

# Determining the Minimal Important Change of Everyday Functioning in Dementia

## Pursuing Clinical Meaningfulness

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

### Correspondence

Dr. Dubbelman  
m.dubbelman@amsterdamumc.nl

*Neurology*® 2022;99:e954-e964. doi:10.1212/WNL.000000000000200781

## Abstract

### Background and Objectives

Decline in everyday functioning is a key clinical change in Alzheimer 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 effect 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 patients with ADRD 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 (IADL) Questionnaire. In the second, quantitative 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 ( $N = 1,629$ ;  $62.4 \pm 9.5$  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 ( $N = 230$ ;  $64.3 \pm 7.7$  years; 39% female; 60% dementia diagnosis) were followed for 1 year, 102 (45%) of whom showed a decline larger than the MIC, after a mean of  $6.7 \pm 3.5$  months. Patients with a dementia diagnosis and more atrophy of the medial temporal lobe had larger odds (odds ratio [OR] = 3.4, 95% 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

### MORE ONLINE

#### Podcast

[NPub.org/Podcast9833](https://npub.org/podcast9833)

From the Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, The Netherlands (M.A.D., M.V., M.C.P., F.G., V.T., W.M.F., P.S., S.A.M.S.); Department of Epidemiology and Data Science, Amsterdam UMC, The Netherlands (C.B.T., W.M.F.); Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston (R.J.J.); Department of Radiology and Nuclear Medicine, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, The Netherlands (F.B., B.N.M.B.); Institutes of Neurology and Healthcare Engineering, University College London, United Kingdom (F.B.); Neurochemistry Laboratory, Department of Clinical Chemistry, Amsterdam Neuroscience, Amsterdam UMC, Vrije Universiteit Amsterdam, The Netherlands (C.T.); and Faculty of Behavioural and Movement Sciences (S.A.M.S.), Clinical Developmental Psychology & Clinical Neuropsychology, Vrije Universiteit Amsterdam, the Netherlands.

Go to [Neurology.org/N](https://www.neurology.org/N) for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

The Article Processing Charge was funded by the authors.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

## Glossary

**ADRD** = Alzheimer disease and related disorders; **A-IADL-Q** = Amsterdam IADL Questionnaire; **A $\beta$**  = amyloid beta<sub>1-42</sub>; **GDS** = Geriatric Depression Scale; **IADLs** = instrumental activities of daily living; **IRT** = item response theory; **MCI** = mild cognitive impairment; **MIC** = minimal important change; **MMSE** = Mini-Mental State Examination; **MTA** = medial temporal atrophy; **OR** = odds ratio; **SCD** = subjective cognitive decline; **WAIS** = Wechsler Adult Intelligence Scale; **ZBI** = Zarit Burden Interview.

than the MIC. Our findings provide guidance in interpreting changes in IADL and may help evaluate treatment effects and monitor disease progression.

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

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, for 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.<sup>7,8</sup> Others have reasoned that clinical meaningfulness comprises prediction of future conversion from normal cognition to mild cognitive impairment (MCI) or dementia.<sup>9</sup> The first definition may overgeneralize and include changes due to noise, whereas 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.<sup>10,11</sup>

The MIC can be determined using anchors,<sup>12</sup> 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 describe the patient now, compared with 1 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.<sup>13</sup> An alternative can be found in a new systematic, qualitative approach<sup>14</sup> in which stakeholders (i.e., patients, caregivers, and clinicians) are asked to compare fictional patient summaries with different levels of impairment

in the area that is being measured. Thresholds are then placed at the first point where the stakeholders indicate that a difference is meaningful.<sup>14</sup> 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.<sup>15,16</sup> 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.

## Methods

Our study comprised 2 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 1 year and which factors were associated with surpassing the MIC threshold.

### Standard Protocol Approvals, Registrations, and Patient Consent

This 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 ([herenonderzoek.nl](https://herenonderzoek.nl)).<sup>17</sup> 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: A-IADL-Q

The A-IADL-Q is an adaptive questionnaire aimed at measuring functional impairment in early dementia.<sup>15</sup> 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.<sup>18,19</sup> The questionnaire consists of 70 items assessing cognitively complex everyday activities. The 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.<sup>16,19</sup> The T-scores have a mean of 50 and a SD of 10 in the memory clinic. Lower scores indicate more impairment.

## Materials: Vignettes

We created 18 vignettes using IRT item parameters that showed the most likely item responses at different total scores, that is, 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.<sup>20</sup> To obtain the optimal balance between distinguishable levels of functional impairment and small distances between the vignettes, they were placed 0.2 SDs apart. We created 6 reference vignettes were spread across the total score distribution, representing different base levels of 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 are given in the eMaterial and eTable 1, [links.lww.com/WNL/C83](https://www.lww.com/WNL/C83).

## Procedures

Survey respondents (both caregivers and clinicians) were randomly branched into 1 of 6 groups, each of which received a different "case" with a unique reference vignette. They were then shown 7 "comparison vignettes," which ranged from  $-8$  to  $+6$  points from the reference vignette.

Following previously outlined procedures,<sup>14</sup> we presented vignettes in pairs, with the reference vignette representing the patient's functioning "1 year ago" and each comparison vignette representing a new situation "now." Respondents judged whether the functioning "now" was better, worse, or the same as "1 year ago" (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.

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 2 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 4-point decrease is judged as meaningful, whereas a 6-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,<sup>21</sup> 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.

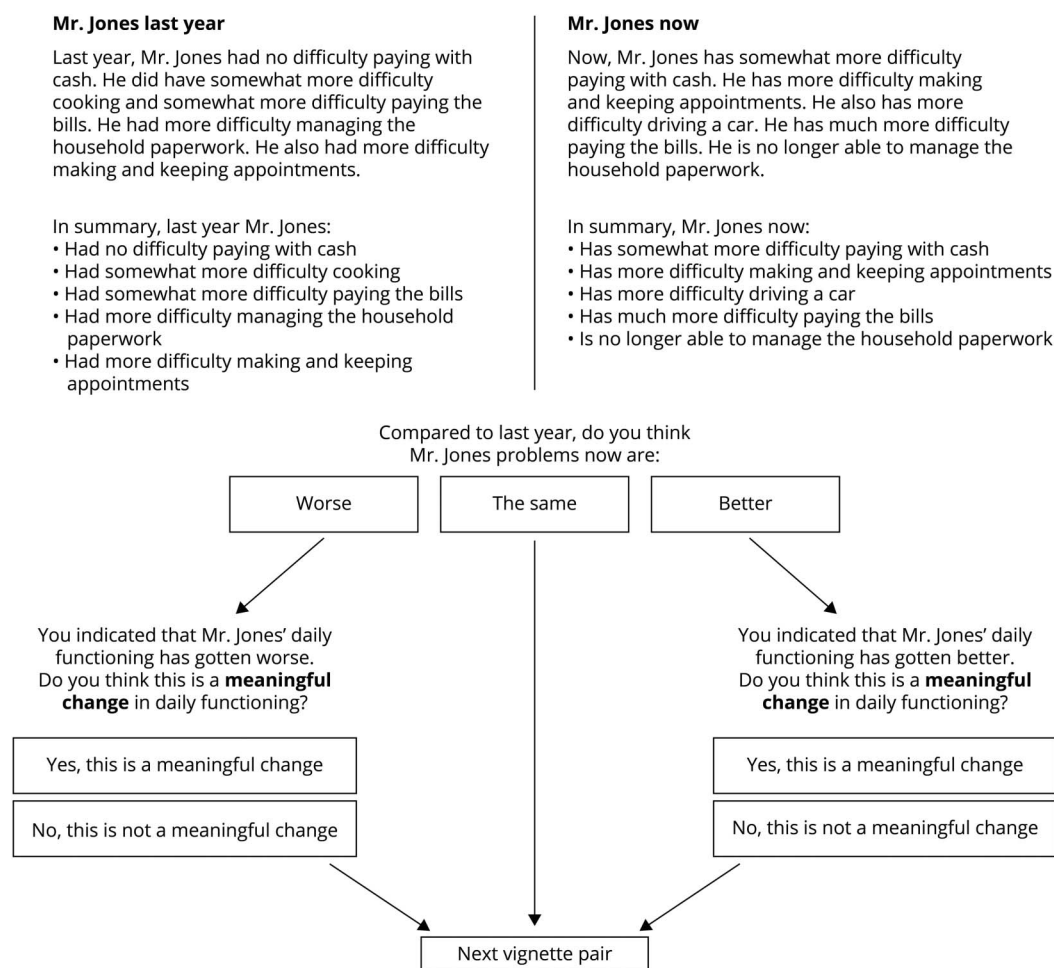
At the baseline screening visit, caregivers completed the A-IADL-Q, while the patients underwent a standard neuropsychological test battery. The screening visit also included a neurologic examination, brain MRI, and a lumbar puncture.<sup>21</sup> Diagnoses were made in a multidisciplinary consensus meeting in which the results from the screening visit were discussed.<sup>21</sup> Clinical diagnoses were made according to the criteria for subjective cognitive decline (SCD), MCI, dementia, Alzheimer disease, frontotemporal dementia, dementia with Lewy bodies, and vascular dementia.<sup>21</sup> Non-Alzheimer disease types of dementia were grouped to avoid small group sizes.

Caregivers were then invited to complete the A-IADL-Q from home at 4 follow-up waves: 3, 6, 9, and 12 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 think the patient declined from baseline and (2) how much burden they experienced from taking care of the patient. These 2 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

### Clinical Measures

A standardized neuropsychological assessment was performed at baseline and included the Dutch version of the

**Figure 1** Example Question From the MIC Survey

First, 2 vignettes are shown side-by-side, with one representing functioning of a fictional patient 1 year ago (the “reference vignette” on the top left, anchored in the example at  $T = 46$ ) and the other representing functioning now (the “comparison vignette” on the top right, anchored in the example at  $T = 42$ ). The respondent is asked to indicate whether they think the problems have worsened, remained the same, or improved from 1 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.

Auditory Verbal Learning Task<sup>22</sup> and the Visual Association Test,<sup>23</sup> to measure episodic memory. The Trail Making Test, Part B,<sup>24</sup> Wechsler Adult Intelligence Scale (WAIS) digit span backward,<sup>25</sup> letter fluency,<sup>26</sup> and Stroop Color-Word Task card III<sup>27</sup> were used to measure executive functioning. Attention and speed were measured using the Trail Making Test, Part A,<sup>24</sup> Stroop Color-Word Task card I,<sup>27</sup> the Letter Digit Substitution Test,<sup>28</sup> and the WAIS digit span forward.<sup>25</sup> Language tasks included the naming portion of the Visual Association Test<sup>23</sup> and the category fluency (animal naming) task.<sup>26</sup>

We calculated Z-scores for the neuropsychological domains: episodic memory, executive functioning, attention/speed of processing, and language. Before 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 SDs of the measures in the entire sample.

The Mini-Mental Examination (MMSE) was used as an indication of general cognitive performance, with higher scores representing better cognition.<sup>29</sup> The 15-item version of the Geriatric Depression Scale (GDS) was used as an indicator for depressive symptoms,<sup>29</sup> 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.<sup>30</sup>

#### Biological Measures

At baseline, patients underwent a standard MRI protocol on a 1.5 or 3 Tesla scanner.<sup>21</sup> 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,<sup>31</sup> the posterior atrophy scale,<sup>32</sup> the global cortical atrophy scale,<sup>33</sup> and the Fazekas scale<sup>34</sup> for white matter hyperintensities. Cerebral microbleeds were counted.

Amyloid beta<sub>1-42</sub> (A $\beta$ ) levels in CSF were measured using ELISA (Innogenetics-Fujirebio, Ghent, Belgium) at the Neurochemistry Laboratory.<sup>35</sup> We dichotomized amyloid status into negative or positive for AD based on our center's cutoff of <813 pg/mL.<sup>36</sup> We also computed the ratio between phosphorylated tau and A $\beta$ . A subset of participants underwent amyloid PET scans, using <sup>11</sup>C-Pittsburgh compound-B, <sup>18</sup>F-flutemetamol, <sup>18</sup>F-florbetapir, or <sup>18</sup>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 to 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 ABI130XL. Patients with either 1 or 2  $\epsilon$ 4 alleles were classified as APOE  $\epsilon$ 4 carriers.

## Statistical Analyses

To obtain MIC thresholds, we averaged individual thresholds separately for each of the 6 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 2.

In the clinical cohort, patients were divided into 3 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. The Tukey 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 with 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 disease genetic risk factors and amyloid biomarkers, and MRI. All factors were investigated individually, with adjustments for sex, education,

baseline age, and syndrome diagnosis (SCD, MCI, or dementia). Analyses were run in R version 4.1.1,<sup>37</sup> using the “nnet” package version 7.3-16 for the multinomial logistic regressions.<sup>38</sup>

## 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.

## Results

### Establishing the MIC

A total of 1,629 caregivers (mean age 62.4  $\pm$  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 (5 neurologists, 5 nurse specialists, 2 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 is given in eTable 2, [links.lww.com/WNL/C83](https://links.lww.com/WNL/C83). 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 ( $p < 0.001$ ). The average MIC threshold for decline was 2.4  $\pm$  1.0 points among all caregivers (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 with 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  $-2.0$ .

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  $\pm$  1.3 points (Table 1). Five clinicians detected a meaningful improvement, with an average threshold of 5.2  $\pm$  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.

**Table 1** Minimal Important Change Thresholds

| Group                        | N     | Decline           |                   | Improvement       |                   |
|------------------------------|-------|-------------------|-------------------|-------------------|-------------------|
|                              |       | Reached threshold | Average threshold | Reached threshold | Average threshold |
| <b>Caregivers</b>            | 1,629 | 1,599 (98.2)      | -2.4 ± 1.0        | 362 (22.2)        | +4.7 ± 1.3        |
| <b>Ms. Smith (T = 54)</b>    | 268   | 265 (98.9)        | -2.3 ± 1.0        | 101 (37.7)        | +4.9 ± 1.2        |
| <b>Mr. Jones (T = 50)</b>    | 260   | 257 (98.8)        | -2.6 ± 1.0        | 141 (54.2)        | +4.9 ± 1.3        |
| <b>Mr. Williams (T = 46)</b> | 284   | 283 (99.6)        | -2.1 ± 0.4        | 10 (3.5)          | +5.4 ± 1.4        |
| <b>Ms. Brown (T = 42)</b>    | 265   | 263 (99.2)        | -2.2 ± 0.8        | 65 (24.5)         | +4.0 ± 1.0        |
| <b>Ms. Johnson (T = 38)</b>  | 282   | 278 (98.6)        | -2.3 ± 1.0        | 29 (10.3)         | +4.2 ± 1.4        |
| <b>Mr. Garcia (T = 34)</b>   | 270   | 253 (93.7)        | -2.9 ± 1.6        | 16 (5.9)          | +4.4 ± 1.1        |
| <b>Clinicians</b>            | 13    | 13 (100.0)        | -2.0 ± 0.0        | 5 (38.5)          | +5.2 ± 1.1        |

Thresholds are displayed as mean ± SD.

### The MIC in Clinical Practice

We included 230 patients (64.3 ± 7.7 years, 39% female) in the clinical cohort. They had diagnoses of SCD (n = 37), MCI (n = 22), AD dementia (n = 81), non-AD dementia (n = 58), or a different diagnosis (n = 36). The 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 decreased. In subsequent analyses, we used the groups as defined at the patient's last completed visit. At the last visit assessment, 104 patients (45%) showed a meaningful decline, whereas 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 (mean 39.0 ± 30.0) for decline than patients who showed no meaningful change (19.3 ± 21.5; mean difference  $p < 0.001$ ) or meaningful improvement (12.1 ± 17.2; mean difference  $p < 0.001$ ). 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  $p < 0.001$ ) and patients who surpassed the MIC for improvement (15.7 ± 23.2; mean difference  $p < 0.001$ ).

Table 2 summarizes the number of patients who surpassed the MIC thresholds for decline and improvement. Overall, the proportion of patients who surpassed the MIC threshold for decline increased with each subsequent visit, whereas the group who showed no meaningful change remained relatively stable. Most patients passed the MIC thresholds consistently across all visits: Only 34 patients (14.8%) inconsistently passed the MIC thresholds, 12 of whom (35.3%) surpassed the MIC for decline initially but ended up not showing a meaningful change, and 10 of whom (29.4%) surpassed the MIC for improvement initially but ended up showing no meaningful change. A breakdown of the number of patients per diagnostic group who surpassed the MIC is given in eTable 3, [links.lww.com/WNL/C83](https://www.lww.com/WNL/C83). Table 3 summarizes 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

**Table 2** MIC per Follow-up Wave

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

Abbreviation: MIC = minimal important change.

An overview of the number of patients surpassing the MIC broken down by diagnostic group is given in the eMaterial, [links.lww.com/WNL/C83](https://www.lww.com/WNL/C83).

**Table 3** Time in Months Until the MIC Threshold Was Reached

| Group           | Decline    |             | Improvement |             |
|-----------------|------------|-------------|-------------|-------------|
|                 | n (%)      | Months      | n (%)       | Months      |
| 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 |

Abbreviations: AD = Alzheimer disease; MCI = mild cognitive impairment; MIC = minimal important change.

reach the MIC threshold for either decline or improvement, after correction for multiple comparisons.

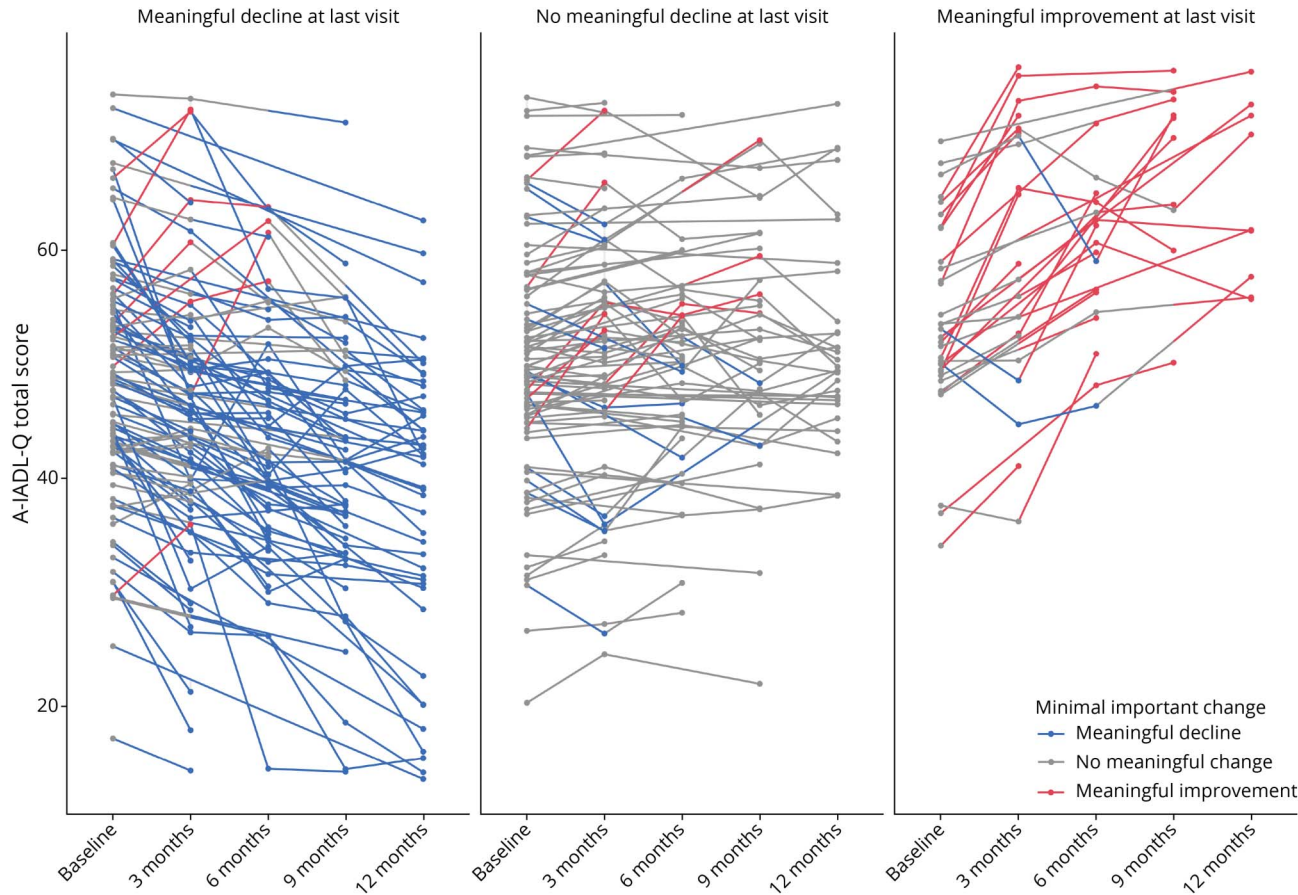
Figure 2 shows a visual representation of the number of patients at each follow-up wave that surpassed the thresholds for

meaningful decline and improvement. 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% CI [1.05–6.12],  $p = 0.039$ ) and less likely to surpass the MIC threshold for improvement (OR = 0.35, 95% CI [0.13–0.94],  $p = 0.037$ ), compared with patients with SCD. Patients with an MTA score of 1.5 were more likely to pass the MIC threshold for decline, compared with patients with an MTA of 0 or 0.5 (OR = 4.97, 95% CI [1.23–19.99],  $p = 0.024$ ). When the caregiver experienced a higher burden, that is, 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],  $p = 0.009$ ). 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; eTable 4, [links.lww.com/WNL/C83](https://www.lww.com/WNL/C83)).

## Discussion

In this study, we involved informal caregivers and clinicians of patients with ADRD to determine what amount of change in

**Figure 2** Change in IADL Functioning Over 1 Year, Stratified in Panels by Change From Baseline to the Last Visit



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 at each visit, compared with their baseline level of functioning. IADL = instrumental activities of daily living.



functional impairments constitutes a clinically meaningful change. We established thresholds for the MIC, 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 US Food and Drug Administration of disease-modifying treatments is contingent on the evidence of a meaningful benefit,<sup>6</sup> yet the interpretation of outcome measures remains difficult,<sup>39</sup> and there is considerable variability in how clinical meaningfulness is defined and investigated. Consensus is yet to be reached.<sup>40</sup> Some methods have methodological and conceptual limitations, including inadequate reliability and validity.<sup>14,41,42</sup> 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,<sup>13,42</sup> which renders them less reliable for investigating the clinical meaningfulness of changes. More importantly, neither method considers input from the target population, although only the individuals themselves, and those who are close to them, can indicate whether a change is impactful. Still, these methods are commonly used in dementia research<sup>8,43-46</sup> possibly because more elaborate qualitative approaches require extensive work. Our study is unique in the field of AD/DRD research in that it uses 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 seems 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. However, it is of interest that the threshold for minimal important improvement was higher when the level of functioning was better, compared with 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 because this is currently not a reality. With the rapid developments in drug development,<sup>47</sup> the exercise of establishing MIC thresholds on outcome measures may need to be repeated because our understandings of what is possible change.

The second part of our study was to apply the MIC thresholds in a real-life data set. Just under half of a nonselected group of memory clinic patients passed the MIC threshold for decline within 1 year and thus showed a meaningful decline, on average within approximately 7 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 MTA were more likely to show a meaningful decline than those with no atrophy. When the caregiver experienced a larger 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,<sup>18</sup> and that worse IADL performance was associated with atrophy in the medial temporal lobe.<sup>48</sup> Studies with other IADL measures related changes in IADL to disease stage,<sup>3</sup> amyloid burden,<sup>49</sup> and executive functioning,<sup>50</sup> 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 our 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 SD in the total score. Had we presented a smaller amount of change (e.g., a tenth of a SD), it is possible that the MIC thresholds 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, nonadherence 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 declined 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.

Acknowledgment

The authors thank all caregivers and clinicians who completed the survey for their time and contributions. S.A.M.S. and P.S. are co-developers of the Amsterdam Instrumental Activities of Daily Living Questionnaire, which is available free of charge for all public health and not-for-profit agencies and can be obtained via [alzheimercentrum.nl/professionals/amsterdam-iadl](http://alzheimercentrum.nl/professionals/amsterdam-iadl).

Study Funding

This study was funded by public-private funding from Health~Holland, Topsector Life Sciences & Health (PPPallowance; LSHM20084-SGF, project DEFEAT-AD, LSHM19051, project OTAPA), and the NIH, as well as license fees from Green Valley, VtV Therapeutics, Alzheon, Vivoryon, and Roche, and honoraria from Boehringer and Toyama of which SAMS was the recipient. All funding is paid to her institution.

Disclosure

M.A. Dubbelman, M. Verrijp, C.B. Terwee, R.J. Jutten, M.C. Postema, F. Barkhof, N.M. van Berckel, F. Gillissen, V. Teeuwen, and C. Teunissen report no disclosures relevant to the manuscript. W.M. van der Flier has received further funding from NWO, EU-FP7, EU-JPND, Alzheimer Nederland, CardioVasculair Onderzoek Nederland, Stichting Dioraphte, Gieskes-Strijbis fund, Stichting Equilibrio, Pasman Stichting, Biogen MA Inc, Boehringer Ingelheim, Life-MI, AVID, Roche BV, Fujifilm, and Combinostics. W.M. van der Flier holds the Pasman chair, has performed contract research for Biogen MA Inc. and Boehringer Ingelheim, is a consultant to Oxford Health Policy Forum CIC, Roche and Biogen MA Inc., and has been an invited speaker at Boehringer Ingelheim, Biogen MA Inc., Danone, Eisai, WebMD Neurology (Medscape). All funding is paid to her institution. P. Scheltens has acquired grant support (for the institution; Alzheimer Center Amsterdam) from GE Healthcare, Danone Research, Piramal, and MERCK. In the past 2 years, he has received consultancy/speaker fees (paid to the institution) from Lilly, GE Healthcare, Novartis, Sanofi, Nutricia, Probiobrug, Biogen, Roche, Avraham, and EIP Pharma. S.A.M. Sikkes was 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. Go to [Neurology.org/N](http://Neurology.org/N) for full disclosures.

Publication History

Received by *Neurology* November 25, 2021. Accepted in final form April 11, 2022. Submitted and externally peer reviewed. The handling editor was Linda Hershey, MD, PhD, FAAN.

Appendix Authors

| Name                    | Location                                                                                                                                             | Contribution                                                                                                                                                                               |
|-------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Mark A. Dubbelman, MSc  | Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands | Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data |
| Merike Verrijp, MSc     | Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands | Drafting/revision of the manuscript for content, including medical writing for content; study concept or design                                                                            |
| Caroline B. Terwee, PhD | Department of Epidemiology and Data Science, Amsterdam UMC, Amsterdam, The Netherlands                                                               | Drafting/revision of the manuscript for content, including medical writing for content; study concept or design; analysis or interpretation of data                                        |
| Roos J. Jutten, PhD     | Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States                                           | Drafting/revision of the manuscript for content, including medical writing for content; study concept or design; analysis or interpretation of data                                        |

## Appendix (continued)

| Name                               | Location                                                                                                                                                                                                                                                                                                                   | Contribution                                                                                                                                                                               |
|------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Merel C. Postema, PhD</b>       | Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands                                                                                                                                                                       | Drafting/revision of the manuscript for content, including medical writing for content                                                                                                     |
| <b>Frederik Barkhof, MD, PhD</b>   | Department of Radiology and Nuclear Medicine, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands; Institutes of Neurology and Healthcare Engineering, University College London, London, United Kingdom                                                                       | Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data                                                                 |
| <b>Bart N.M. Berckel, PhD, MD</b>  | Department of Radiology and Nuclear Medicine, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands                                                                                                                                                                              | Drafting/revision of the manuscript for content, including medical writing for content                                                                                                     |
| <b>Freek Gillissen</b>             | Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands                                                                                                                                                                       | Drafting/revision of the manuscript for content, including medical writing for content                                                                                                     |
| <b>Vivianne Teeuwen</b>            | Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands                                                                                                                                                                       | Drafting/revision of the manuscript for content, including medical writing for content                                                                                                     |
| <b>Charlotte Teunissen</b>         | Neurochemistry Laboratory, Department of Clinical Chemistry, Amsterdam Neuroscience, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands                                                                                                                                                               | Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data                                                                 |
| <b>Wiesje M. van de Flier, PhD</b> | Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands; Department of Epidemiology and Data Science, Amsterdam UMC, Amsterdam, The Netherlands                                                                               | Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data                          |
| <b>Philip Scheltens, PhD, MD</b>   | Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands                                                                                                                                                                       | Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data |
| <b>Sietske A.M. Sikkes, PhD</b>    | Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands; Faculty of Behavioural and Movement Sciences, Clinical Developmental Psychology & Clinical Neuropsychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands | Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data |

## References

- Scheltens P, De Strooper B, Kivipelto M, et al. Alzheimer's disease. *The Lancet*. 2021; 397(10284):1577-1590. doi: 10.1016/s0140-6736(20)32205-4
- Marshall GA, Aghajyan 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
- 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;12(1):138. doi: 10.1186/s13195-020-00706-2
- 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;6(4):223-227. doi: 10.14283/jpad.2019.30
- Liu KY, Schneider LS, Howard R. The need to show minimum clinically important differences in Alzheimer's disease trials. *The Lancet Psychiatry*. 2021;8(11):1013-1016. doi: 10.1016/s2215-0366(21)00197-8
- Food and Drug Administration. Early Alzheimer's disease: developing drugs for treatment - guidance for industry. 2018. Accessed June 3, 2021. <https://www.fda.gov/files/drugs/published/Alzheimer's-Disease-Developing-Drugs-for-Treatment-Guidance-for-Industry.pdf>
- Siemers E, Holdridge KC, Sundell KL, Liu-Seifert H. Function and clinical meaningfulness of treatments for mild Alzheimer's disease. *Alzheimers Dement (Amst)*. 2016;2:105-112. doi: 10.1016/j.dadm.2016.02.006
- 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*. 2017; 13(10):1098-1106. doi: 10.1016/j.jalz.2017.02.005
- Papp KV, Buckley R, Mormino E, et al. Clinical meaningfulness of subtle cognitive decline on longitudinal testing in preclinical AD. *Alzheimers Dement*. 2020;16(3): 552-560. doi: 10.1016/j.jalz.2019.09.074
- 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*. 2006;4:54. doi: 10.1186/1477-7525-4-54
- 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*. 2021;30(10):2729-2754. doi: 10.1007/s11136-021-02925-y
- Devi 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*. 2020;369:m1714. doi: 10.1136/bmj.m1714
- 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*. 2015;68(5):518-524. doi: 10.1016/j.jclinepi.2014.07.008
- 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*. 2017; 26(11):2961-2971. doi: 10.1007/s11136-017-1625-2
- Sikkes SA, de Lange-de Klerk ES, Pijnenburg YA, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. *Alzheimers Dement*. 2012;8(6):536-543. doi: 10.1016/j.jalz.2011.08.006
- 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
- 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
- Koster N, Knol DL, Uitdehaag BM, Scheltens P, Sikkes SAM. The sensitivity to change over time of the Amsterdam IADL Questionnaire((c)). *Alzheimers Dement*. 2015;11(10):1231-1240. doi: 10.1016/j.jalz.2014.10.006
- 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
- 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*. 2017;26(3):565-586. doi: 10.1007/s11136-016-1468-2
- 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
- Saan R, Deelman B. *De 15-Woordentest A en B (Een Voorlopige Handleiding)*. Afdeling Neuropsychologie; 1986.
- 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*. 2002; 73(2):126-133. doi: 10.1136/jnnp.73.2.126
- Tombaugh TN. Trail making test A and B: normative data stratified by age and education. *Arch Clin Neuropsychol*. 2004;19:203-214. doi: 10.1016/S0887-6177(03)00039-8
- Wechsler D. *WAIS-III Administration and Scoring Manual*. The Psychological Corporation; 1997.
- 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*. 2006; 12(1):80-89. doi: 10.1017/S1355617706060115

27. Stroop JR. Studies of interference in serial verbal reactions. *J Exp Psychol Gen*. 1992; 121(1):15.
28. 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*. 2006;28(6):998-1009. doi: 10.1080/13803390591004428
29. Folstein MF, Folstein SE, and McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3): 189-198.
30. Yesavage JA, Sheikh JL. Geriatric Depression Scale (GDS) - recent evidence and development of a shorter version. *Clin Gerontologist*. 1986;5(1-2):165-173. doi: 10.1300/J018v05n01\_09
31. Zarit SH, Reeve KE, Bach-Peterson J. Relatives of the impaired elderly: correlates of feelings of burden. *Gerontologist*. 1980;20(6):649-655. doi: 10.1093/geront/20.6.649
32. 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*. 1995;242(9):557-560.
33. Koedam EL, Lehmann M, van der Flier WM, et al. Visual assessment of posterior atrophy development of a MRI rating scale. *Eur Radiol*. 2011;21(12):2618-2625. doi: 10.1007/s00330-011-2205-4
34. 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):268-272. doi: 10.1159/000117270
35. 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*. 1987; 149(2):351-356. doi: 10.2214/ajr.149.2.351
36. 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*. 2010;56(2):248-253. doi: 10.1373/clinchem.2009.130518
37. Tijms BM, Willems EAJ, Zwan MD, et al. Unbiased approach to counteract upward drift in cerebrospinal fluid amyloid- $\beta$  1-42 analysis results. *Clin Chem*. 2018;64(3): 576-585. doi: 10.1373/clinchem.2017.281055
38. R: A Language and Environment for Statistical Computing. Version 4.1.0; 2021. Available at: R-project.org/.
39. Venables WN, Ripley BD. *Modern Applied Statistics with S*. 4th ed. Springer; 2002.
40. Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. *BMJ*. 1998;316(7132):690-693. doi: 10.1136/bmj.316.7132.690
41. 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
42. Wyrwich KW, Bullinger M, Aaronson N, Hays RD, Patrick DL, Symonds T. Estimating clinically significant differences in quality of life outcomes. *Qual Life Res*. 2005; 14(2):285-295. doi: 10.1007/s11136-004-0705-2
43. 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*. 2010;63(7):760-766.e1. doi: 10.1016/j.jclinepi.2009.09.009
44. 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*. 2008;61(8):827-831. doi: 10.1016/j.jclinepi.2007.10.022
45. Schrag A, Schott JM, Alzheimer's Disease Neuroimaging Initiative. What is the clinically relevant change on the ADAS-Cog? *J Neurol Neurosurg Psychiatry*. 2012;83(2): 171-173. doi: 10.1136/jnnp-2011-300881
46. 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
47. Insel PS, Weiner M, Mackin RS, et al. Determining clinically meaningful decline in preclinical Alzheimer disease. *Neurology*. 2019;93(4):e322-e333. doi: 10.1212/WNL.0000000000007831
48. 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
49. 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*. 2018;2675:109-116. doi: 10.1016/j.neurobiolaging.2018.11.018
50. Lilamand M, Cesari M, Cantet C, Payoux P, Andrieu S, Vellas B. 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*. 2018;66(10):1940-1947. doi: 10.1111/jgs.15497
51. Marshall GA, Rentz DM, Frey MT, Locascio JJ, Johnson KA, Sperling RA. Executive function and instrumental activities of daily living in mild cognitive impairment and Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):300-308. doi: 10.1016/j.jalz.2010.04.005