88, 1.02]	0.185	1.00 [0.89, 1.13]	0.996
GDS[3]	0.97 [0.88, 1.07]	0.538	1.00 [0.88, 1.15]	0.950
ZBI[3]	1.01 [0.98, 1.05]	0.462	0.89 [0.82, 0.97]	0.009
APOE ε4 carrier[3]	0.98 [0.62, 1.55]	0.942	1.25 [0.66, 2.38]	0.491
Aβ positive[3]	1.39 [0.63, 3.04]	0.414	0.98 [0.32, 3.02]	0.970
P-tau/Aβ ratio[3]	0.76 [0.01, 93.49]	0.912	0.10 [0.00, 604.58]	0.599
Episodic memory z-score[3]	0.88 [0.58, 1.32]	0.530	1.02 [0.54, 1.93]	0.960
Executive functioning z-score[3]	0.75 [0.45, 1.24]	0.259	2.35 [0.95, 5.81]	0.064

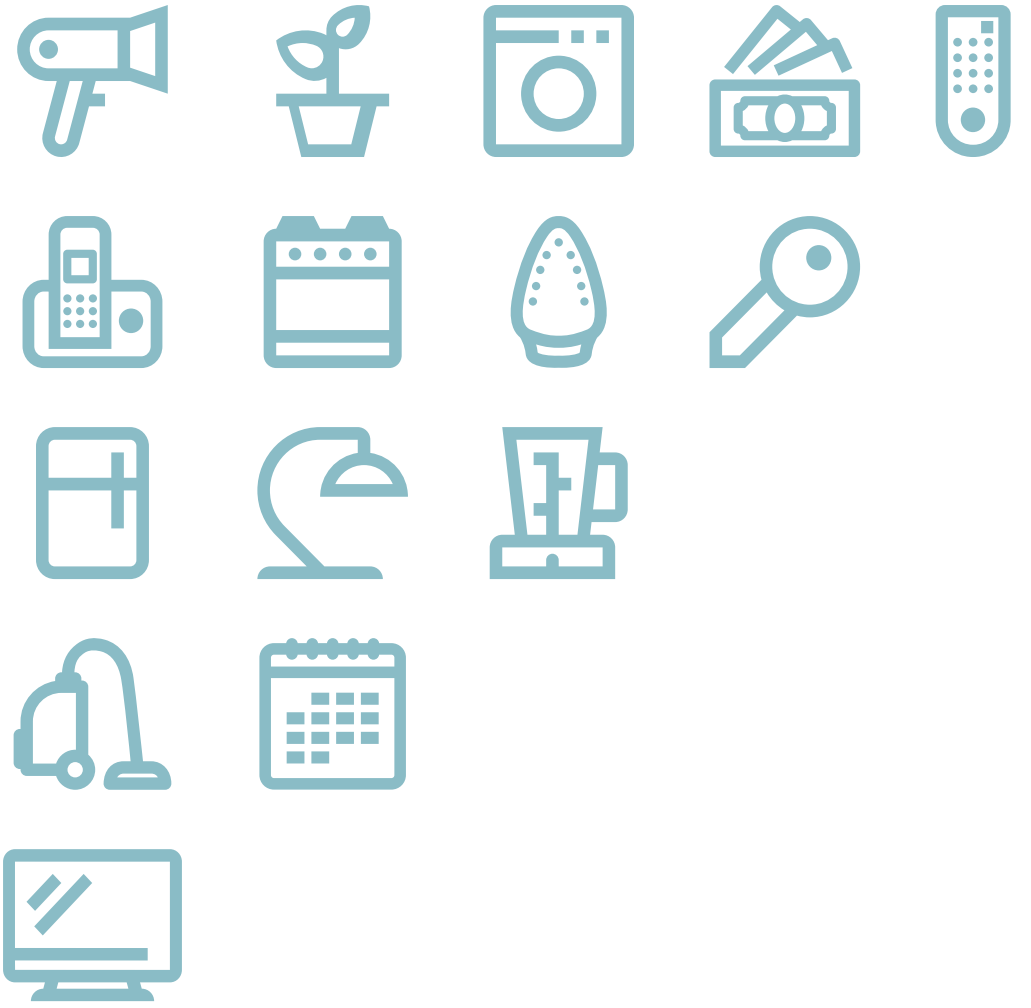
Factor	MIC for decline		MIC for improvement	
	OR [95%CI]	P	OR [95%CI]	P
Attention/speed z-score[3]	0.93 [0.63, 1.37]	0.720	1.50 [0.75, 3.00]	0.258
Language z-score[3]	0.88 [0.57, 1.35]	0.548	0.99 [0.51, 1.92]	0.985
MTA[3,4]	1.80 [0.66, 4.89]	0.248	1.21 [0.34, 4.36]	0.771
1	4.97 [1.23, 19.99]	0.024	0.72 [0.07, 7.92]	0.788
1.5	1.78 [0.61, 5.19]	0.293	0.47 [0.08, 2.79]	0.403
2+				
PA[3,5]	0.83 [0.29, 2.39]	0.725	0.40 [0.11, 1.42]	0.156
1	0.74 [0.23, 2.39]	0.610	0.40 [0.09, 1.78]	0.227
2				
Fazekas[3,6]	2.16 [0.85, 5.47]	0.105	0.40 [0.12, 1.34]	0.137
1	2.64 [0.77, 9.06]	0.124	1.20 [0.27, 5.39]	0.809
2				
GCA[3]	1.15 [0.42, 3.19]	0.783	0.56 [0.15, 2.01]	0.370
Microbleeds[3]	0.90 [0.32, 2.51]	0.838	1.02 [0.26, 4.06]	0.976

[1] Single model; [2] reference category: subjective cognitive decline; [3] OR adjusted for age, sex, education, and syndrome diagnosis; [4] reference category: MTA ≤ 0.5; [5] reference category: PA = 0; [6] reference category: Fazekas = 0.

Abbreviations: APOE, apolipoprotein; Aβ, amyloid beta; GCA, global cortical atrophy; GDS, Geriatric Depression Scale; MCI, mild cognitive impairment; MIC, minimal important change; MMSE, Mini-Mental State Examination; OR, odds ratio; PA, posterior atrophy; p-tau, phosphorylated tau; ZBI, Zarit Burden Interview; 95%CI, 95% confidence interval.

References

1. Jutten RJ, Peeters CFW, Leijdesdorff SMJ, et al. Detecting functional decline from normal aging to dementia: Development and validation of a short version of the Amsterdam IADL Questionnaire. Alzheimers Dement (Amst). 2017;8:26-35. doi:10.1016/j.dadm.2017.03.002
2. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. Alzheimers Dement. Nov 2012;8(6):536-43. doi:10.1016/j.jalz.2011.08.006
3. Reise SP, Waller NG. Item response theory and clinical measurement. Annu Rev Clin Psychol. 2009;5:27-48. doi:10.1146/annurev.clinpsy.032408.153553
4. Morgan EM, Mara CA, Huang B, et al. Establishing clinical meaning and defining important differences for Patient-Reported Outcomes Measurement Information System (PROMIS((R))) measures in juvenile idiopathic arthritis using standard setting with patients, parents, and providers. Qual Life Res. Mar 2017;26(3):565-586. doi:10.1007/s11136-016-1468-2



CHAPTER 9

SUMMARY AND GENERAL DISCUSSION

9.1 Summary of findings

In this thesis, we aimed to extend the knowledge about the measurement of changes in everyday functioning in the context of Alzheimer's disease and related disorders. This may help clinicians and researchers identify when changes in everyday functioning occur and when they matter, as well as benefit the monitoring of disease progression and evaluation of potential treatment effects.

In this final chapter, we will summarize the key findings of this thesis and place them in the context of current literature. We also make suggestions for future research.

Self- vs. study partner-report

In chapters 2 and 3 we described how cognitively normal participants self-reported their everyday functioning and compared it to how their study partners reported the participants' everyday functioning. With this, we aimed to capture a form of disease (or impairment) awareness in individuals who may be on the trajectory to developing Alzheimer's disease or a related disorder, but who do not yet have any overt clinical symptoms. In **chapter 2**, we invited a large sample of cognitively normal adults and asked them to rate their own level of everyday functioning on the Amsterdam IADL Questionnaire. We also invited a study partner to complete the same questionnaire. Unsurprisingly, we found that, at the group level, the majority of both participants and study partners reported a high level of everyday functioning, indicating that they experienced no impairment. Interestingly, when we looked at the level of agreement between participants and study partners, we found that it was only moderate, i.e., that participants and study partners reported differently on the participants' functioning. Subsequently, we divided our sample into three groups to distinguish between participants who reported a similar level of functioning as their study partner, participants who reported a lower level of functioning and participants who reported a higher level of functioning than their study partner. Approximately one quarter of participants reported lower functioning while about one fifth reported better functioning. We then investigated several demographic, cognitive and psychosocial factors in relation to these groups and found that participants reported lower levels of functioning (i.e., more difficulty) than their study partners in the presence of depressive symptoms and memory complaints. From these findings, we concluded that self- and study partner-report measures of everyday functioning may not be interchangeable, and that the level

of awareness of a participant needs to be considered even in cognitively normal older adults. In **chapter 3**, we investigated both the participant self- and study partner-report of everyday functioning among cognitively normal older adults in relation to cerebral tau burden, a marker of Alzheimer's disease pathology. Participants underwent tau and amyloid positron emission tomography scans and completed the Alzheimer's Disease Cooperative Study Activities of Daily Living – Prevention Instrument. We found that tau in two brain regions that are relevant for Alzheimer's disease was related to study partner-reported difficulties in everyday functioning, but not to self-reported difficulties. We also found that participants were more likely to self-report more difficulties than their study partners when tau burden was greater. We concluded that these findings may hint at an altered awareness of functional changes among cognitively normal older adults with underlying Alzheimer's disease pathology.

Biases in the measurement of everyday functioning

The report of difficulties with everyday functioning is not only determined by the cognitive status of a participant. Many other factors may affect how everyday functioning is reported and some of these may lead to inaccurate or biased reports of participants' or patients' levels of everyday functioning. We investigated this in **chapter 4**. First, in **chapter 4.1**, we combined data from studies using the Amsterdam IADL Questionnaire that were performed in eight countries in Europe and North America. The country in which the study was performed was then investigated as a source of systematic bias, based on the observation that certain everyday activities are country specific. For example, in the United States, using checks and balancing a checkbook is a normal part of daily life, whereas this is much less common in many European countries. In the Netherlands, people eat sliced bread for lunch, whereas in France, Spain, and many other countries, it is customary to have a hot meal. From this example, it is apparent that activities of daily living may have a different relevance for people from different countries. Furthermore, it is conceivable that a participants' country of residence may affect how they report difficulties with certain activities. The Amsterdam IADL Questionnaire is scored using item response theory, which models the relationship between a person's ability on a construct and the likelihood of their endorsement of a given response category on an item measuring the ability in question. In the case of the Amsterdam IADL Questionnaire, the ability being measured is daily functioning. The information from the item parameters enables us to analyze systematic differences in the relationship between a person's ability

and the endorsement of response categories on the items of the questionnaire, in an analysis called 'differential item functioning'. We analyzed differential item functioning for the country where the participant was from, as well as for age, sex, and education, and found no evidence for systematic bias by any of these factors, suggesting that the questionnaire is not biased by country, age, sex, or education. From this, we concluded that the Amsterdam IADL Questionnaire may be used across countries, age ranges, sexes, and education level. One important caveat, however, is that the participants included in this study were all from Western countries, which still share a lot of common cultural features. As data become available from other regions of the world, a reiteration of this work is necessary to investigate whether the systematic bias is not present in countries that are culturally more different from the West. In **chapter 4.2**, we investigated the influence of differences between German-speaking Switzerland and the Netherlands. In addition, we analyzed test–retest reliability over two to four weeks in cognitively normal participants and participants with mild cognitive impairment and mild dementia and performed construct validation. The results demonstrated no systematic bias in the Swiss adaptation, showed a good test–retest reliability and confirmed the construct validity of the questionnaire as previously established.

Scoring techniques

Not only the outcome measure itself but potentially also the type of scoring may influence how everyday functioning is measured. In **chapter 5** we compared and aligned two methods for scoring the Amsterdam IADL Questionnaire: item response theory and classical test theory scoring. Both scoring methods are currently used: item response theory in research and classical test theory in clinical practice. Differences between these scoring methods include the scale, the underlying theory, and the preciseness of measurement at extreme levels of the construct that is being measured. Because it is impractical to have two scales in use, we aimed to align the classical test theory-based scores with the item response theory-based scores. To justify this alignment, we first examined to what extent the two types of scorings were comparable. In a sample of predominantly cognitively normal individuals, we scored the Amsterdam IADL Questionnaire using both techniques, and investigated the relationship between demographic and psychosocial factors and the two scores, as well as the changes in the scores over time. While we found no major differences between the scoring techniques, item response theory-based scores do have the advantage that they account

for item difficulty and the likelihood of someone endorsing a response option given their overall level of functioning. Another advantage of the item response theory-based scores was that they showed a substantially smaller ceiling effect than the classical test theory-based scores, which implies item response theory is a better option for measuring the extreme levels of the construct of everyday functioning. After establishing that the two scoring techniques were sufficiently similar, the second step we took in this chapter was to align classical test theory-based scores and item response theory-based scores and place them on the same scale. We simulated a large number of responses to the Amsterdam IADL Questionnaire, calculated both types of scores, and then created a alignment (or 'cross-walk') table where each unique classical test theory-based score was linked to the item response theory-based scale. With the alignment of the two scoring methods, classical test theory-based scores can be placed on the same scale as item response theory-based scores, allowing comparison of results, eliminating the need for separate cutoffs and interpretation guidelines, and facilitating use in clinical practice. We recommended item response theory be used to calculate the total scores for the Amsterdam IADL Questionnaire, and that when this type of scoring is not available or feasible, the cross-walk table is used to convert classical test theory-based scores to the item response theory-based *T*-score scale.

Measuring changes over time in everyday functioning

Measuring everyday functioning at a single point in time can only provide so much information. Understanding how the performance of everyday activities changes over time across diseases and disease stages helps monitor disease progression and can help inform clinicians and patients about imminent changes in functioning. In **chapter 6** we endeavored to obtain a better understanding of how everyday functioning changes along the Alzheimer's disease continuum. The National Institute on Aging and Alzheimer's Association presented a new research framework in 2018 that defines Alzheimer's disease biologically.[1] Anyone with abnormal accumulation of the amyloid-beta protein is considered to be on the Alzheimer's continuum. The research framework also defined clinical stages, representing asymptomatic (preclinical), early symptomatic (late preclinical) and prodromal Alzheimer's disease, followed by increasingly severe stages of Alzheimer's disease dementia. In our investigation, we focused mainly on the first three stages, and merged the dementia stages. We pooled data from several cohorts in the United States and The Netherlands, combining various everyday functioning outcome measures into

one. Compared to healthy controls with no amyloid or symptoms, individuals who were in the early symptomatic stage of Alzheimer's disease showed a faster decline in everyday functioning. The predominant assumption currently is that functional impairment marks the transition from the prodromal to the dementia stage, but our findings imply that changes in daily functioning occur earlier than that. We also found that the rate of decline in everyday functioning increased with each successive clinical stage, such that those who were in the dementia stage declined faster, on average, than those in the prodromal stage, and that those in the prodromal stage declined faster than those in the early symptomatic stage.

In addition to overall everyday functioning, we also investigated the performance over time of two specific activities, namely preparing a hot beverage and managing the paperwork. From the Amsterdam IADL Questionnaire item parameters, we construed that preparing a hot beverage is a relatively easy activity, i.e., one that an individual remains capable of doing until a greater overall level of functional impairment is reached. Conversely, managing the paperwork is an activity that people may find more difficult to perform early on, when overall impairment is more limited still. We found that the more difficult item showed a stronger decrease in performance, even in early disease stages. In later disease stages, the easier item showed more change. While this seems logical, it is important to guarantee that an instrument covers both the easier and the more difficult ends of the spectrum of everyday functioning.

Rates of change in everyday functioning among memory clinic patients are highly heterogeneous. This heterogeneity exists both within and across diagnostic groups, as well as between and within disease stages. Some patients show very little progression of impairment in everyday functioning, whereas other patients decline rapidly. Using a sample of memory clinic patients with various diagnoses, in **chapter 7**, we aimed to characterize the heterogeneity in change over time in everyday functioning. Using a longitudinal latent class analysis, we attempted to identify groups of patients who showed a similar change in everyday functioning over time. The analysis showed that there were seven classes of patients, but when we analyzed the differences in rates of change between the classes, we distinguished three overarching groups: patients who remained stable, patients who showed a slow decline and patients who showed a steep decline in everyday functioning. We subsequently investigated what baseline characteristics, including demographics, psychosocial factors, cognitive functioning, brain imaging and Alzheimer's disease biomarkers, were associated with the groups we identified. We found that patients

with subjective cognitive decline were more likely to remain stable, whereas patients with Alzheimer's disease dementia or dementia with Lewy bodies were most likely to show a rapid decline in everyday functioning. Those who performed worse on various cognitive domains also were more likely to experience a rapid decline in everyday functioning. Other baseline characteristics, such as magnetic resonance imaging, Alzheimer's disease biomarkers and family history may also provide useful information for making an accurate prognosis for future functional decline.

Determining when change impacts daily life

It has thus been established that changes in everyday functioning emerge in early disease stages and that some individuals show more decline than others. A vital question, however, remains: when do changes in everyday functioning impact daily life? We aimed to answer that question in **chapter 8**. First, in **chapter 8.1**, we described a new qualitative method for establishing meaningful scoring categories for the Amsterdam IADL Questionnaire. The scoring categories were predetermined to represent no, mild, moderate, and severe problems in everyday functioning. We organized focus groups with informal caregivers of dementia patients and asked them to rank nine short clinical summaries of fictional patients from least to most impairment in everyday functioning. The summaries were created using the item response theory parameters for a selection of items from the Amsterdam IADL Questionnaire. From these parameters, we derived the most likely response categories for items at different levels of overall functioning. Each summary was assigned a random name, sex, and age, to make it seem like a fictional patient. After sorting the summaries from least to most impaired, participants placed bookmarks between them to demarcate the thresholds between no and mild, mild and moderate, and moderate and severe problems. Individual sorting and bookmark placements were then discussed to achieve consensus. Clinicians working in various memory clinics in The Netherlands completed an online questionnaire where they too placed bookmarks to create the same scoring categories. While we did find some individual variation in categorizations, caregivers and clinicians generally agreed on the thresholds, particularly about the distinction between 'no' an 'mild' problems. This important finding suggests that the emergence of mild problems may be noticed simultaneously by both caregivers and clinicians. With these substantive scoring categories, the interpretation of Amsterdam IADL Questionnaire scores may be facilitated, as an intuitive label with a clear definition may be attached to any given score. The categories can also help clinicians to improve communication with patients and caregivers.

In **chapter 8.2**, we embarked to establish the minimal important change in everyday functioning: the smallest amount of change that has a meaningful impact on patients' lives. This study comprised two parts. First, in the qualitative part of the study, we invited informal caregivers of Alzheimer's disease and related disorders patients from the community to participate in a survey in which they judged various situations representing changes in everyday functioning. These situations were formed by presenting two summaries side-by-side: one reference summary that remains the same, and one comparison summary that changes in each situation. The difference in scores between the two summaries varied. For each situation, the respondent was asked to indicate whether they thought there was a change in the overall level of everyday functioning. If they thought there was, they were then asked whether they thought the change would make an important difference or not. We placed thresholds at the situation with the smallest score difference for the minimally important change. Second, in the quantitative part of the study, we followed a group of unselected memory clinic patients for a year. We determined that approximately half of them showed a decline that was larger than the minimally important change. Patients with a dementia diagnosis and more atrophy of the medial temporal lobe had a larger risk of passing the minimally important change threshold than patients with subjective cognitive decline and no atrophy. We concluded that we were able to operationalize clinically meaningful decline in everyday functioning and that the usefulness of our thresholds for minimally important change was supported by the findings from the quantitative part of the study. The findings of this chapter provide guidance in interpreting changes in everyday functioning and may help evaluate treatment effects as well as monitor disease progression.

9.2 General discussion

The concept of 'instrumental activities of daily living' (IADL) was first introduced in 1969, when Lawton and Brody published their IADL scale.[2] In the following decades, the measurement of IADL in the context of dementia became widespread, yet a clear definition of the concept lacked.[3,4] In 2014, Sikkes and De Rotrou defined IADL as 'intentional and complex everyday activities for which multiple cognitive processes are necessary, particularly high-level controlled processes'.[3] The goal of this thesis was to extend our knowledge about IADL in Alzheimer's disease and related disorders, and to improve our understanding of when changes emerge, in what ways they can be measured, and what they mean for the patient and their caregiver.

The majority of the work discussed here made use of the Amsterdam IADL Questionnaire, which is a relatively new measure of everyday functioning, published ten years ago.[5] Previous work on the Amsterdam IADL Questionnaire has shown that the questionnaire has good content and construct validity,[6] that it has diagnostic accuracy (criterion validity),[7] good test–retest reliability,[5] and that it is responsive to changes over time.[8] These psychometric qualities were all deemed relevant by the consensus-based standards for the selection of health measurement instruments (COSMIN) initiative.[9] The work described in this dissertation adds important information on possible measurement error (**chapter 4**), the meaning of scores (**chapter 8.1**) as well as meaning of decline in scores (**chapter 8.2**) and reinforced evidence on internal consistency (**chapter 4**), test–retest reliability (**chapter 4.2**) and responsiveness (chapters 5, 6 and 7). Figure 1 shows what is now known about the psychometric quality of the Amsterdam IADL Questionnaire.

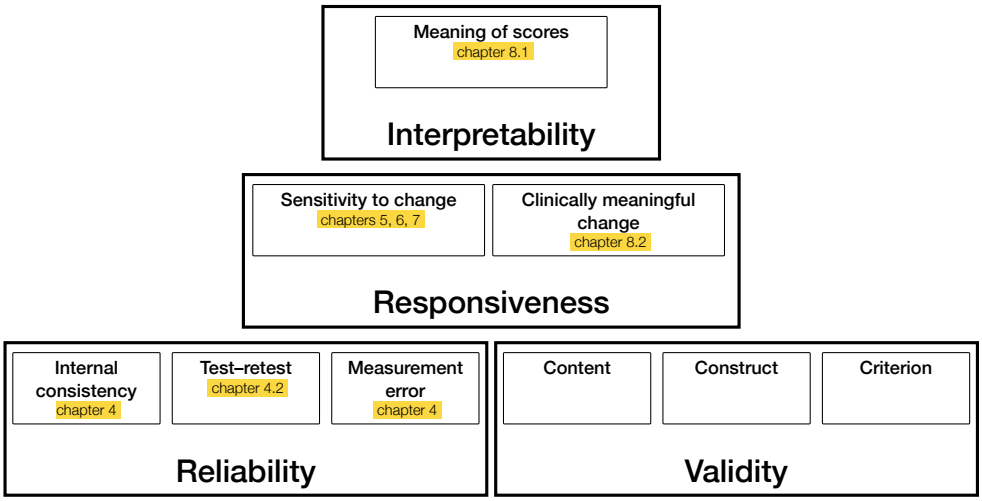


Figure 1 | Investigated psychometric properties of the Amsterdam IADL Questionnaire. Highlighted in yellow are the contributions of the present work to what is known about the psychometric properties of the Amsterdam IADL Questionnaire.

For clinical practice as well as for trials, it is pertinent that psychometrically strong outcome measures be used. This may be even more important when cognitive and functional symptoms are minimal, for which outcome measures need to be able to detect more subtle impairments. With the focus on early disease stages,[10] this seems especially relevant for clinical trials. In 2021, there were 152 trials investigating

disease-modifying drugs for Alzheimer's disease;[11] a substantial increase from the 36 trials five years prior.[12] Most trials rely on primary outcome measures with a limited score range and proven unsatisfactory psychometric qualities, such as the Alzheimer's Disease Assessment Scale–Cognitive Subscale[13,14] (for cognition) and the Clinical Dementia Rating[15] (for functional impairment). Many trials fail to show clinical benefits,[16] which may be attributable to the intervention under investigation. However, it is difficult to rule out the possibility that the null findings, at least partly, stem from the limited sensitivity of the most commonly used outcome measures to more subtle changes in function and cognition.[17,18] In fact, heterogeneity in disease progression may obfuscate real treatment effects, as the results of anti-amyloid trials so far have mostly fallen within the ranges of change due to random variation.[18]

The findings presented in this thesis transcend the Amsterdam IADL Questionnaire, in the sense that they not only provide evidence for the psychometric quality of the Amsterdam IADL Questionnaire, but that they can also inform us more generally about the construct of IADL, or everyday functioning, as a whole. Below, we will discuss three major topics our work contributed to: the ways in which everyday functioning can be measured, when changes in everyday functioning occur, and when changes are meaningful for patients and their caregivers.

The ways to measure everyday functioning

A matter of who you ask

Because dementia is characterized by a diminished disease insight, most questionnaires rely on collateral report by someone who is close to the patient, such as a partner or adult child, also called a study partner. Indeed, study partner report of cognitive problems has been related to progression to mild cognitive impairment and dementia among those who were cognitively normal.[19] Because disease insight is supposedly intact in early disease stages,[20] researchers have investigated the value of self-report by the subject.[21-23] In chapters 2 and 3 we evaluated self- and study partner-report in two separate cohorts of cognitively normal adults. A potential problem with relying on a single source is that subjective report, either by the participant themselves or by their study partner, may be affected by factors such as mood and anxiety, as well as caregiver stress.[24] These personal factors may affect the actual functioning, according to the ICF model,[25] but may also influence the report of functioning. Our findings showed that, even though the participants

were unimpaired and thus assumed to have an intact insight into their cognitive and functional performance, there were differences between participants and their study partners. This is in line with previous findings:[26-29] it seems that information from both the person in question and someone close to them on the report of cognition in everyday life has added value over the use of either one of those sources independently. Discrepancies between self- and study partner-report have previously been suggested as an indicator of subjective cognitive decline and mild cognitive impairment.[29] As such, participant and study partner differences in the report of everyday functioning constitute promising avenue for future research.

A matter of what you ask

There is a large diversity among people with dementia in terms of their age, sex, educational attainment, cultural background and race and ethnicity. With that, it seems natural that a one-size outcome measure might not fit all.[30] For years, researchers have called for a better inclusion of underrepresented groups.[31,32] There is a need for evidence that an outcome measure accurately reflects the construct it intends to measure, regardless of the patient's individual characteristics. In fact, when Lawton and Brody developed the first scale for measuring instrumental activities of daily living in 1969, they included different activities for men and women, because "for women, the maintenance of earlier life levels of adequacy in such tasks as shopping, cooking, and manner of doing laundry may be the best means of assessing general competence. While the list of such representative activities is smaller for men, one can still differentiate their performance of tasks such as use of transportation, or handling money, as the basis for measuring competence".[2] While this separation now sounds dated, it is still conceivable that sex differences, as well as age, education level and the country in which one lives, affect the relationship between the overall level of functioning and the amount of difficulties with specific activities. Indeed, previous work with other functional outcome measures has shown that there are differences between ages, sexes and cultures.[33,34] Culture-adjusted versions of functional outcome measures are thus warranted, and it is therefore of importance to use a thorough translation and cross-cultural adaptation process such as the one proposed by Beaton and colleagues,[35] in which good care is taken to ensure that the adapted items accurately reflects the original items, while also being comprehensible and relevant in the new setting. The work we present in **chapter 4** suggests that bias induced by cross-cultural adaptation in the Amsterdam IADL Questionnaire was limited, which may be attributable to

the elaborate cross-cultural adaptation process we employed when translating the questionnaire. Although we did not find any substantial bias caused by the various factors we investigated, we cannot assume that an instrument developed in one country can be used across the world without potentially needing adaptations. This is especially the case for measures of everyday functioning because the activities of daily living differ from country to country. Even within areas that are generally considered to have 'one culture', like Western Europe, there are differences in customs between countries. An anecdotal example of this is that it is customary to eat hot meals for lunch in countries like France and Spain, while people living in The Netherlands more often eat bread. This demonstrates that differences in daily practices between countries and cultures are omnipresent, and that they should be in some way accounted for when cross-culturally adapting a questionnaire.

A matter of how you score

Finally, we expected that item response theory would enable more precise measurement of everyday functioning, especially at the extremes of the scale, that is, when there is severe impairment or very subtle impairment.[36] Floor and ceiling effects are often found in functional outcome measures,[37] hampering adequate assessment of very mild or very severe impairment. In **chapter 5** we discovered that item response theory scores only marginally outperformed classical test theory scores in the detection of subtle changes in everyday functioning, specifically in individuals who were cognitively normal. Other studies have suggested that item response theory should be used when investigating change over time, as the measurement is more precise and therefore better able to reliably show subtle changes.[38] Item response theory-based scores also had a notably smaller ceiling effect than classical test theory-based scores, which argues for the use of item response theory scoring when assessing the extreme abilities of everyday functioning. In later disease stages, it appeared to matter less which scoring method was used. Indeed, it is an established fact that item response theory and classical test theory scores of the same instrument correlate very highly,[36] which was also the case for the Amsterdam IADL Questionnaire. A major advantage of the use of item response theory in the development of the Amsterdam IADL Questionnaire, is that elaborate item-level information is available. Without this information, we would not have been able to investigate systematic item bias in **chapter 4**, nor use the qualitative method to give meaning to the scores as well as determine the minimal important change in **chapter 8**. With the alignment

of the two scoring methods, we performed in **chapter 5**, in the future, all scores can be placed on a single scale. This eliminates the need for separate cutoffs and interpretation guidelines and allows for clinicians to directly apply the findings that we presented in **chapter 8.1** in practice.

When changes emerge

The second major theme we aimed to address in this thesis concerns the emergence of subtle impairments in everyday functioning. Traditionally, impairments in everyday life have been described in dementia.[1] Over the past decade, studies have shown that more subtle impairments in everyday functioning not only emerge before the onset of dementia,[39,40] but that they track with cognitive changes[41] and may predict future clinical progression to dementia.[42] We showed in chapters 2, 3, 5 and 6 that even among cognitively normal older adults, subtle difficulties in everyday functioning exist. While it is difficult to compare findings between studies because of differing definitions of 'difficulties in everyday functioning', other studies have shown subtle impairments in daily functioning among cognitively normal older adults as well.[40,43,44] In **chapter 5**, we showed that item response theory scores declined marginally over time among cognitively normal older adults who were included in a study that was designed to represent a population ready for participation in disease-modifying treatment studies. In **chapter 6**, we showed that individuals in the late preclinical stage of Alzheimer's disease show a steeper rate of decline than cognitively normal older adults with no underlying Alzheimer's disease, a finding that adds to the growing body of literature showing that decline in function occurs in individuals who are on a trajectory to dementia.[39,42,45] An important aspect of this study is that it combined data from different instruments. Interestingly, we did not find any substantial differences between the instruments that were used, suggesting that the trajectories we found apply to the construct of everyday functioning, rather than to any specific instrument used to measure the construct. This opens the door to combining established findings and performing collaborative research across countries and instruments. It also encourages harmonization efforts, which are already taking place in other areas, such as the study of subjective cognitive decline.[46]

In various studies included in this thesis, we observed a large heterogeneity in patterns of change in everyday functioning among memory clinic patients. In **chapter 7** we argued that, for an adequate prognosis, we need to understand what

characteristics are associated with different rates of decline in everyday functioning. Heterogeneity in clinical decline in Alzheimer's disease has been described in previous studies.[47-49] We were able to characterize some of the heterogeneity in rates of functional decline by identifying three groups of patients who showed a similar change over time. Similar clusters were found in overall clinical decline in another study:[50] a group of stable patients, a group of patients showing slow progression, and a group of patients showing fast progression. In line with expectations, patients in more severe disease stages and who performed worse on cognitive tests were more likely to show fast progression. Other characteristics were also related to fast progression of functional impairment and included age and atrophy. Together, these characteristics may be used by clinicians to identify those patients who are at an increased risk of showing functional decline.

Grouping patients with similar trajectories of functional change may help reduce some of the heterogeneity, but still does not entirely account for individual variation in changes over time. Individualized prediction models have been designed for determining risk of conversion from mild cognitive impairment to dementia based on biomarkers,[51,52] and these models may be applied to the prediction of future functional decline as well.

When changes matter

The third and final topic we address in this thesis, is the meaningfulness of changes in everyday functioning. In **chapter 8.2**, we added the aspect of clinical meaningfulness to changes in everyday functioning. Some patients may show some change over time, but it may not always affect their quality of life. With the guidance document from the United States Food and Drug Administration calling for proof of a clinically meaningful benefit of disease-modifying treatments in order to obtain full approval,[53] it is important to know when changes actually have an impact on the patient's life. As there is no consensus yet on what constitutes meaningful change,[54,55] the interpretation of outcome measures remains difficult.[56] Other attempts have been made to establish clinical meaningfulness of changes for cognitive and functional outcome measures,[57,58] but they rely on either statistical differences or anchors, neither of which accurately reflect what constitutes important change for the patient and their caregivers. Now that the Amsterdam IADL Questionnaire has threshold values for meaningful decline and meaningful improvement, it should have an advantage over other functional outcomes that do not possess such information.

A caveat to consider in this, is that it was more complicated to determine meaningful improvement than it was to determine meaningful decline. Currently, as there are no treatment options that show a clinical benefit, it may be hard for people to imagine what a meaningful improvement might look like. A question we raised in **chapter 8.2** was whether the absence of a meaningful decline could be considered a clinical benefit—in other words, if a potential new drug is able to prevent someone from experiencing a meaningful decline in everyday functioning, may we regard that as a successful treatment? Until it is better established what clinical improvements in the context of Alzheimer's disease and related disorders may look like, we might have to make do with viewing no meaningful decline as a clinical benefit. Ultimately, of course, the goal should be to achieve an actual improvement.

Clinical implications

The findings presented here have implications for clinical practice in addition to clinical trials. From chapters 2 and 3 we may conclude that assessing everyday functioning in cognitively normal older adults is relevant, because this population is not entirely free of functional difficulties. When a study partner is not available, it may be worthwhile to ask the patient or participant to rate their own level of everyday functioning. Furthermore, obtaining information from both the patient and a partner, relative or friend about patient's level of functioning, may provide additional insight into the level of everyday functioning. It seems that the value of the combined assessment by self and study partner might be greater than the sum of its two parts.

With the results from **chapter 8.1**, we provide guidance for interpreting the total scores of the Amsterdam IADL Questionnaire. Using the labels of 'no', 'mild', 'moderate' and 'severe impairment' adds a layer of meaningfulness to the scores, that the raw total scores lack. Good communication about what test results mean is a current unmet need in clinical practice,[59,60] and these labels may facilitate this communication. The categories were established by both clinicians and caregivers, making them applicable to experience both from the perspective of the clinician, as well as the patient and their caregivers.

The findings presented in chapters 6 and 7 show when changes in everyday functioning may occur and characterize some of the heterogeneity in individual

patterns of change in function over time. This reinforces the idea that impairment in everyday functioning is not limited to the dementia stage of neurodegenerative diseases, but that changes need to be monitored along the entire disease spectrum. When combining this with the results shown in **chapter 8.2**, clinicians may be better able to identify people who experience a meaningful change in everyday functioning. The thresholds for meaningful decline and improvement may also be used in the evaluation of newly developed disease-modifying treatments, as a way of establishing a clinically meaningful benefit.

Future directions

While we've added evidence for the psychometric quality of the Amsterdam IADL Questionnaire, there is still more to be investigated. More elaborate validation of the self-reported version of the Amsterdam IADL Questionnaire is warranted, specifically its sensitivity to change over time. The calculation of norm scores would improve the clinical value of the Amsterdam IADL Questionnaire, as they provide a reference with which to compare observed performance. Norm scores could also be used to establish cutoff values for distinguishing normal from abnormal performance in everyday functioning.

The work in this thesis relies on questionnaires, which provide self-perceived evaluation of a person's level of functioning. Self-perceived difficulties with everyday functioning, either as reported by the person themselves, or by someone close to them, have inherent meaning. Subjective complaints should not be dismissed when they cannot be objectified using existing tests. However, as many intrinsic factors, such as depressive mood or anxiety, may lead a person to report cognitive and functional complaints, we still need to understand better what part of self-perceived difficulties in everyday functioning relates to actual performance. When this relationship is better understood, we may be able to disentangle complaints due to the person's mental state from complaints stemming from an underlying disease that might progressively worsen. This may, in turn, improve prognostic accuracy. One way of achieving this deeper understanding, is by investigating the differences between self- and study partner-report. Questions that remain unanswered include which of the reports relates more strongly to objective performance and cognition, what the differences can tell us about underlying disease, and whether the difference between self- and study partner-report changes over time.

Another interesting avenue for future research is the comparison of performance-based measures of everyday functioning with patient or observer-reported outcomes. There are various performance-based tools for measuring everyday functioning, such as the Day-Out Task,[61] the Breakfast Task,[62] and the Naturalistic Action Test.[63] With the almost endless possibilities offered by smartphones and computers, the future of the assessment of everyday functioning may lie in using these technologies to obtain objective and precise measurements of difficulties in everyday functioning. An initiative that aims to do this is the Virtual Reality Functional Capacity Assessment Tool, which was shown to correlate strongly with study partner-reported IADL functioning, and seemed able to discriminate between individuals with mild cognitive impairment and individuals who were cognitively normal.[64] Further, combining objective with subjective measures may provide information on disease insight by both patient and observer, and can teach use about the unique value of subjective report of everyday functioning. It is important that all steps carried out with the patient and observer-reported outcome measures are reiterated for existing and new performance-based tests, as evidence on psychometric properties of performance-based assessments is currently lacking.[65]

A vital limitation that applies to this work, and more generally, to a large portion of clinical research in the West, is that groups who are underrepresented in society are also underrepresented in research. Babulal and colleagues have called for an improvement in the representation of ethnoracially diverse groups in Alzheimer's disease and related disorders research.[31] Activities of daily living are culturally specific, and they may also be different between ethnoracial groups. Therefore, in order to make findings more widely applicable, it is the responsibility of the scientific community to make a better effort in including participants from various backgrounds.

9.3 Conclusion

By investigating how to measure everyday functioning, when changes in everyday functioning emerge, and what these changes mean for the patient and their caregiver, this thesis contributed important new insights that may help shape the clinical practice and research endeavors of the future.

Together, the findings presented in this thesis may be used to inform decisions on the design of clinical trials aiming to assess the efficacy of new disease-modifying

treatments. Our conclusions may also help clinicians in identifying individuals at greater risk for decline in everyday functioning as well as in communicating with patients and their caregivers about impairments in daily life. Ultimately, this work may improve care and clinical research and, by extension, contribute to the continuing endeavor to reduce the suffering caused by Alzheimer's disease and related disorders.

References

1. Jack CR, Jr., Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement. Apr 2018;14(4):535-562. doi:10.1016/j.jalz.2018.02.018
2. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. Autumn 1969;9(3):179-86.
3. Sikkes SAM, De Rotrou J. A qualitative review of instrumental activities of daily living in dementia: what's cooking? Neurodegener Dis Manag. 2014;4(5):393-400. doi:10.2217/nmt.14.24
4. Sikkes SAM. Measuring IADL in dementia. Vrije Universiteit Amsterdam; 2011.
5. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. Alzheimers Dement. Nov 2012;8(6):536-43. doi:10.1016/j.jalz.2011.08.006
6. Sikkes SAM, Knol DL, Pijnenburg YAL, 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(1):35-41. doi:10.1159/000346277
7. Sikkes SAM, Pijnenburg YAL, 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. Dec 2013;26(4):244-50. doi:10.1177/0891988713509139
8. Koster N, Knol DL, Uitdehaag BM, Scheltens P, Sikkes SAM. The sensitivity to change over time of the Amsterdam IADL Questionnaire((c)). Alzheimers Dement. Oct 2015;11(10):1231-40. doi:10.1016/j.jalz.2014.10.006
9. Mokkink LB, Terwee CB, Patrick 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. Jul 2010;63(7):737-45. doi:10.1016/j.jclinepi.2010.02.006
10. Harrison J, Maruff P. Measuring the mind: assessing cognitive change in clinical drug trials. Expert Rev Clin Pharmacol. Jul 2008;1(4):471-3. doi:10.1586/17512433.1.4.471
11. Cummings J, Lee G, Zhong K, Fonseca J, Taghva K. Alzheimer's disease drug development pipeline: 2021. Alzheimers Dement (N Y). 2021;7(1):e12179. doi:10.1002/trc2.12179
12. Cummings J, Morstorf T, Lee G. Alzheimer's drug-development pipeline: 2016. Alzheimers Dement (N Y). Nov 2016;2(4):222-232. doi:10.1016/j.trci.2016.07.001
13. Grochowalski JH, Liu Y, Siedlecki KL. Examining the reliability of ADAS-Cog change scores. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn. Sep 2016;23(5):513-29. doi:10.1080/13825585.2015.1127320
14. Cogo-Moreira H, Krance SH, Black SE, et al. Questioning the Meaning of a Change on the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog): Noncomparable Scores and Item-Specific Effects Over Time. Assessment. Sep 2021;28(6):1708-1722. doi:10.1177/1073191120915273
15. Lim WS, Chong MS, Sahadevan S. Utility of the clinical dementia rating in Asian populations. Clin Med Res. Mar 2007;5(1):61-70. doi:10.3121/cmr.2007.693
16. Ackley SF, Zimmerman SC, Brenowitz WD, et al. Effect of reductions in amyloid levels on cognitive change in randomized trials: instrumental variable meta-analysis. BMJ. Feb 25 2021;372:n156. doi:10.1136/bmj.n156

17. Jutten RJ, Harrison JE, Lee Meeuw Kjoe PR, et al. Assessing cognition and daily function in early dementia using the cognitive-functional composite: findings from the Catch-Cog study cohort. Alzheimers Res Ther. May 15 2019;11(1):45. doi:10.1186/s13195-019-0500-5
18. Jutten RJ, Sikkes SAM, Van der Flier WM, et al. Finding Treatment Effects in Alzheimer Trials in the Face of Disease Progression Heterogeneity. Neurology. Jun 1 2021;96(22):e2673-e2684. doi:10.1212/WNL.0000000000012022
19. Nosheny RL, Jin C, Neuhaus J, et al. Study partner-reported decline identifies cognitive decline and dementia risk. Ann Clin Transl Neurol. Dec 2019;6(12):2448-2459. doi:10.1002/acn3.50938
20. Hanseeuw BJ, Scott MR, Sikkes SAM, et al. Evolution of anosognosia in alzheimer's disease and its relationship to amyloid. Ann Neurol. Feb 2020;87(2):267-280. doi:10.1002/ana.25649
21. Amariglio RE, Sikkes SAM, Marshall GA, et al. Item-Level Investigation of Participant and Study Partner Report on the Cognitive Function Index from the A4 Study Screening Data. J Prev Alzheimers Dis. 2021;8(3):257-262. doi:10.14283/jpad.2021.8
22. Tomaszewski Farias S, Giovannetti T, Payne BR, et al. Self-perceived Difficulties in Everyday Function Precede Cognitive Decline among Older Adults in the ACTIVE Study. J Int Neuropsychol Soc. Jan 2018;24(1):104-112. doi:10.1017/S1355617717000546
23. Rueda AD, Lau KM, Saito N, et al. Self-rated and informant-rated everyday function in comparison to objective markers of Alzheimer's disease. Alzheimers Dement. Sep 2015;11(9):1080-9. doi:10.1016/j.jalz.2014.09.002
24. Martyr A, Nelis SM, Clare L. Predictors of perceived functional ability in early-stage dementia: self-ratings, informant ratings and discrepancy scores. Int J Geriatr Psychiatry. Aug 2014;29(8):852-62. doi:10.1002/gps.4071
25. World Health Organization. International Classification of Functioning, Disability and Health (ICF). Accessed 13 Apr, 2022. https://who.int/standards/classifications/international-classification-of functioning-disability-and-health
26. Farias ST, Mungas D, Jagust W. Degree of discrepancy between self and other-reported everyday functioning by cognitive status: dementia, mild cognitive impairment, and healthy elders. Int J Geriatr Psychiatry. Sep 2005;20(9):827-34. doi:10.1002/gps.1367
27. Okonkwo OC, Wadley VG, Griffith HR, et al. Awareness of deficits in financial abilities in patients with mild cognitive impairment: going beyond self-informant discrepancy. Am J Geriatr Psychiatry. Aug 2008;16(8):650-9. doi:10.1097/JGP.0b013e31817e8a9d
28. Pol MC, Buurman BM, de Vos R, de Rooij SE. Patient and proxy rating agreements on activities of daily living and the instrumental activities of daily living of acutely hospitalized older adults. J Am Geriatr Soc. Aug 2011;59(8):1554-6. doi:10.1111/j.1532-5415.2011.03514.x
29. Ryu SY, Kim A, Kim S, et al. Self- and informant-reported cognitive functioning and awareness in subjective cognitive decline, mild cognitive impairment, and very mild Alzheimer disease. Int J Geriatr Psychiatry. Jan 2020;35(1):91-98. doi:10.1002/gps.5224
30. Chiu HF, Lam LC. Relevance of outcome measures in different cultural groups--does one size fit all? Int Psychogeriatr. Jun 2007;19(3):457-66. doi:10.1017/S1041610207004838
31. Babulal GM, Quiroz YT, Albensi BC, et al. Perspectives on ethnic and racial disparities in Alzheimer's disease and related dementias: Update and areas of immediate need. Alzheimers Dement. Dec 13 2018;doi:10.1016/j.jalz.2018.09.009
32. Black R, Greenberg B, Ryan JM, et al. Scales as outcome measures for Alzheimer's disease. Alzheimers Dement. Jul 2009;5(4):324-39. doi:10.1016/j.jalz.2009.05.667

33. Fleishman JA, Spector WD, Altman BM. Impact of differential item functioning on age and gender differences in functional disability. J Gerontol B Psychol Sci Soc Sci. Sep 2002;57(5):S275-84. doi:10.1093/geronb/57.5.S275

34. Tennant A, Penta M, Tesio L, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care. Jan 2004;42(1 Suppl):I37-48. doi:10.1097/01.mlr.0000103529.63132.77

35. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976). Dec 15 2000;25(24):3186-91.

36. Reise SP, Waller NG. Item response theory and clinical measurement. Annu Rev Clin Psychol. 2009;5:27-48. doi:10.1146/annurev.clinpsy.032408.153553

37. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, Scheltens P, Uitdehaag BM. A systematic review of Instrumental Activities of Daily Living scales in dementia: room for improvement. J Neurol Neurosurg Psychiatry. Jan 2009;80(1):7-12. doi:10.1136/jnnp.2008.155838

38. Gorter R, Fox JP, Twisk JW. Why item response theory should be used for longitudinal questionnaire data analysis in medical research. BMC Med Res Methodol. Jul 30 2015;15:55. doi:10.1186/s12874-015-0050-x

39. 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. Oct 1 2012;2(5):483-491. doi:10.2217/nmt.12.55

40. Marshall GA, Sikkes SAM, Amariglio RE, et al. Instrumental activities of daily living, amyloid, and cognition in cognitively normal older adults screening for the A4 Study. Alzheimers Dement (Amst). 2020;12(1):e12118. doi:10.1002/dad2.12118

41. Marshall GA, Aghjayan SL, Dekhtyar M, et al. Activities of daily living measured by the Harvard Automated Phone Task track with cognitive decline over time in non-demented elderly. J Prev Alzheimers Dis. 2017;4(2):81-86. doi:10.14283/jpad.2017.10

42. Sikkes SAM, Visser PJ, Knol DL, et al. Do instrumental activities of daily living predict dementia at 1- and 2-year follow-up? Findings from the Development of Screening guidelines and diagnostic Criteria for Predementia Alzheimer's disease study. J Am Geriatr Soc. Dec 2011;59(12):2273-81. doi:10.1111/j.1532-5415.2011.03732.x

43. Crimmins EM, Kim JK, Sole-Auro A. Gender differences in health: results from SHARE, ELSA and HRS. Eur J Public Health. Feb 2011;21(1):81-91. doi:10.1093/eurpub/ckq022

44. Scheel-Hincke LL, Moller S, Lindahl-Jacobsen R, Jeune B, Ahrenfeldt LJ. Cross-national comparison of sex differences in ADL and IADL in Europe: findings from SHARE. Eur J Ageing. Mar 2020;17(1):69-79. doi:10.1007/s10433-019-00524-y

45. Farias ST, Chou E, Harvey DJ, et al. Longitudinal trajectories of everyday function by diagnostic status. Psychol Aging. Dec 2013;28(4):1070-5. doi:10.1037/a0034069

46. Rabin LA, Smart CM, Crane PK, et al. Subjective cognitive decline in older adults: An overview of self-report measures used across 19 international research studies. J Alzheimers Dis. Sep 24 2015;48:S63-S86. doi:10.3233/JAD-150154

47. Tschanz JT, Corcoran CD, Schwartz S, et al. Progression of cognitive, functional, and neuropsychiatric symptom domains in a population cohort with Alzheimer dementia: the Cache County Dementia Progression study. Am J Geriatr Psychiatry. Jun 2011;19(6):532-42. doi:10.1097/JGP.0b013e3181faec23

48. Haaksma ML, Rizzuto D, Leoutsakos JS, et al. Predicting Cognitive and Functional Trajectories in People With Late-Onset Dementia: 2 Population-Based Studies. J Am Med Dir Assoc. Nov 2019;20(11):1444-1450. doi:10.1016/j.jamda.2019.03.025

49. van Loenhoud AC, van der Flier WM, Wink AM, et al. Cognitive reserve and clinical progression in Alzheimer disease: A paradoxical relationship. Neurology. Jul 23 2019;93(4):e334-e346. doi:10.1212/WNL.0000000000007821

50. Haaksma ML, Calderon-Larranaga A, Olde Rikkert MGM, Melis RJF, Leoutsakos JS. Cognitive and functional progression in Alzheimer disease: A prediction model of latent classes. Int J Geriatr Psychiatry. Aug 2018;33(8):1057-1064. doi:10.1002/gps.4893

51. van Maurik IS, Zwan MD, Tijms BM, et al. Interpreting Biomarker Results in Individual Patients With Mild Cognitive Impairment in the Alzheimer's Biomarkers in Daily Practice (ABIDE) Project. JAMA Neurol. Dec 1 2017;74(12):1481-1491. doi:10.1001/jamaneurol.2017.2712

52. van Maurik IS, Vos SJ, Bos I, et al. Biomarker-based prognosis for people with mild cognitive impairment (ABIDE): a modelling study. The Lancet Neurology. 2019;18(11):1034-1044. doi:10.1016/s1474-4422(19)30283-2

53. Early Alzheimer's disease: developing drugs for treatment - guidance for industry. 2018.

54. Rentz DM, Wessels AM, Annapragada AV, et al. Building clinically relevant outcomes across the Alzheimer's disease spectrum. Alzheimers Dement (N Y). 2021;7(1):e12181. doi:10.1002/trc2.12181

55. Rentz DM, Wessels AM, Bain LJ, Weber CJ, Carrillo MC. Clinical meaningfulness addressed at Alzheimer's Association Research Roundtable. Alzheimers Dement. May 2020;16(5):814. doi:10.1002/alz.12111

56. Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. BMJ. Feb 28 1998;316(7132):690-3. doi:10.1136/bmj.316.7132.690

57. Andrews JS, Desai U, Kirson NY, Zichlin ML, Ball DE, Matthews BR. Disease severity and minimal clinically important differences in clinical outcome assessments for Alzheimer's disease clinical trials. Alzheimers Dement (N Y). 2019;5:354-363. doi:10.1016/j.trci.2019.06.005

58. Insel PS, Weiner M, Mackin RS, et al. Determining clinically meaningful decline in preclinical Alzheimer disease. Neurology. Jul 23 2019;93(4):e322-e333. doi:10.1212/WNL.0000000000007831

59. Visser LNC, Kunneman M, Murugesu L, et al. Clinician-patient communication during the diagnostic workup: The ABIDE project. Alzheimers Dement (Amst). Dec 2019;11:520-528. doi:10.1016/j.dadm.2019.06.001

60. Fruijtier AD, Visser LNC, Bouwman FH, et al. What patients want to know, and what we actually tell them: The ABIDE project. Alzheimers Dement (N Y). 2020;6(1):e12113. doi:10.1002/trc2.12113

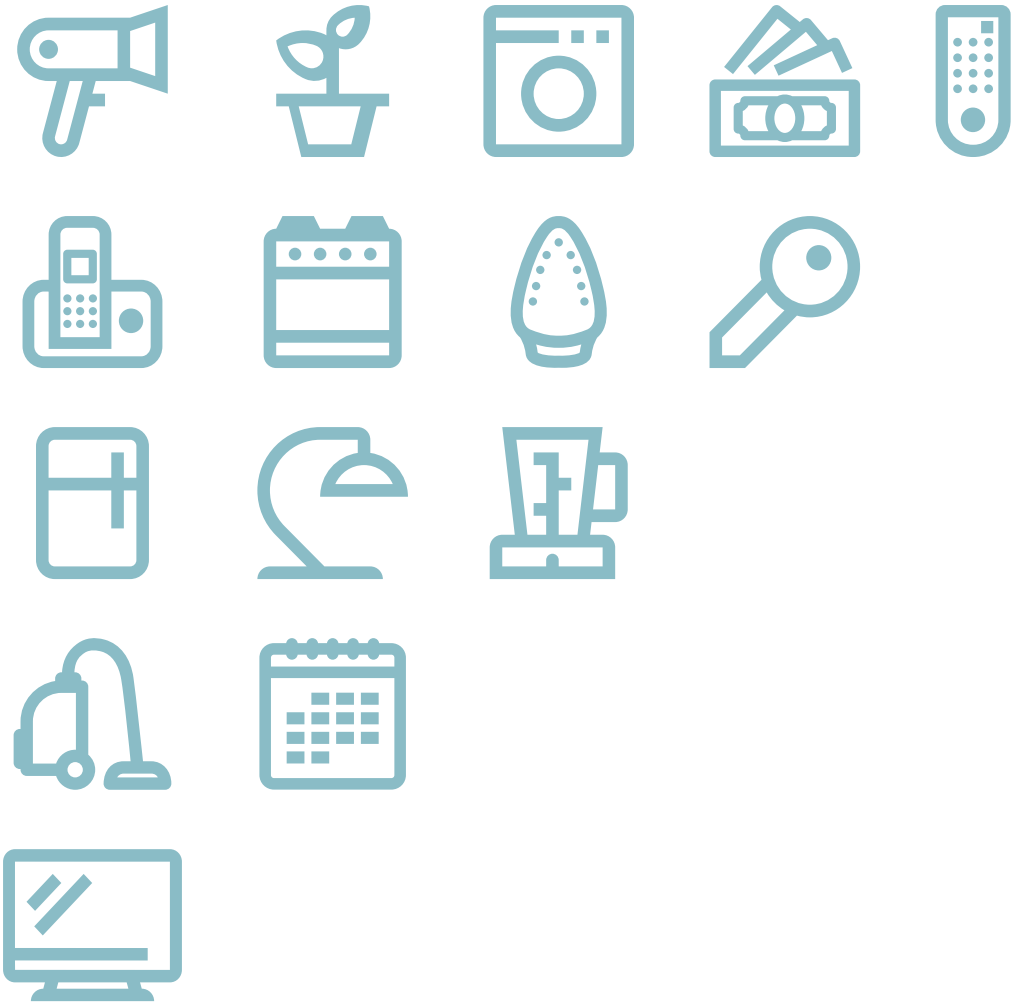
61. Schmitter-Edgecombe M, McAlister C, Weakley A. Naturalistic assessment of everyday functioning in individuals with mild cognitive impairment: the day-out task. Neuropsychology. Sep 2012;26(5):631-41. doi:10.1037/a0029352

62. Craik FI, Bialystok E. Planning and task management in older adults: cooking breakfast. Mem Cognit. Sep 2006;34(6):1236-49. doi:10.3758/bf03193268

63. Rycroft SS, Giovannetti T, Divers R, Hulswit J. Sensitive performance-based assessment of everyday action in older and younger adults. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn. Mar 2018;25(2):259-276. doi:10.1080/13825585.2017.1287855

64. Atkins AS, Khan A, Kelly SE, et al. P4-340: Performance-Based Assessment of Iadl Functioning in Mci and Mild Ad Using the Virtual Reality Functional Capacity Assessment Tool (Vrfcat): A Pilot Study. Alzheimer's & Dementia. 2018;14(7S_Part_30)doi:10.1016/j.jalz.2018.07.163

65. Wesson J, Clemson L, Brodaty H, Reppermund S. Estimating functional cognition in older adults using observational assessments of task performance in complex everyday activities: A systematic review and evaluation of measurement properties. Neurosci Biobehav Rev. Sep 2016;68:335-360. doi:10.1016/j.neubiorev.2016.05.024



“Sun always brings the light
Time always brings a change
We planted seeds inside
That were dancing in the rain
Growing up slowly over the hill
Winter is over
Here come the daffodils”

— Alicia Keys, ‘Daffodils’ (2021)

APPENDIX

Nederlandse samenvatting

In dit proefschrift hebben wij geprobeerd meer grip te krijgen op het meten van de eerste tekenen van cognitieve achteruitgang in het dagelijks leven bij beginnende dementie. In dit hoofdstuk zullen we de bevindingen van de onderzoeken waarop dit proefschrift berust, bespreken. Allereerst plaatsen we de onderzoeken in de algemene context, daarna bespreken we het doel van dit proefschrift en vatten we alle onderzoeken samen. Als laatst volgt een conclusie en kijken we kort naar de toekomst.

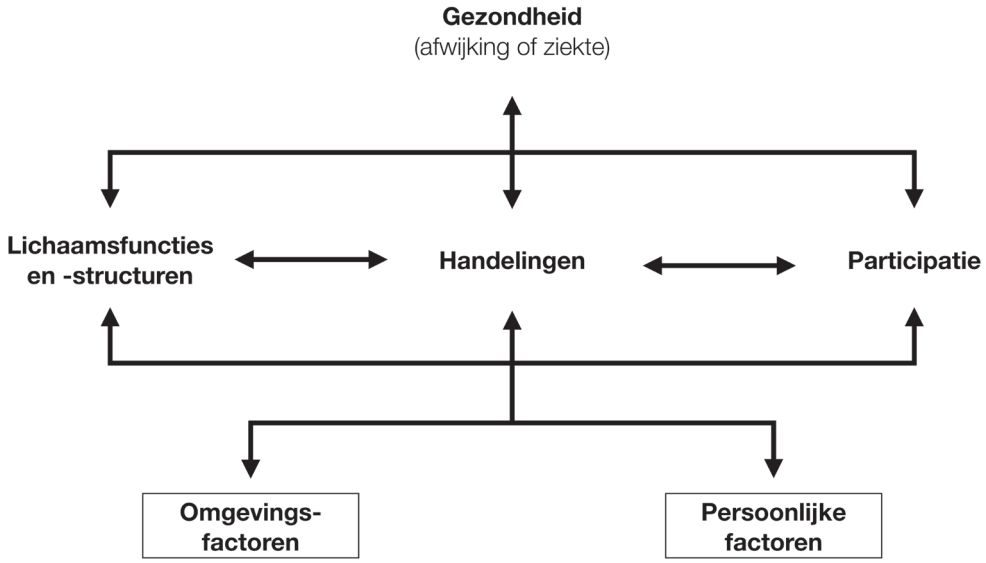
Inleiding: Dementie en de ziekte van Alzheimer

Bij neurodegeneratieve aandoeningen wordt de structuur van de hersenen langzaam aangetast. Dit leidt ertoe dat de hersenfuncties, zoals het geheugen, de aandacht en de planningsvaardigheden, geleidelijk afnemen. We gebruiken deze hersenfuncties elke dag zonder dat we daarbij stilstaan. Wanneer de hersenfuncties dusdanig zijn beperkt dat iemand niet meer zelfstandig kan functioneren in het dagelijks leven, spreken we van dementie. Op dit moment leven er in Nederland ongeveer 290.000 mensen met dementie en men verwacht dat dit aantal dankzij de vergrijzing verder zal stijgen naar meer dan een half miljoen in 2040.[1] Dementie is volgens het Centraal Bureau voor de Statistiek de snelst groeiende doodsoorzaak in Nederland. Vanwege de toenemende afhankelijkheid van anderen, wordt er door de persoon met dementie een steeds groter beroep gedaan op familie en andere naasten. Hierdoor vormt dementie een last voor de persoon zelf, zijn of haar omgeving, en uiteindelijk op de samenleving als geheel.

Dementie vormt het eindstadium van verschillende ziekten, waarvan de ziekte van Alzheimer de meest bekende en meest voorkomende is.[2] Alzheimer en gerelateerde ziekten ontwikkelen zich gedurende een groot aantal jaren. In toenemende mate worden deze ziekten biologisch gedefinieerd;[3] dat wil zeggen: wanneer bepaalde biologische processen, zoals de opstapeling van schadelijke eiwitten in de hersenen, aanwezig zijn, spreekt men van een ziekte. Biologische processen die een bepaalde ziekte karakteriseren, worden 'biomarkers' genoemd. Iemand kan een ziekte hebben—vastgesteld op basis van positieve biomarkers—, zonder dat hij of zij daar zelf iets van merkt. Dit eerste stadium, ook wel het preklinische stadium genoemd, duurt bij de ziekte van Alzheimer naar schatting tien tot vijftien jaar.[4] De eerste kenmerken van de ziekte uiten zich doorgaans in het ontstaan van subjectieve klachten van het geheugen,[5,6] gevolgd door milde objectiveerbare stoornissen.[7] Nu het steeds makkelijker wordt om biomarkers te bepalen, richt het klinisch onderzoek zich steeds meer op deze vroege

stadia, waarbij het ziekteproces onderweg is, maar de hersenfuncties nog niet zodanig zijn aangetast dat er sprake is van dementie. De overgang van het ene ziektestadium naar het andere is afhankelijk van veranderingen in de cognitie.

Cognitie is een verzamelterm voor de hersenfuncties die we gebruiken om onze weg door het dagelijks leven te vinden. Het omvat onder meer het geheugen, de aandacht, de taal en de mentale flexibiliteit. In de klinische praktijk is het gebruikelijk om de cognitie te meten door middel van neuropsychologische tests, die vaak zijn ontworpen om één specifiek onderdeel, of domein, van de cognitie te meten. Echter, in het dagelijks leven komt het zelden voor dat we voor een taak—bijvoorbeeld het boodschappen doen—slechts gebruikmaken van één cognitief domein. Mede daardoor zijn traditionele neuropsychologische tests maar beperkt ecologisch valide: ze vertalen zich niet goed naar het dagelijks leven. Cognitie in het dagelijks leven kan worden gevangen in de uitvoering van 'instrumentele algemene dagelijkse levensverrichtingen', of IADL. Dit zijn activiteiten waarvoor we meerdere cognitieve domeinen aanroepen, zoals het doen van de administratie, het gebruiken van een smartphone en het maken van afspraken. Volgens de Internationale Classificatie van het menselijk Functioneren (ICF) staat het uitvoeren van alledaagse handelingen centraal en wordt de mate waarin iemand daartoe in staat is bepaald door een combinatie van de gezondheid en omgevings- en persoonlijke factoren (Figuur 1).



Figuur 1 | Internationale Classificatie van het menselijk Functioneren

Het meten van problemen in het dagelijks leven berust vaak op vragenlijsten die door de naaste van een patiënt of onderzoeksdeelnemer worden ingevuld.[8] Dit is van oudsher het geval omdat bij gevorderde cognitieve beperkingen, iemand zelf niet meer goed in staat is om op zijn of haar eigen functioneren te reflecteren. Nu onderzoek ook vaak mensen betrekt die niet cognitief beperkt zijn, wordt er in toenemende mate gebruikgemaakt van zelfrapportage door de patiënt of onderzoeksdeelnemer in kwestie.

Om een goede uitkomstmaat te ontwikkelen moet een aantal psychometrische kwaliteiten worden onderzocht, waarvan betrouwbaarheid (de consistentie van een uitkomstmaat) en validiteit (de mate waarin de uitkomstmaat meet wat hij beoogt te meten) de belangrijkste zijn.[9] Voor het accuraat meten van het dagelijks functioneren is het belangrijk dat de uitkomstmaten aangetoond goede psychometrische kwaliteiten heeft, hetgeen voor veel bestaande uitkomstmaten voor de IADL nu niet het geval is.[10]

Doel van dit proefschrift

Het doel van dit proefschrift was om een beter beeld te krijgen van hoe we het dagelijks functioneren kunnen meten en welke klinische waarde deze metingen hebben. In het eerste deel beschreven we verschillende manieren voor het meten van cognitie in het dagelijks leven bij cognitief gezonde mensen, waarbij we specifiek keken naar de verschillen in rapportage door de deelnemers zelf en door hun naasten. In het tweede deel onderzochten we hoe verschillende factoren, waaronder het land waarin iemand woont, van invloed kunnen zijn op hoe we het dagelijks functioneren meten. In het derde deel van dit proefschrift onderzochten we het meten van veranderingen in het dagelijks functioneren, waarbij we keken of een bepaalde manier van het scoren van een vragenlijst meer gevoelig kan zijn voor achteruitgang. Daarnaast onderzochten we in welk stadium van de ziekte van Alzheimer de eerste achteruitgang in het dagelijks functioneren te meten is. Ook bekeken of we groepen patiënten konden onderscheiden die een vergelijkbare verandering in het dagelijks functioneren doormaakten en probeerden we deze groepen te karakteriseren aan de hand van klinische gegevens. Als laatst hebben we vastgesteld wanneer veranderingen in het dagelijks functioneren als belangrijk worden ervaren en wat er samenhangt met belangrijke veranderingen.

Samenvatting van de studies

In hoofdstuk 2 onderzochten we hoe cognitief normale mensen hun eigen dagelijks functioneren beoordelen. We keken ook naar de mate van inzicht van deelnemers in hun dagelijks functioneren, door de verschillen te onderzoeken tussen hoe de deelnemers zelf en hoe hun naasten oordeelden over het functioneren van de deelnemer. Ook onderzochten we welke factoren geassocieerd waren met de mate van inzicht. Hiervoor nodigden we een grote groep gezonde vrijwilligers uit: deze mensen hadden zelf geen diagnose dementie. We vroegen de deelnemers de Amsterdam IADL vragenlijst in te vullen over hun eigen dagelijks functioneren (zelfrapportage). Daarnaast vroegen we hen iemand uit hun omgeving te vragen om dezelfde vragenlijst in te vullen (naasterapportage). We veronderstelden ten eerste dat er weinig mensen zouden zijn die problemen rapporteren in het dagelijks functioneren. Ten tweede verwachtten we dat de zelf- en naasterapportage grotendeels overeen zouden komen. Met andere woorden: we verwachtten dat de mate van zelfinzicht in het dagelijks functioneren onder gezonde vrijwilligers goed zou zijn. We vonden inderdaad dat de meerderheid van de deelnemers geen problemen hadden in het functioneren, zowel volgens henzelf als volgens hun naasten. Echter, er was slechts een matige overeenstemming tussen zelf- en naasterapportage. Ongeveer een kwart van de deelnemers rapporteerde zelf méér problemen dan hun naasten en iets minder dan een vijfde rapporteerde juist minder problemen. Dit eerste zou kunnen worden gezien als een verhoogd inzicht, waarbij iemand zelf veel problemen opmerkt die zijn of haar omgeving niet bemerkt. Het tweede kan als verminderd inzicht worden gezien: waar de omgeving zegt dat er problemen zijn, ervaart de persoon dit zelf niet zo. We vonden dat mensen die depressieve en geheugenklachten hadden, een grotere kans hadden om meer problemen te rapporteren dan hun naaste. Dit zou erop kunnen duiden dat interne factoren zoals stemming een invloed kunnen hebben op hoe iemand op zijn of haar eigen functioneren reflecteert. Uit onze bevindingen maakten we op dat zelf- en naasterapportage niet zondermeer door elkaar gebruikt kunnen worden, zelfs niet in gezonde vrijwilligers. Dit bevestigt de noodzaak om ook altijd een naaste te vragen hoe een deelnemer of patiënt functioneert, vooral wanneer de populatie waarin men is geïnteresseerd vaker bepaalde klachten heeft.

In hoofdstuk 3 onderzochten we de associatie tussen het dagelijks functioneren en één van de biologische kenmerken van de ziekte van Alzheimer: het tau-eiwit. Dit deden we in een groep deelnemers van een groot Amerikaans onderzoek onder

cognitief normale volwassenen. Alle deelnemers ondergingen een hersenscan waarbij een klein beetje radioactief materiaal wordt geïnjecteerd: een *positron emission tomography*, of PET, scan. Het radioactieve materiaal bindt zich aan het tau-eiwit in de hersenen en kan door de scanner in beeld gebracht worden. Hierdoor wordt zichtbaar hoeveel tau-eiwit aanwezig is en waar in de hersenen het zich bevindt. We richtten ons in onze analyses op twee specifieke hersengebieden die belangrijk zijn voor de werking van het geheugen. Deelnemers en hun naasten hadden voor het onderzoek daarnaast een vragenlijst ingevuld over het dagelijks functioneren. We vonden een relatie tussen de hoeveelheid tau-eiwit in beide onderzochte hersengebieden en problemen in het dagelijks functioneren, zoals dat werd gerapporteerd door de studiepartner. Opmerkelijk genoeg, vonden we deze relatie niet voor de zelfgerapporteerde problemen in het dagelijks functioneren. Net als in hoofdstuk 2 vonden we dat er verschillen waren tussen hoe de deelnemers zelf over hun functioneren rapporteerden en hoe hun naasten dat deden. Deelnemers die meer tau-eiwit hadden, rapporteerden hun problemen vaak anders dan hun naasten. Mensen die meer schadelijk tau-eiwit in hun hersenen hebben, rapporteren meer klachten in het dagelijks functioneren dan hun naasten. Tegelijkertijd vonden we ook het tegengestelde effect: er waren ook mensen met meer tau-eiwit die juist minder problemen zeiden te ondervinden dan hun naasten. Beide bevindingen duiden mogelijk op een veranderd ziekte-inzicht, al moet toekomstig onderzoek nog uitwijzen hoe de vork precies in de steel zit.

Deel II van het proefschrift staat in het teken van verschillende factoren die van invloed kunnen zijn op het meten van het dagelijks functioneren. In hoofdstuk 4.1 onderzochten we hoe een aantal van deze factoren de manier waarop mensen de Amsterdam IADL vragenlijst invullen kan beïnvloeden. De Amsterdam IADL vragenlijst maakt gebruikt van de itemresponstheorie om de totaalscores te gebruiken. Itemresponstheorie berust op wiskundige modellen en houdt rekening met de kans dat iemand een bepaald antwoord op een vraag geeft, gegeven zijn of haar algemene functioneringsniveau.[11] Met deze gegevens kan ook onderzocht worden of er systematische verschillen bestaan tussen mensen uit verschillende groepen met hetzelfde algemene functioneringsniveau in de kans dat iemand een bepaald antwoord geeft. Dit kan worden gedaan met een analyse die *differential item functioning*, of DIF, wordt genoemd. Om DIF te onderzoeken, maakten we gebruik van een gecombineerde dataset met gegevens van ruim 3500 mensen uit acht

landen: Nederland, Spanje, Frankrijk, de Verenigde Staten, het Verenigd Koninkrijk, Griekenland, Servië en Finland. Al deze mensen deden mee aan onderzoeken naar cognitieve achteruitgang, ofwel omdat ze zelf klachten hadden, ofwel omdat ze vanwege genetische of vasculaire factoren een verhoogd risico hadden op cognitieve achteruitgang. Als eerst onderzochten we of er systematische verschillen waren tussen de deelnemers uit de verschillende landen. Hoewel enkele vragen wel statistisch significante DIF bleken te hebben, waren de effecten van deze DIF dermate klein dat ze praktisch niet betekenisvol en dus verwaarloosbaar waren. Verder onderzochten we deze systematische verschillen ook tussen mannen en vrouwen, oudere en jongere deelnemers en deelnemers met een laag en een hoog opleidingsniveau. Hier bleken geen verschillen te zijn. We concludeerden dat de Amsterdam IADL vragenlijst niet duidelijk beïnvloed wordt door deze verschillende factoren, wat de vergelijkbaarheid van gegevens tussen onderzoeken met een verscheidenheid aan deelnemers bevordert. Hierbij moet opgemerkt worden dat de deelnemers uit de onderzoeken die in deze studie werden gebruikt een redelijk homogene groep vormden; zij kwamen vrijwel allemaal uit Europese landen, waren grotendeels blank en relatief hoog opgeleid. Toekomstig onderzoek met deelnemers uit andere delen van de wereld en met een andere achtergrond moet uitwijzen of de tot nu toe gevonden resultaten breder kunnen worden gegeneraliseerd. In hoofdstuk 4.2 herhaalden we deze stappen specifiek in een groep mensen uit Duitstalig Zwitserland. Ook hier vonden we geen aanwijzingen dat er verschillen waren tussen hoe Nederlandse en Zwitserse deelnemers de vragen beantwoorden. Behalve de invloed van het land waarin iemand woont op hoe hij of zij de vragenlijst invult, onderzochten we in dit hoofdstuk ook de samenhang met andere constructen en de test–hertestbetrouwbaarheid. Dit was in eerder onderzoek al aangetoond,[12,13] maar nog niet eerder met deze vertaling. Voor de Zwitsers-Duitse vertaling vonden we een vergelijkbare samenhang met externe constructen en ook de test–hertestbetrouwbaarheid bleek goed.

Deel III van dit proefschrift richtte zich op het meten van veranderingen in het dagelijks functioneren. In hoofdstuk 5 vergeleken we twee methoden om de score van de Amsterdam IADL vragenlijst te berekenen, met als doel om te onderzoeken of één van de twee gevoeliger is voor subtiele achteruitgang. De ene methode, itemresponstheorie, weegt als het ware de moeilijkheid van elke activiteit mee. De naaste van iemand die in het algemeen veel moeite heeft met dagelijkse handelingen, zal met een grotere waarschijnlijkheid aangeven dat diegene veel

moeite heeft met bijvoorbeeld het uitvoeren van de administratie dan de naaste van iemand die weinig problemen heeft in het dagelijks leven. De andere methode is eenvoudiger te berekenen door het gemiddelde te nemen van de antwoorden op alle vragen. We vonden dat beide methoden zeer vergelijkbare resultaten gaven, al leek de complexere methode wel een minder sterk plafondeffect te laten zien. Als laatste stap maakten we omrekentabellen waarmee de twee scoringsmethoden aan elkaar gelinkt kunnen worden. Dit betekent dat de complexere scores benaderd kunnen worden met de eenvoudige manier van scoren. Het voordeel hiervan is dat er nog slechts één schaal nodig is en dat het werk dat we uitvoeren breder toepasbaar is. In hoofdstuk 6 lieten we zien dat het dagelijks functioneren in de ziekte van Alzheimer geleidelijk achteruitgaat en dat deze achteruitgang steeds harder gaat, naarmate de ziekte zich verder ontwikkelt. Op basis van klinische gegevens, kunnen mensen met de ziekte van Alzheimer worden ingedeeld in verschillende stadia. Van oudsher werd gedacht dat het dagelijks functioneren pas in het dementiestadium beperkt was, maar wij lieten zien dat er zelfs sprake is van een lichte achteruitgang in het dagelijks functioneren wanneer mensen subjectieve geheugenklachten hebben, maar nog geen geobjectiveerde cognitieve beperkingen. Dit kan betekenen dat de cognitieve tests die veel worden gebruikt in de vroegste ziektestadia nog niet erg gevoelig zijn, terwijl maten van dagelijks functioneren dat mogelijk wel zijn. Onze bevindingen pleiten er dan ook voor om het dagelijks functioneren ook in het begin van de ziekte te meten.

Met de kennis dat het dagelijks functioneren afneemt naarmate de ziekte van Alzheimer zich ontwikkelt, resteert de vraag welke mensen harder achteruitgaan en welke mensen langer stabiel blijven. Dit onderzochten we in hoofdstuk 7. We lieten een computeralgoritme naar patronen zoeken in de herhaalde metingen van de Amsterdam IADL vragenlijst in een grote groep patiënten. Het algoritme bepaalde dat er zeven verschillende groepen patiënten waren, die elk een net iets andere verandering over tijd lieten zien. Bij het karakteriseren van de groepen hebben we een aantal samengenomen, omdat zij in grote lijnen dezelfde patronen lieten zien. Zo bleven er drie groepen over: een groep patiënten die stabiel blijft, een groep patiënten die langzaam achteruitgaat en een groep patiënten die hard achteruitgaat. Als volgende stap probeerden we vast te stellen welke gegevens die bij het eerste bezoek zijn gemeten konden voorspellen wie in welke groep terechtkwam. We vonden dat de patiënten die stabiel bleven hogere scores hadden op de neuropsychologische tests, met name die die gericht zijn op de

mentale flexibiliteit, maar ook dat zij minder hersenschade hadden. Patiënten die het hardst achteruitgingen hadden al bij hun eerste bezoek aan de geheugenpoli meer hersenschade en lagere scores op de neuropsychologische tests.

In hoofdstuk 8 keken we naar de klinische relevantie van veranderingen in het dagelijks functioneren. In hoofdstuk 8.1 voerden we een kwalitatief onderzoek uit waarbij we de naasten van mensen met dementie uitnodigden voor focusgroepen. Zij kregen verschillende casussen voorgelegd van fictieve patiënten, waarin het dagelijks functioneren werd beschreven. We vroegen de deelnemers de casussen te sorteren van degene die de minste problemen had naar degene die de meeste problemen had. Vervolgens vroegen we ze boekenleggers te plaatsen tussen de casussen op de grens van geen en milde, milde en matige, en matige en ernstige problemen. Met name aan de uiteinden van het spectrum, wanneer er geen of juist ernstige problemen zijn, was er overeenstemming tussen de deelnemers. In hoofdstuk 8.2 keken we vervolgens naar de klinische relevantie van veranderingen in het dagelijks functioneren. We bepaalden drempelwaarden voor de zogenaamde '*minimally important change*', de kleinste belangrijke verandering, voor zowel achteruitgang als verbetering in het dagelijks functioneren. Dit deden we opnieuw met casussen van fictieve patiënten, die we voorlegden aan een grote groep naasten van mensen met dementie. We legden dezelfde casussen ook voor aan een aantal clinici van verschillende geheugenpoliklinieken in Nederland. De drempelwaarden die we vonden hebben we vervolgens toegepast op een groep patiënten van Alzheimercentrum Amsterdam. Zij werden een jaar lang gevolgd en hun naasten vulden elke drie maanden de Amsterdam IADL vragenlijst in. We vonden dat ongeveer de helft van de patiënten binnen één jaar een klinisch relevante achteruitgang liet zien. Patiënten die waren gediagnosticeerd met dementie en patiënten die meer schade hadden in Alzheimer-gerelateerde hersengebieden hadden een grotere kans om een klinisch relevante achteruitgang door te maken dan patiënten met subjectieve klachten en geen hersenschade.

Conclusie

Op basis van de in dit proefschrift beschreven onderzoeken kunnen we allereerst concluderen dat het meten van het dagelijks functioneren van groot belang is, niet alleen bij dementie, maar ook in de vroegere stadia van de ziekte van Alzheimer en aanverwante aandoeningen. In deze vroege stadia is het van toegevoegde waarde meerdere bronnen te raadplegen, dat wil zeggen: door een

vragenlijst aan zowel de onderzoeksdeelnemer als aan zijn of haar naaste voor te leggen, kan extra informatie worden gewonnen op een manier waarbij het geheel groter is dan de som van de losse onderdelen. Verder is het belangrijk een goede, sensitieve maat te hebben, die niet sterk wordt beïnvloed door factoren als culturele achtergrond, leeftijd of geslacht en die daarnaast gevoelig is voor klinisch relevante veranderingen. Voor veel uitkomstmaten is dit helaas nog niet of onvoldoende onderzocht, maar zonder deze kennis kunnen we niet zeker zijn dat de resultaten die we vinden daadwerkelijk een weerspiegeling zijn van de cognitieve veranderingen die plaatsvinden.

De uitkomsten van de onderzoeken die we in dit proefschrift beschrijven, dragen bij aan de kennis over de effecten van cognitieve achteruitgang in het dagelijks leven. Daarnaast hebben onze bevindingen belangrijke implicaties voor zowel klinisch gebruik als onderzoek. Nu het eerste alzheimermedicijn in de Verenigde Staten is goedgekeurd, en er naar verwachting in de toekomst meer zullen volgen, is het van extra groot belang dat we de mogelijke effecten van medicamenteuze behandelingen goed kunnen evalueren. Hiervoor moeten we niet alleen effecten zien op de eiwitophopingen, maar ook op klinische uitkomstmaten, zoals het dagelijks functioneren. Als een nieuw middel een klinisch relevante achteruitgang kan voorkómen, of zelfs een klinisch relevante verbetering kan bewerkstelligen, kunnen we daadwerkelijk onze hoop vestigen op een toekomst waarin niemand meer hoeft te lijden aan de ziekte van Alzheimer.

Referenties

1. Alzheimer Nederland. Factsheet cijfers en feiten over dementie. Updated February 2021. Accessed November 19, 2021. https://www.alzheimer-nederland.nl/factsheet-cijfers-en-feiten-over-dementie
2. Scheltens P, De Strooper B, Kivipelto M, et al. Alzheimer's disease. The Lancet. 2021;doi:10.1016/s0140-6736(20)32205-4
3. Jack CR, Jr., Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement. Apr 2018;14(4):535-562. doi:10.1016/j.jalz.2018.02.018
4. Vermunt L, Sikkes SAM, van den Hout A, et al. Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease in relation to age, sex, and APOE genotype. Alzheimers Dement. Jul 2019;15(7):888-898. doi:10.1016/j.jalz.2019.04.001
5. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers & Dementia. May 2011;7(3):280-292. doi:10.1016/j.jalz.2011.03.003
6. Jessen F, Amariglio RE, Buckley RF, et al. The characterisation of subjective cognitive decline. The Lancet Neurology. 2020;19(3):271-278. doi:10.1016/s1474-4422(19)30368-0
7. Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild cognitive impairment: a concept in evolution. J Intern Med. Mar 2014;275(3):214-28. doi:10.1111/joim.12190
8. Sikkes SAM, De Rotrou J. A qualitative review of instrumental activities of daily living in dementia: what's cooking? Neurodegener Dis Manag. 2014;4(5):393-400. doi:10.2217/nmt.14.24
9. Mokkink LB, Terwee CB, Patrick 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. Jul 2010;63(7):737-45. doi:10.1016/j.jclinepi.2010.02.006
10. Sikkes SAM, de Lange-de Klerk ES, Pijnenburg YAL, Scheltens P, Uitdehaag BM. A systematic review of Instrumental Activities of Daily Living scales in dementia: room for improvement. J Neurol Neurosurg Psychiatry. Jan 2009;80(1):7-12. doi:10.1136/jnnp.2008.155838
11. Reise SP, Waller NG. Item response theory and clinical measurement. Annu Rev Clin Psychol. 2009;5:27-48. doi:10.1146/annurev.clinpsy.032408.153553
12. Jutten RJ, Peeters CFW, Leijdesdorff SMJ, et al. Detecting functional decline from normal aging to dementia: Development and validation of a short version of the Amsterdam IADL Questionnaire. Alzheimers Dement (Amst). 2017;8:26-35. doi:10.1016/j.dadm.2017.03.002
13. Sikkes SAM, Knol DL, Pijnenburg YAL, 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(1):35-41. doi:10.1159/000346277

List of publications

Dubbelman, M.A., Verrijp, M., Facal, D., Sánchez-Benavides, G., Brown, L.J.E., van der Flier, W.M., Jokinen, H., Lee, A., Leroi, I., Lojo-Seoane, C., Milošević, V., Molinuevo, J.L., Pereiro Rozas, A.X., Ritchie, C., Salloway, S., Stringer, G., Zygouris, S., Dubois, B., Epelbaum, S., Scheltens, P., Sikkes, S.A.M. (2020). The influence of diversity on the measurement of functional impairment: An international validation of the Amsterdam IADL Questionnaire in eight countries. *Alzheimers Dement (Ams)*, 12*(1)*:e12021. DOI: 10.1002/dad2.12021.

Bruderer-Hofstetter, M., **Dubbelman, M.A.,** Meichtry, A., Koehn, F., Münzer, T., Jutten, R.J., Scheltens, P., Sikkes, S.A.M., Niedermann, K. (2020). Cross-cultural adaptation and validation of the Amsterdam Instrumental Activities of Daily Living Questionnaire Short Version German for Switzerland. *Health Qual Life Outcomes*, *18*:323. DOI: 10.1186/s12955-020-01576-w.

Dubbelman, M.A., Jutten, R.J., Tomaszewski Farias, S.E., Amariglio, R.E., Buckley, R.F., Visser, P.J., Rentz, D.M., Jonhson, K.A., Properzi, M.J., Schultz, A., Donovan, N., Gatchel, J.R., Teunissen, C.E., van Berckel, B.N.M., van der Flier, W.M., Sperling, R.A., Papp, K.V., Scheltens, P., Marshall, G.A., Sikkes, S.A.M. (2020). Decline in cognitively complex everyday activities accelerates along the Alzheimer's disease continuum. *Alzheimers Res Ther*, *12*:138. DOI: 10.1186/s13195-020-00706-2.

Verrijp, M., **Dubbelman, M.A.,** Visser, L.N.C., Jutten, R.J., Nijhuis, E.W., Zwan, M.D., van Hout, H.P.J., Scheltens, P., van der Flier, W.M., Sikkes, S.A.M. (2022). Everyday functioning in a community-based volunteer population: Differences between participant- and study partner-report. *Front Aging Neurosci*, *13*:761932. DOI: 10.3389/fnagi.2021.761932.

Dubbelman, M.A., Terwee, C.B., Verrijp, M., Visser, L.N.C., Scheltens, P., Sikkes, S.A.M. (2022). Giving meaning to the scores of the Amsterdam Instrumental Activities of Daily Living Questionnaire: A qualitative study. *Health Qual Life Outcomes*, *20*:47. DOI: 10.1186/s12955-022-01958-2.

Dubbelman, M.A., Verrijp, M., Terwee, C.B., Jutten, R.J., Postema, M.C., Barkhof, F., van Berckel, B.N.M., Gillissen, F., Teeuwen, V., Teunissen, C.E., van der Flier, W.M., Scheltens, P. Sikkes, S.A.M. (2022). Pursuing clinical meaningfulness: Determining the minimal important change of everyday functioning in dementia. *Neurology*, published online ahead of print. DOI: 10.1212/WNL.0000000000200781.

Dubbelman, M.A., Sanchez, J., Schultz, A.P., Rentz, D.M., Amariglio, R.E., Sikkes, S.A.M., Sperling, R.A., Johnson, K.A., Marshall, G.A. (2022). Everyday functioning and entorhinal and inferior temporal tau in cognitively normal older adults. *J Prev Alz Dis*, published online ahead of print. DOI: 10.14283/jpad.2022.58.

Dubbelman, M.A., Postema, M.C., Jutten, R.J., Harrison, J.E., Ritchie, C.W., Aleman, A., de Jong, F.J., Schalet, B.D., Terwee, C.B., van der Flier, W.M., Scheltens, P., Sikkes, S.A.M. Comparing and aligning scores based on item response theory and classical test theory for the Amsterdam Instrumental Activities of Daily Living Questionnaire: A longitudinal study. *Under review.*

Dubbelman, M.A., Caelers, M., Jutten, R.J., Lemstra A.W., Teunissen, C.E., van Berckel, B.N.M., Barkhof, F., van der Flier, W.M., Scheltens, P., Sikkes, S.A.M. Identifying and characterizing heterogeneity in everyday functioning progression among memory clinic patients. *In preparation.*

Not in this thesis

Sikkes, S.A.M., **Dubbelman, M.A.,** Verrijp, M., Jutten, R.J., Gillissen, F., Pijnenburg, Y.A.L., & Scheltens, P. (2019). Het meten van problemen in het dagelijks leven bij beginnende dementie: De Amsterdam IADL-vragenlijst. *Tijdschrift voor Neuropsychologie*, *14*(2), 126–137.

Visser, L.N.C., **Dubbelman, M.A.,** Verrijp, M., Wanders, L., Pelt, S., Zwan, M.D., Wouters, H., Sikkes, S.A.M., van Hout, H.P.J., van der Flier, W.M. (2021). Online cognitive testing at home using the Cognitive Online Self-Test Amsterdam (COST-A): Establishing psychometric properties and normative data. *Alzheimers Dement (Ams)*, *13*(1):e12234. DOI: 10.1002/dad2.12234.

Rabin, L.A., Sikkes, S.A.M., Tommet, D., Jones, R.N., Crane, P.K., Elbulok-Charcape, M., **Dubbelman, M.A.,** Koscik, R.L., Amariglio, R.E., Buckley, R.F., Boada, M., Chételat, G., Dubois, B., Ellis, K.A., Gifford, K.A., Jefferson, A.L., Jessen, F., Johnson, S.C., Katz, M.J., Lipton, R.B., Luck, T., Maruff, P., Molinuevo, J.L., Perrotin, A., Petersen, R.C., Rami, L., Reisberg, B., Rentz, D.M., Riedel-Heller, S.G., Risacher, S.L., Rodriguez-Gomez, O., Sachdev, P.S., Saykin, A.J., Scarmeas, N., Smart, C., Snitz, B.E., Sperling, R.A., van der Flier, W.M., van Harten, A.C., Wagner, M., Wolfsgruber, S. & the Alzheimer's Disease Neuroimaging Initiative, on behalf of the ISTAART Subjective Cognitive Decline Professional Interest Area. Linking self-perceived cognitive functioning questionnaires across aging studies: Statistical harmonization and analysis of item properties from the Subjective Cognitive Decline Initiative. *Under review.*

Sikkes, S.A.M., **Dubbelman, M.A.,** Jones, R.N., Tommet, D., Rabin, L.A. Subjective cognitive decline in relationship to amyloid accumulation and clinical progression: An international harmonization and validation effort. *In preparation.*

Dubbelman, M.A., Sikkes, S.A.M., Ebenau, J.L., Van Leeuwenstijn, M.S.S.A, Kroeze, L.A., Trieu, C., van Berckel, B.N.M., Teunissen, C.E., van Harten, A.C., van der Flier, W.M.F. Changes in self- and study partner-perceived cognitive functioning in relation to amyloid status and future clinical progression: Findings from the SCIENCe project. *Submitted.*

List of theses of the Alzheimer Center Amsterdam

1. L. Gootjes: Dichotic listening, hemisphere connectivity and dementia (September 14, 2004)
2. K. van Dijk: Peripheral nerve stimulation in Alzheimer's disease (January 16, 2005)
3. R. Goekoop: Functional MRI of cholinergic transmission (January 16, 2006)
4. R. Lazeron: Cognitive aspects in multiple sclerosis (July 3, 2006)
5. N.S.M. Schoonenboom: CSF markers in dementia (November 10, 2006)
6. E.S.C. Korf: Medial temporal lobe atrophy on MRI: risk factors and predictive value (November 22, 2006)
7. B. van Harten: Aspects of subcortical vascular ischemic disease (December 22, 2006)
8. B. Jones: Cingular cortex networks: role in learning and memory and Alzheimer's disease related changes (March 23, 2007)
9. L. van de Pol: Hippocampal atrophy from aging to dementia: a clinical and radiological perspective (May 11, 2007)
10. Y.A.L. Pijnenburg: Frontotemporal dementia: towards an earlier diagnosis (July 5, 2007)
11. A. Bastos Leite: Pathological ageing of the brain (November 16, 2007)
12. E.C.W. van Straaten: Vascular dementia (January 11, 2008)
13. R.L.C. Vogels: Cognitive impairment in heart failure (April 11, 2008)
14. J. Damoiseaux: The brain at rest (May 20, 2008)
15. G.B. Karas: Computational neuro-anatomy (June 19, 2008)
16. F.H. Bouwman: Biomarkers in dementia: longitudinal aspects (June 20, 2008)
17. A.A. Gouw: Cerebral small vessel disease on MRI: clinical impact and underlying pathology (March 20, 2009)
18. H. van der Roest: Care needs in dementia and interactive digital information provisioning (October 12, 2009)
19. C. Mulder: CSF biomarkers in Alzheimer's disease (November 11, 2009)
20. W. Henneman: Advances in hippocampal atrophy measurement in dementia: beyond diagnostics (November 27, 2009)
21. S.S. Staekenborg: From normal aging to dementia: risk factors and clinical findings in relation to vascular changes on brain MRI (December 23, 2009)
22. N. Tolboom: Imaging Alzheimer's disease pathology in vivo: towards an early diagnosis (February 12, 2010)
23. E. Altena: Mapping insomnia: brain structure, function and sleep intervention (March 17, 2010)
24. N.A. Verwey: Biochemical markers in dementia: from mice to men. A translational approach (April 15, 2010)
25. M.I. Kester: Biomarkers for Alzheimer's pathology: monitoring, predicting and understanding the disease (January 14, 2011)
26. J.D. Sluimer: Longitudinal changes in the brain (April 28, 2011)
27. S.D. Mulder: Amyloid associated proteins in Alzheimer's disease (October 7, 2011)
28. S.A.M. Sikkes: Measuring IADL in dementia (October 14, 2011)

29. A. Schuitemaker: Inflammation in Alzheimer's disease: in-vivo quantification (January 27, 2012)
30. K. Joling: Depression and anxiety in family caregivers of persons with dementia (April 2, 2012)
31. W. de Haan: In a network state of mind (November 2, 2012) (Cum laude)
32. D. van Assema: Blood-brain barrier P-glycoprotein function in ageing and Alzheimer's disease (December 7, 2012)
33. J.D.C. Goos: Cerebral microbleeds: connecting the dots (February 6, 2013)
34. R. Ossenkoppele: Alzheimer PEThology (May 8, 2013)
35. H.M. Jochemsen: Brain under pressure: influences of blood pressure and angiotensin-converting enzyme on the brain (October 4, 2013)
36. A.E. van der Vlies: Cognitive profiles in Alzheimer's disease: Recognizing its many faces (November 27, 2013)
37. I. van Rossum: Diagnosis and prognosis of Alzheimer's disease in subjects with mild cognitive impairment (November 28, 2013)
38. E.I.S. Møst: Circadian rhythm deterioration in early Alzheimer's disease and the preventative effect of light (December 3, 2013)
39. M.A.A. Binnewijzend: Functional and perfusion MRI in dementia (March 21, 2014)
40. H. de Waal: Understanding heterogeneity in Alzheimer's disease: A neurophysiological perspective (April 25, 2014)
41. W. Jongbloed: Neurodegeneration: Biochemical signals from the brain (May 8, 2014)
42. E.L.G.E. Poortvliet-Koedam: Early-onset dementia: unraveling the clinical phenotypes (May 28, 2014)
43. A.C. van Harten: The road less traveled: CSF biomarkers for Alzheimer's disease: predicting earliest cognitive decline and exploring microRNA as a novel biomarker source (November 7, 2014)
44. A.M. Hooghiemstra: Early-onset dementia: With exercise in mind (December 3, 2014)
45. L.L. Sandberg-Smits: A cognitive perspective on clinical manifestations of Alzheimer's disease (March 20, 2015)
46. F.H. Duits: Biomarkers for Alzheimer's disease, current practice and new perspectives (April 1, 2015)
47. S.M. Adriaanse: Integrating functional and molecular imaging in Alzheimer's disease (April 7, 2015)
48. C. Möller: Imaging patterns of tissue destruction - Towards a better discrimination of types of dementia (May 1, 2015)
49. M. Del Campo Milán: Novel biochemical signatures of early stages of Alzheimer's disease (June 19, 2015)
50. M.R. Benedictus: A vascular view on cognitive decline and dementia: relevance of cerebrovascular MRI markers in a memory clinic (January 20, 2016)
51. M.D. Zwan: Visualizing Alzheimer's disease pathology: implementation of amyloid PET in clinical practice (March 3, 2016)
52. E. Louwersheimer: Alzheimer's disease: from phenotype to genotype (June 21, 2016)
53. W.A. Kudrop: The frontal lobe syndrome: a neuropsychiatric challenge (September 23, 2016)

54. E.G.B. Vijverberg: The neuropsychiatry of behavioral variant frontotemporal dementia and primary psychiatric disorders: similarities and dissimilarities (September 22, 2017)
55. F.T. Gossink: Late-onset behavioral changes differentiating between bvFTD and psychiatric disorders in clinical practice (April 20, 2018)
56. M.A. Engels: Neurophysiology of dementia (May 18, 2018)
57. S.C.J. Verfaillie: Neuroimaging in subjective cognitive decline: incipient Alzheimer's disease unmasked (September 12, 2018)
58. M. ten Kate: Neuroimaging in predementia Alzheimer's disease (September 13, 2018)
59. H.F.M. Rhodius-Meester: Optimizing use of diagnostic tests in memory clinics: the next step (September 24, 2018)
60. E.A.J. Willemse: Optimizing biomarkers in cerebrospinal fluid: how laboratory reproducibility improves the diagnosis of Alzheimer's disease (October 18, 2018)
61. E. Konijnenberg: Early amyloid pathology – Identical twins, two of a kind? (June 25, 2019)
62. A.E. Leeuwis: Connecting heart and brain: vascular determinants of cognitive impairment and depressive symptoms (July 2, 2019)
63. J. den Haan: Imaging the retina in Alzheimer's disease (September 12, 2019)
64. A.C. van Loenhoud: Cognitive reserve in Alzheimer's disease: a perspective on the flourishing and withering of the brain (September 18, 2019)
65. R.J. Jutten: Capturing changes in cognition: refining the measurement of clinical progression in Alzheimer's disease (September 20, 2019)
66. N. Legdeur: Determinants of cognitive impairment in the oldest-old (October 8, 2019)
67. R. Slot: Subjective cognitive decline: predictive value of biomarkers in the context of preclinical Alzheimer's disease (November 14, 2019)
68. N. Scheltens: Understanding heterogeneity in Alzheimer's disease – a data-driven approach (December 17, 2019)
69. L. Vermunt: Secondary prevention for Alzheimer's disease – timing, selection and endpoint of clinical trials (March 13, 2020)
70. L.M.P. Wesselman: Lifestyle and brain health – exploring possibilities of an online intervention in non-demented elderly (April 1, 2020)
71. I.S. van Maurik: Interpreting biomarker results in patients with mild cognitive impairment to estimate prognosis and optimize decision making (May 12, 2020)
72. E. Dicks: Gray matter covariance networks in Alzheimer's disease: Edging towards a better understanding of disease progression (September 9, 2020)
73. J.J. van der Zande: A sharper image of dementia with Lewy bodies: The role of imaging and neurophysiology in DLB, and the influence of concomitant Alzheimer's disease pathology (September 21, 2020)
74. I. van Steenoven: Cerebrospinal fluid biomarkers in dementia with Lewy bodies – towards a biological diagnosis (September 22, 2020)
75. N. Beker: Cognition in centenarians: Evaluation of cognitive health and underlying factors in centenarians from the 100-plus Study (October 2, 2020)
76. F.A. de Leeuw: Nutrition and metabolic profiles in Alzheimer's disease (November 30, 2020)
77. T. Timmers: Tau PET across the Alzheimer's disease continuum (December 2, 2020)

78. E.E. Wolters: Untangling tau pathology using PET (December 2, 2020)
79. A.S. Doorduijn: Nutrition, the unrecognized determinant in Alzheimer's disease (January 12, 2021)
80. A. de Wilde: Visualizing brain amyloid-β pathology: Toward implementation of amyloid imaging in daily memory clinic practice (March 17, 2021)
81. C. Groot: Heterogeneity in Alzheimer's disease: A multi-modal perspective (April 6, 2021) (Cum laude)
82. D. Bertens: The use of biomarkers in non-demented AD patients for clinical trial design and clinical practice (April 14, 2021)
83. J. Reimand: Discordance between amyloid-β PET and CSF biomarkers: Clinical and pathophysiological consequences (May 12, 2021)
84. L. Collij: The AMYPAD project: Toward the next stage in amyloid PET imaging (July 1, 2021)
85. C.T. Briels: Evaluation and implementation of functional cerebral biomarkers in Alzheimer's disease (September 15, 2021)
86. M.M.J. Linszen: Understanding hallucinations outside the context of psychotic disorders (September 27, 2021)
87. N. Tesi: The genetics of cognitively healthy centenarians (September 28, 2021)
88. L.M. Reus: Triangulating heterogeneity in dementia: An interaction between genetic variation, biological correlates and clinical symptoms (September 28, 2021)
89. R. Babapour-Mofrad: The use of biofluid biomarkers in dementia: Implementation in clinical practice and breaking new grounds (October 8, 2021)
90. E.H. Thijssen: Blood-based biomarkers for Alzheimer's disease: assay development and clinical validation (November 15, 2021) (Cum laude)
91. A.C. Baakman: Innovation in cholinergic enhancement for Alzheimer's disease (November 17, 2021)
92. I.M.W. Verberk: Blood-based biomarkers for Alzheimer's disease coming to fruition (December 8, 2021) (Cum laude)
93. J.F. Leijenaar: Streamlining VCI: Heterogeneity and treatment response (January 11, 2022)
94. B.D.C. Boon: Deciphering neuropathological heterogeneity in Alzheimer's disease: Beyond plaques and tangles (January 14, 2022)
95. E.H. Singleton: The behavioral variant of Alzheimer's disease: A clinical, neuroimaging and neuropathological perspective (June 2, 2022)
96. M. van de Beek: The many faces of dementia with Lewy bodies: Impact on clinical manifestation, disease burden and disease progression (September 9, 2022)
97. W. Pelkmans: Imaging neurodegeneration across the Alzheimer's disease continuum (October 3, 2022)
98. J.L. Ebenau: From subjective to objective cognitive decline: Using biomarkers to understand the earliest stages of Alzheimer's disease (October 4, 2022)
99. H. Uluğut Erkoyun: Challenges in phenotyping the frontotemporal dementia subtypes (November 8, 2022)
100. **M.A. Dubbelman: For good measure: Detecting and defining changes in everyday functioning in Alzheimer's disease and related disorders (November 15, 2022)**

Portfolio

Courses	Year(s)	ECTS
Friday afternoon PhD program (Alzheimer Center Amsterdam)	2018–2022	8.0
BROK course	2019	1.5
Mulitlevel analyse (EpiDM)	2019	2.0
Research integrity	2020	2.0
Teaching		
Workgroup teacher, Regressietechnieken (EpiDM)	2019	2.0
Workgroup tutor, Minor Hot Topics in Neurology & Psychiatry	2020	2.0
Teacher's assistant, Age & Age-related Disorders	2020	5.5
Supervision of 1 Bachelor's and 7 Master's students	2020–2022	
International conferences		
Alzheimer's Association International Conference, online Oral presentation, poster presentation	2020	2.0
Clinical Trials on Alzheimer's disease, online Oral presentation	2020	2.0
Alzheimer's Association International Conference, Denver, USA Oral presentation, poster presentations	2021	2.0
Clinical Trials on Alzheimer's Disease, Boston, USA Poster presentation	2021	
Alzheimer's Association International Conference, San Diego, USA Oral presentation	2022	2.0
National conferences		
Analysis and interpretation of epidemiological data: New insights and opportunities, Amsterdam	2019	0.3
VU Science Exchange Day, Amsterdam	2019	0.2
Alzheimer Nederland Mix & Match, Utrecht	2019, 2022	
Dementie Update, Amsterdam	2019, 2020	
Amsterdam Neuroscience Annual Meeting, online	2020	0.3
NL Research Integrity Network 'INSPIRE' symposium, Utrecht Invited speaker	2020	
Other		
Clinical work as a neuropsychologist, Alzheimer Center Amsterdam	2019–2022	10.0
Research fellowship at Massachusetts General Hospital and Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA	2021	8.0
	Total	**49.8**

Dankwoord

Dit is het dan: na ruim 250 pagina's inleidingen, tabellen en referenties, geschreven met bloed, zweet en veel tranen, volgt nu het hoofdstuk dat mensen daadwerkelijk lezen. Dit proefschrift had niet tot stand kunnen komen zonder de hulp en steun van velen. Allereerst wil ik de patiënten en deelnemers, maar natuurlijk ook hun naasten, bedanken. Met hun deelname aan de vele onderzoeken waarop dit werk berust hebben zij een bijdrage van onkenbare waarde geleverd aan het streven naar een wereld zonder alzheimer. Hieronder wil ik daarnaast een aantal mensen in het bijzonder danken.

Prof.dr. Scheltens, beste Philip, als ik je drie halve zinnen gaf, begreep je niet alleen waar ik mee bezig was, maar kon je ook precies het advies geven dat ik nodig had om verder te kunnen. Daardoor waren onze kort maar krachtige overleggen enorm waardevol. Bedankt voor het vertrouwen dat je vanaf het begin in mij hebt gehad.

Dr. Sikkes, lieve Sietske, met jouw onuitputtelijke bron aan creatieve ideeën ben je echt een ware inspiratie. Je bent een onderzoeker in hart en nieren—en af en toe, geheel passend, misschien een beetje een verstrooide professor en ik heb dan ook heel veel van je geleerd. Met jou als begeleider heb ik daarnaast ook met veel onderzoekers kunnen samenwerken en mooie carrièremogelijkheden gehad. Ik ben je daarvoor zeer dankbaar en kijk ernaar uit om met je te blijven werken.

Dr. Marshall, dear Gad, thank you for having me at CART. Even though things went a bit differently from how we both hoped, I would not have missed the opportunity to work with and learn from you. I am very excited to continue our collaboration in the (near) future.

Prof.dr. van der Flier, beste Wiesje, ik heb grote bewondering voor hoe je bij zoveel projecten betrokken bent en toch overal inhoudelijke en doortastende input voor geeft. Bedankt voor het meedenken en -schrijven aan mijn projecten en voor de wetenschappelijk bezielende leiding van ons mooie centrum.

Prof.dr. Terwee, beste Caroline, we hebben in een aantal projecten met elkaar samengewerkt waarbij je altijd een kritische blik wierp op de klinimetrische aspecten van deze onderzoeken. Ik heb in onze overleggen veel van je geleerd; grote dank daarvoor.

Leden van de leescommissie, members of the doctorate committee: prof.dr. Yolande Pijnenburg, prof.dr. Bernard Uitdehaag, prof.dr. Hanneke Hulst, prof. dr. Dorene Rentz, prof.dr. Sarah Tomaszewski Farias en dr. Wieneke Mokkink, hartelijk bedankt voor de tijd die u allen heeft gestoken in het beoordelen van dit proefschrift. I would like to thank you very much for your time and effort in critically reviewing this thesis.

Mijn dank aan Alzheimer Nederland en het Amsterdam UMC Jonge Talentfonds is groot, want mede door de beurzen die zij mij hebben gegeven, heb ik tijdens mijn PhD naar Boston kunnen gaan. I am very grateful to Boston, Cambridge, Mass General Brigham, CART and every kind and special person I met and spent time with in the six months I lived in Boston (and after). I experienced in Boston even more so than back home how important it is to do things outside of work.

Lieve Merike, samen hebben we als twee belangrijke bouwstenen van “team Sikkes” veel doorstaan; van het werken aan artikelen, vertalingen checken, scoren voor en overleggen met de industrie tot het (eindeloos) heen en weer mailen met data scientists. Bedankt voor al je hulp, goede ideeën, klaagmomentjes, de gifjes die je in Teams stuurde vlak voordat ik mijn scherm moest delen, en niet te vergeten: bedankt voor de geweldige wijnproeverijen. Roos, bedankt voor je begeleiding en ideeën bij allerlei projecten. Je hebt mij geïnspireerd om zoveel mogelijk uit mijn PhD-traject te halen en hebt me hierin af en toe ook ‘voorgevolgd’. Ik zie je straks graag weer bij the Owl’s Nest of Lamplighter. Lois, ik ben onder de indruk van hoe je zo lang twee zware en bijna fulltimebanen hebt gecombineerd. Wat goed dat je nu iets doet dat beter bij je past—en wat was het fijn iemand te hebben om smart mee te delen. Linda W, Nina, Quirien, Josephine, Merel, Linda L, Rosanne, Angela, Sanne, Elke, Evelien, Lisa, Matthijs, Selina en Sophie, veel dank voor jullie kritische vragen en scherpe opmerkingen tijdens onze onderzoeksoverleggen. Elke, Lisa-Marie, Niels, Caroline, Melina, Marie-Christine, Marijn, Nadia en Camila, het was leuk om als (gedeeltelijk) begeleider met jullie zulke uiteenlopende projecten te hebben gedaan. Dank jullie wel, danke schön, grazie mille, muchas gracias.

Bunker 3, mijn thuis op werk. Mijn tijd in het AC was lang niet zo leuk en bijzonder geweest zonder onze te gekke bunker. Promoveren is niet altijd makkelijk en daardoor is een plek waar je je even kan omdraaien om te spuwen (of te luisteren hoe het bij anderen is en je realiseren dat we allemaal tegen ongeveer dezelfde dingen aanlopen) des te belangrijker. Lieve Marleen, Jori, Jarith, Jay, Hülya, Marijn,

Kirsten, Suzan, MP, Lisa, Jochum, Suzie en Janna, bedankt voor de vele mooie herinneringen aan onze tijd. Jori, je bent echt een fantastisch mens. Ik ben blij dat je straks naast me zit!!1

Alle andere geweldige collega's uit het Alzheimercentrum; het is al vaak geschreven, maar we zijn met te veel om jullie allemaal persoonlijk te noemen. Bedankt voor de fijne sfeer in ons centrum, de gezellige lunches en borrels, koffies en (oh ja) ook de wetenschappelijke discussies. Een bijzondere shoutout voor Astrid, Karin, Diana, Mardou en de volledige bezetting van het projectbureau over de jaren. Heleen en Emma, we hebben ons toch maar mooi opgewerkt vanuit het raamloze stagiairehok; ik ben trots op de harde kern van de 'interndisciplinarity'. Sterre, wat fijn dat er iemand is die de PhD memes net zo kan waarderen als ik. Ik hoop dat we die nog lang kunnen blijven uitwisselen.

Piet, met jou gesapro (voor wie geen Marie-Claire is: gezellig samen promoveren) was me een waar genoegen. Ik kijk met weemoed terug op de treinritten om 7:26, het nachtbraken in VUmc en natuurlijk het racen of schieten in die vijf minuten dat we zo nu en dan niet aan het werk waren. Emma en Ineke, bedankt ook voor de steun in de ESAD-intervisiegroep. Denis, er is denk ik niemand anders waarmee ik gesprekken kan voeren over politiek of geschiedenis, maar tegelijkertijd als een idioot kan klinken. Bedankt voor je oprechte interesse in wat ik doe. Alyssa (Lyssos), bedankt voor de vele gezellige avonden. Ik kijk ernaar uit om snel weer bij te praten. 'Oude' vrienden van het CGU: Marije, Jolien en Coen, wat leuk dat we elkaar na al die jaren nog steeds met enige regelmaat zien, in Leuven, Leiden, Utrecht, of waar dan ook. Ilona en Josh, Dore en Adam, onze zondagse Discordsessies in coronatijden en daarna waren een zeer welkome afleiding, bedankt daarvoor. Inês, um grande beijinho.

Mijn eeuwige dank gaat verder uit naar: Andy Field, mijn medebestuursleden van jongNVN, koffie, the 'Gram, miss Rona, de 'High impact PhD memes', MARINA, Parks and Recreation, mijn continu groeiende Whatsapp-stickercollectie (nieuwe altijd welkom), stackoverflow, RollerCoaster Tycoon, wijn, Duolingo, PlayStation, NS, Netflix, de broodbelegcombinatie spicy mango hummus, oude kaas, vegetarische grilworst en sambal, RStudio, de obductofoon, de subreddit r/EarthPorn, Delta (je boy heeft Silver Medallion status), Alicia Keys, de torenhoge publicatiekosten (voor mijn reeds gepubliceerde papers bij elkaar ruim €16.000) voor het letterlijk in een iets mooiere PDF gieten van ons harde werk, Café Sushi, reviewer #2, Wordle,

ISTAART (zeg: ijstaart), de ‘random videos I watch at 2am’ YouTube video’s, mijn iMac en, last but not least: the tidyverse, R’s real MVP.

Lieve tante Stance, bedankt voor de gezellige avonden met goede discussies waarin je mij, mijn zus en mijn neven en nichten bij elkaar bracht. Ik hoop jullie allemaal snel weer een keer te zien. Lisa, bedankt voor het ontwerpen van de omslag van mijn proefschrift. Lieve Josien, je bent een soort bonustante; bedankt voor je interesse in wat ik doe. Tantes Karen, Judith en Rineke en oom Pieter, bedankt voor jullie interesse in mijn werk.

Laura, mijn allerliefste zusje, ik ben enorm trots op hoe je je eigen pad in het leven bewandelt en zou willen dat ik af en toe wat meer van het leven genoot zoals jij. Bedankt dat je naast me op het podium wil staan. Lieve papa, lieve mama, bedankt voor jullie onvoorwaardelijke steun en vertrouwen in al mijn keuzes, van Portugees via psychologie en PhD tot postdoc aan de andere kant van de oceaan. We hoeven het niet hardop te zeggen want het is een gegeven, maar ik hou van jullie.

"Don't underestimate me, 'cause one day you're gonna see you're in a losing battle, babe, you'll never stop me be me 'cause whatever you give life, you're gonna get back"

— MARINA, 'Venus Fly Trap' (2021)

Abbreviations

AA, Alzheimer's Association; Aβ, amyloid beta; AD, Alzheimer's disease; ADC, Amsterdam Dementia Cohort; ADCS ADL-PI, Alzheimer's Disease Cooperative Study Activities of Daily Living Prevention Instrument; ADNI, Alzheimer's Disease Neuroimaging Initiative; ADRD, Alzheimer's disease and related disorders; A-IADL-Q, Amsterdam Instrumental Activities of Daily Living Questionnaire; A-IADL-Q-SV, Amsterdam Instrumental Activities of Daily Living Questionnaire Short Version; APOE, apolipoprotein; AUC, area under the curve; AVLT, auditory verbal learning task; A4, Anti-Amyloid Treatment in Asymptomatic Alzheimer's; B, estimate; BMI, body mass index; CAMCOG, Cambridge Cognitive Examinations; CAT, computerized adaptive testing; CatchCog, Capturing Changes in Cognition; CDR, Clinical Dementia Rating; CFA, confirmatory factor analysis; CFI, comparative fit index; CI, confidence interval; COSMIN, consensus-based standards for the selection of health measurement instruments; COST-A, Cognitive Online Self-Test Amsterdam; CSF, cerebrospinal fluid; CTT, classical test theory; DIA-S, Depression in Old Age Scale; DIF, differential item functioning; DLB, dementia with Lewy bodies; ECog, Everyday Cognition; EDTA, ethylenediamine tetraacetic acid; EEG, electroencephalography; ELISA, enzyme-linked immunosorbent assay; EMIF, European Medical Information Framework; EPAD(-LCS), European Prevention of Alzheimer's Disease (Longitudinal Cohort Study); FAB, Frontal Assessment Battery; FAQ, Functional Activities Questionnaire; FDA, Food and Drug Administration; FLAIR, fluid attenuated inversion recovery; FTD, frontotemporal dementia; GA, global atrophy; GDS, Geriatric Depression Scale; HABS, Harvard Aging Brain Study; HSD, honest significant difference; IADL, instrumental activities of daily living; ICC, intraclass correlation coefficient; ICF, International Classification of Functioning, Disability and Health; INSIGHT, Investigation of Alzheimer's predictors in subjective memory complainers; IQCODE, Informant Questionnaire for Cognitive Decline in the Elderly; IQR, interquartile range; IRT, item response theory; ISCED, international standard classification of education; LCLMM, latent class linear mixed model; LDST, letter–digit substitution test; LEARN, Longitudinal Evaluation of Amyloid Risk and Neurodegeneration; LMM, linear mixed model; M, median; MC, Monte Carlo; MIC, minimally important change; MCI, mild cognitive impairment; MD *(**chapter 4.2**)*, mild dementia; MD *(**chapter 8.1**)*, more difficulty; MMD, much more difficulty; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; MR, magnetic resonance; MRI, magnetic resonance imaging; MSDR, mean-to-standard deviation ratio; MTA, medial temporal atrophy; NACC, National Alzheimer's

Coordinating Center; NC, normal cognition; ND, no difficulty; NIA, National Institute on Aging; OLR, ordinal logistic regression; OR, odds ratio; PA, parietal atrophy; PET, positron emission tomography; PiB, Pittsburgh compound-B; PROM, patient-reported outcome measure; PVC, partial volume corrected; RCFT, Rey Complex Figure Test; RMSEA, root mean standard error of approximation; SAMS, Software Architecture for Mental health Self-management; SCD, subjective cognitive decline; SD, standard deviation; SDC, smallest detectable change; SEM, standard error of measurement; SMD, somewhat more difficulty; SUVr, standardized uptake volume ratio; TICS-m, modified Telephone Interview for Cognitive Status; TMT, Trail Making Test; UN, unable to perform; VAT, Visual Association Test; VOSP, Visual Object and Space Perception; WAIS, Wechsler Adult Intelligence Scale; ZBI, Zarit Burden Interview.

Author affiliations

André Aleman: Department of Neurosciences, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Rebecca E. Amariglio: Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States of America; Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, **MA, United States of America**

Bart N.M. van Berckel: Department of Radiology and Nuclear Medicine, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Laura J.E. Brown: Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Science Centre, Manchester, United Kingdom

Marina Bruderer-Hofstetter: School of Health Professions, Institute of Physiotherapy, Zurich University of Applied Sciences, Winterthur, Switzerland; Department of Health Sciences and Medicine, University of Lucerne, Lucerne, Switzerland

Rachel F. Buckley: Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States of America; Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States of America; Melbourne School of Psychological Sciences, **University of Melbourne, Melbourne, Australia**

Michael Caelers: KPMG Advisory N.V., Amstelveen, The Netherlands

Nancy Donovan: Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States of America; Department of Psychiatry, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, United States of America

Mark A. Dubbelman: Alzheimer Center Amsterdam, Department of Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The Netherlands; Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands

Bruno Dubois: Department of Neurology, Institut de la Mémoire et de la Maladie d'Alzheimer (IM2A) of the Pitié-Salpêtrière Hospital; ARAMIS, Sorbonne University, Inria de Paris, Institut du cerveau et de la moelle épinière (ICM), Paris, France

Stéphane Epelbaum: Department of Neurology, Institut de la Mémoire et de la Maladie d'Alzheimer (IM2A) of the Pitié-Salpêtrière Hospital; ARAMIS, Sorbonne University, Inria de Paris, Institut du cerveau et de la moelle épinière (ICM), Paris, France

David Facal: Department of Developmental Psychology, University of Santiago de Compostela, A Coruña, Spain

Wiesje M. van der Flier: Alzheimer Center Amsterdam, Department of Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The Netherlands; Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands; Department of Epidemiology and Biostatistics, Amsterdam **UMC, Amsterdam, The Netherlands**

Jennifer R. Gatchel: Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States of America

John E. Harrison: Alzheimer Center Amsterdam, Department of Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The Netherlands; Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands; Metis Cognition Ltd, Wiltshire, United Kingdom; Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

Hein P.J. van Hout: Department of General Practice and Medicine for Older Persons, Amsterdam Institute for Public Health Research, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Keith A. Johnson: Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States of America; Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States of America

Hanna Jokinen: Clinical Neurosciences, Neurology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Finland

Frank-Jan de Jong: Department of Neurology, Erasmus Medical Center, Rotterdam, The Netherlands

Roos J. Jutten: Alzheimer Center Amsterdam, Department of Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The Netherlands; Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands; Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, United States of America

Florian Koehn: Geriatrische Klinik St. Gallen, St. Gallen, Switzerland

Athene Lee: Butler Hospital, Warren Alpert Medical School of Brown University, Providence, Rhode Island, United States of America

Afina W. Lemstra: Alzheimer Center Amsterdam, Department of Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The Netherlands; Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands

Iracema Leroi: Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Science Centre, Manchester, United Kingdom

Cristina Lojo-Seoane: Department of Developmental Psychology, University of Santiago de Compostela, A Coruña, Spain

Gad A. Marshall: Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States of America; Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States of America

André Meichtry: School of Health Professions, Institute of Physiotherapy, Zurich University of Applied Sciences, Winterthur, Switzerland

Vuk Milošević: Clinic of Neurology, Clinical Center Niš, Niš, Serbia

José Luís Molinuevo: BarcelonaBeta Brain Research Center, Pasqual Margall Foundation, Barcelona, Spain

Thomas Münzer: Geriatrische Klinik St. Gallen, St. Gallen, Switzerland; Department of Geriatrics and Aging Research, University Hospital and University of Zurich, Zurich, Switzerland

Karin Niedermann: School of Health Professions, Institute of Physiotherapy, Zurich University of Applied Sciences, Winterthur, Switzerland

Kathryn V. Papp: Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States of America; Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States of America

Arturo X. Pereiro Rozas: Department of Developmental Psychology, University of Santiago de Compostela, A Coruña, Spain

Merel C. Postema: Alzheimer Center Amsterdam, Department of Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The Netherlands; Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands

Michael J. Properzi: Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States of America

Dorene M. Rentz: Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States of America; Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States of America

Craig Ritchie: University of Edinburgh, Edinburgh, United Kingdom

Stephen Salloway: Butler Hospital, Warren Alpert Medical School of Brown University, Providence, Rhode Island, United States of America

Justin Sanchez: Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, United States of America; Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, United States of America

Gonzalo Sánchez-Benavides: Barcelonaβeta Brain Research Center, Pasqual Margall Foundation, Barcelona, Spain

Benjamin D. Schalet: Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, United States of America

Reisa A. Sperling: Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, United States of America; Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America

Gemma Stringer: Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Science Centre, Manchester, United Kingdom

Philip Scheltens: Alzheimer Center Amsterdam, Department of Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The Netherlands; Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands

Aaron P. Schultz: Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States of America

Sietske A.M. Sikkes: Alzheimer Center Amsterdam, Department of Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The Netherlands; Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands; Department of Clinical, Neuro- & Developmental Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; Department of Epidemiology and Biostatistics, Amsterdam UMC, Amsterdam, The Netherlands

Caroline B. Terwee: Department of Epidemiology and Data Science, Amsterdam UMC, Amsterdam, The Netherlands

Charlotte E. Teunissen: Neurochemistry Laboratory, Department of Clinical Chemistry, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Sarah E. Tomaszewski-Farias: Department of Neurology, University of California Davis, Davis, CA, United States of America

Merike Verrijp: Alzheimer Center Amsterdam, Department of Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The Netherlands; Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands

Leonie N.C. Visser: Alzheimer Center Amsterdam, Department of Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The Netherlands; Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands; Division of Clinical Geriatrics, Center for Alzheimer Research, Department of Neurobiology, Care Sciences and Society, Karolinska Institute, Stockholm, Sweden

Pieter Jelle Visser: Alzheimer Center Amsterdam, Department of Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The Netherlands; Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands; Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastrict University, Maastricht, The Netherlands

Marissa D. Zwan: Alzheimer Center Amsterdam, Department of Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The Netherlands; Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands

Stelios Zygouris: School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece & Network Aging Research, Heidelberg University, Heidelberg, Germany

About the author

Mark Dubbelman was born on August 27th, 1993, in Utrecht, The Netherlands. In 2011, he enrolled in Portuguese Language and Culture at Utrecht University. Upon his return from a semester at the University of Coimbra, Portugal, Utrecht University discontinued the program. Students already in the program were allowed to complete their studies at Leiden University, where Mark simultaneously enrolled in Psychology. He finished his Psychology program at Utrecht University in 2016, but not before studying at the University of New Hampshire, United States, for one semester. He obtained his Master's degree in Clinical Neuropsychology at VU University Amsterdam in 2018.

During his first research internship in the Department of Psychiatry of the Utrecht University Medical Center, his interest in doing research grew, which led him to pursue a second research internship at the Alzheimer Center Amsterdam in the team of dr. Sikkes. He worked as a PhD student under supervision of dr. Sikkes and prof. Scheltens from 2018 to 2022, and he spent six months at Brigham and Women's Hospital, Harvard Medical School in Boston, United States. Harnessing years of experience working at a coffee shop in Utrecht central station, Mark dedicated himself to cleaning the Alzheimer Center Amsterdam's new coffee machine. He also serves as a member of the junior executive committee of the Dutch Neuropsychological Society and as student/postdoc member of the executive committee of the Subjective Cognitive Decline Professional Interest Area of the International Society to Advance Alzheimer's Research and Treatment.

"Not all those who wander are lost"

— J.R.R. Tolkien, 'The Fellowship of the Ring' (1954)

“fim e ponto final”

— José Saramago, ‘O Ano da Morte de Ricardo Reis’ (1984)