# Supplementary Material – Additional File 4

## Table S.1. Overview of measures

| Measure | N of publications meeting criteria† | Original intended…  a) construct  b) target population  c) context | Version | Measurement characteristics (refers only to the apathy component of the scale) | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Mode of administration & other administration information | Recall Period | Number of items | Scoring and Response options\* |
| AD-RD [40,41] | 1 [40] | 1. Mood 2. Moderate to severe AD 3. Research or clinical | n/a | Interviewer-judgement, informed by observation and patient and carer interview | 7 days | 5 | Items rated for frequency on Likert scale (1 to 5, all options described) |
| AES  [33] | 9 [33,53–57,71,78,89] | 1. Apathy 2. People with various clinical disorders or apathy, (with MMSE over 10 for patient reported version) 3. Clinical | AES-C | Clinician-rated based on semi-structured interview with patient and observations. Bachelor level raters can conduct with 4-6 hours experience.  10 to 20 minutes to administer | 4 weeks | 18 | Items rated on Likert scale (1 to 4; all options described), and quantifiable items rated 1 to 4 based 0, 1-2, 2-3, 3 or more quantifiable instances. Requires verbal or nonverbal evidence of intensity.  Total score is sum of item scores. Range 18 to 72. |
| AES-I | Informant-report via paper and pencil  10 to 20 minutes to administer. | 4 weeks | 18 | Likert scale (1 to 4; all options described).  Total score is sum of item scores.  Range 18 to 72. |
| AES-I (16 item versions) | Informant-report via paper and pencil | 4 weeks | 16 | Likert scale (1 to 4; all options described).  Total score is sum of item scores. Range 18 to 64 |
| AES-S | Self-report via interview (recommended) or paper and pencil  10 to 20 minutes to administer | 4 weeks | 18 | Likert scale (1 to 4; all options described).  Total score is sum of item scores. Range 18 to 72. |
| AES-12PD | Self-report | 4 weeks | 12 | Likert scale (1 to 4; all options described).  Total score is sum of item scores. Range 18 to 48. |
| AI [34] | 3 [34,72,80] | 1. Apathy 2. Older adults with brain disorders 3. Clinical | AI-C | Clinician opinion based on observations, and participant and informant answers to the AI when available.  At least 20 minutes of observation | Since beginning of the disease, last clinical assessment, or other defined time period e.g. last four weeks. | 3 | Likert scale (0 to 4; 3 options described)  Total score is the sum of item scores. Range 0 to 12 |
| AI-I | Informant-report via interview | Since beginning of the disease or an otherwise specified time point | 3 | Screening questions: (Yes=0 or No) with follow-up questions rated on Likert scale (Frequency: 1 to 4; Severity: 1 to 3; all options described )  Item score is Frequency x Severity. Range 0 to 12.  Total score is the sum of items scores. Range 0 to 36. |
| AI-S | Self-report via interview | Since beginning of the disease or an otherwise specified time point | 3 | Screening questions: 0=“Yes”; “No” with follow up question rated on a visual scale (1 to 12; end-points described).  Total score is the sum of item scores. Range 0 to 36. |
| AMI [35] | 1 [67] | 1. Apathy 2. Healthy adults 3. Research | n/a | Self-report via paper & pencil | 2 weeks | 18 | Likert Scale (0 to 4; all options described).  Total score is sum of item scores. Range 0 to 72. |
| AS [36] | 8 [36,58–61,77,90,91] | 1. Apathy 2. Parkinson’s Disease 3. Clinical | AS-HC | Self-report via paper and pencil | 4 weeks | 11 | Likert scale: (0 to 3; all options described). Total score is sum of item scores. Range 0 to 33 |
| AS-I | Informant report via interview  ~ 10 minutes to administer | 4 weeks | 14 | Likert Scale (0 to 3; all options described). Total score is sum of item scores. Range 0 to 42. |
| AS-S | Self-report via interview | 4 weeks | 14 | Likert Scale (0 to 3; all options described). Total score is sum of item scores. Range 0 to 42. |
| AS-S (13 item version) | Self-report via interview | 4 weeks | 13 | Likert scale: (0 to 3; all options described). Total score is sum of item scores. Range 0 to 39 |
| BMDS [42] | 1 [42] | 1. Neuropsychiatric symptoms (behaviour & mood disturbances) 2. Dementia 3. Research | n/a | Informant report via interview | - | 11 | Likert scale (0 to 4; all options described)  Total score is sum of item scores. Range 0 to 44. |
| BSSD [43] | 1 [43] | 1. Neuropsychiatric symptoms (behavioural syndromes in AD) 2. AD 3. Clinical | n/a | Clinician-judgement based on information from interview with informant and informed by clinician observations | 1 week | 7 | Likert scale (0 to 6; all options described).  Total score is not specified but presumable sum of item scores. |
| DAIR [37] | 1 [37] | 1. Apathy 2. Dementia (mild-moderate) 3. Research and clinical | n/a | Interviewer-judgement based on informant reports. In person or over the phone.  ~ 30 minutes administration time | 1 month | 16 | Main items rated on Likert scale by informant: (0 to 3; all options described) with follow-up questions to determine if this was a change in apathy rated by the interviewer (no change; increase; decrease)  Total score is sum of all items reflecting a change (more apathetic), divided by the number of items completed. |
| DAS [38] | 5 [62,63,74–76] | 1. Apathy 2. Neurodegenerative diseases specifically with motor disability 3. Research and clinical | DAS-I | Informant reported via online or paper and pencil  ~ 5 minutes to administer | 1 month | 24  (8 per subscale) | Likert scale (0 to 3; all options described).  ‘Executive’, ‘Initiation’ and ‘Emotional’ subscales are scored by summing all items in sub-scale. Range 0 to 24.  Total score is the sum of the subscale scores. Range 0 to 72. |
| DAS-S | Self-reported via online or paper and pencil  ~ 5 minutes to administer | 1 month | 24  (8 per subscale) | Likert scale (0 to 3; all options described).  ‘Executive’, ‘Initiation’ and ‘Emotional’ subscales are scored by summing all items in sub-scale. Range 0 to 24.  Total score is the sum of the subscale scores. Range 0 to 72. |
| b-DAS | Informant reported via online or paper and pencil  >5 minutes to administer | 1 month | 9  (3 per subscale) | Likert scale (0 to 3; all options described).  ‘Executive’, ‘Initiation’ and ‘Emotional’ subscales are scored by summing all items in sub-scale. Range 0 to 9.  Total score is the sum of the subscale scores. Range 0 to 27.  (an awareness deficit rating is also present but not included in the total score) |
| DEX [44]^ | 1 [81] | - | - | - | - | - | - |
| FrSBe [45]^ | 2  [64,68] | - | FrSBe-14a | - | - | 14 | - |
| FrSBe-11a | - | - | 11 | - |
| FrSBe-6a | - | - | 6 | - |
| GDS  [46,47] | 2  [69,92] | 1. Depression 2. Older adults 3. Clinical screening | GDS-3a | Self-reported via paper and pencil (interviewer administered if required) | 1 week | 3 | Responses (Yes/No) that indicate depression are scored 1.  Total score is sum of items. Range 0 to 3 |
| GDS-6a | Self-reported via paper and pencil (interviewer administered if required) | 1 week | 6 | Responses (Yes/No) that indicate depression are scored 1.  Total score is sum of items. Range 0 to 3 |
| GIP [48]^ | 1  [82] | - | GIP-subscale | - | - | - | - |
| GIP-domain | - | - | - | - |
| GIP-9a (subscale of the GIP-28) | Observation by health professional | 2 to 3 weeks | 9 | Likert scale (options not described) |
| IMD [49] | 1  [49] | 1. ‘Mental decline’ or ‘impairment’ 2. Older adults, particularly with dementia 3. Research. (Possibly also for clinical evaluation of progression but should not be used for diagnosis) | n/a | Informant reported | Not reported | 3 | Items are rated using categories that are associated with weighted scores depending on the item.  0=“Absent”; 2/3=“Mild-moderate / discontinuous symptoms”; 4/5/6=“Severe / continuous symptoms”  Total score is sum of item scores. Range 0 to 15 |
| KBCI [50,104] | 1  [93] | 1. Behaviour change 2. Traumatic Brain Injury 3. Clinical and research | KBCI-8a | Informant reported via paper and pencil | Not reported | 8 | Likert scale (all options described) Total score is the sum item scores but the scores attributed to the Likert scale and therefore the range is unspecified. |
| KBCI-10a | Informant reported via paper and pencil | Not reported | 10 | Likert scale (all options described) Total score is the sum item scores but the scores attributed to the Likert scale and therefore the range is unspecified. |
| LARS [39] | 3  [65,70,83] | 1. Apathy 2. Parkinson’s Disease 3. Clinical and research? | LARS-C | Interviewer-judgement informed by patient self-report and interviewer observations during the interview with the patient | 4 weeks | 33 | Four items are based on 3 or 5 point Likert scales (all options described)  For the remaining items, patient responses are categorised by the interviewer as 1 or -1 (all options described). Items are scored 0 if they are rated ‘N/A’ or the interviewer was not able to categorise the reply.  Total score is the sum item scores. Range -36 to 36. |
| LARS-I | Interviewer-judgement informed by informant-responses during the interview with the informant | 4 weeks | 33 | Five items are based on 3 or 5 point Likert scales (all options described)  For the remaining items, informant responses are categorised by the interviewer as 1 or -1 (all options described). Items are scored 0 if they are rated ‘N/A’ or the interviewer was not able to categorise the reply.  Total score is the sum item scores. Range -36 to 36. |
| NPI [51] | 12  [51,66,73,79,84–88,94–96] | 1. Neuropsychiatric symptoms 2. Dementia 3. Research and clinical | NPI (original) | Informant rated via interview | 1 month (and represents a change from behaviour before the illness) | 1 (but rated for frequency and severity) | Screening question (Yes=0; No), with follow-up questions using Likert scales, regarding severity (1 to 3; all options described) and frequency (1 to 4; all options described).  Total score is Frequency x Severity  (a distress rating is also present but not included in total score) |
| NPI-A | Informant rated via interview | 1 month (and represents a change from behaviour before the illness) | - | Each item is rated for frequency on the same Likert scale as the original NPI.  Total score is the sum of frequency scores.  (Severity is also rated for the overall domain as per the original NPI procedure, but not included in the total score) |
| NPI-C | Clinician-judgement, informed by information from the NPI with an informant and patient as well as other relevant information about the patient. Clinicians must have a minimum of two years’ experience of NPSs in people with dementia | 4 weeks | 11 | Each item is scored individually by informants, employing the Likert method as the original NPI, regarding frequency, severity and distress. Total score is the summation of frequency and severity item scores.  A clinical rating method is also required: Each item is also rated by a clinician based on their clinical impressions, informed by the interview with the patient and informant, clinical notes and other carers, rated on Likert scale (0 to 3). Total score is the sum of these clinician rated item scores.  Two separate total scores are obtained: one from the informant, one from the clinician. |
| UPDRS [52]^ | 4  [97–100] | - | UPRDS | - | - | 1 | Likert scale (0 to 4; all options described). No total score calculation required as only 1 item present. |
| MDS-UPDRS | Rater-judgement informed by interview with patient and / or informant | 1 week | 1 | Likert scale (0 to 4; all options described). No total score calculation required as only 1 item present. |

† Number does not include development article where development article did not meet the inclusion criteria, even if it was later assessed for purposes of content validity

\* Reverse coding is not included here

^ Unable to obtain development article for rating

- Unable to obtain information

Abbreviations: AD-RD, Alzheimer's Disease and Related Dementias Mood Scale; AES Apathy Evaluation Scale; AES-12PD, Apathy Evaluation Scale for Parkinson Disease; AES-C, Apathy Evaluation Scale Clinician; AES-I, Apathy Evaluation Scale Informant; AES-S, Apathy Evaluation Scale Self; AI, Apathy Inventory; AI-C, Apathy Inventory Clinician; AI-I, Apathy Inventory Informant; AI-S, Apathy Inventory Self; AMI, Apathy Motivation Index; AS, Apathy Scale; AS-HC, Apathy Scale Home Care; AS-I, Apathy Scale Informant; AS-S, Apathy Scale Self; b-DAS, brief-Dimensional Apathy Scale; BMDS, Behavioural and Mood Disturbance Scale; BSSD, Behavioral Syndromes Scale for Dementia; DAIR, Dementia Apathy Interview Rating; DAS, Dimensional Apathy Scale; DAS-I, Dimensional Apathy Scale Informant; DAS-S, Dimensional Apathy Scale Self; DEX, Dysexecutive Questionnaire; FrSBe, Frontal Systems Behavior Scale; FrSBe-6a, Frontal Systems Behavior Scale 6-item apathy subscale; FrSBe-11a, Frontal Systems Behavior Scale 11-item apathy subscale; FrSBe-14a, Frontal Systems Behavior Scale 14-item apathy subscale; GDS, Geriatric Depression Scale apathy; GIP, Behavioral Rating Scale for Psychogeriatric Inpatients; IMD, Index of Mental Decline; KBCI, Key Behaviors Change Inventory; KBCI-8a, Key Behaviors Change Inventory 8 item apathy subscale; KBCI-10a, Key Behaviors Change Inventory 10 item apathy subscale; LARS, Lille Apathy Rating Scale; MDS-UPDRS, Movement Disorder Society-Sponsored Revision of the Unified Parkinson’s Disease Rating Scale; NPI, Neuropsychiatric Inventory; NPI-A, Neuropsychiatric Inventory Alternative; NPI-C, Neuropsychiatric Inventory Clinician; UPDRS, Unified Parkinson’s Disease Rating Scale

## Table S.2. Overview of studies

| Reference | Measure | Language of measure | Measurement properties investigated | Residential status | N | Population  (N of each subgroup, or % where N not possible to calculate) | Mean age  (SD, range) | Gender  (% Male) | Cognitive status  Mean MMSE (SD, range) unless otherwise stated | Apathy score  Mean (SD, range) |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| [40] | AD-RD | English^ | Reliability (test-retest). | "Approximately half lived in low-income housing." No confirmation from correspondence. | N=39 | nr | 79.33 (9.22 ; 55 to 96) | 49% | 17.21 (5.98, 3 to 24) | AD-RD apathy: 10.57 (3.88) |
| Development (pilot study) | "Conducted in a dementia-specific day center and two skilled nursing facilities."  No confirmation from correspondence | N=45 | Cognitive Impairment (type not specified) | 79.00 (8.37; 61 to 94) | 45% | 7.88 (6.47; 0 to 23) | nr |
| [40,41] | AD-RD | English^ | Development (item elicitation via interviews) | Nursing home and day care. No confirmation from correspondence regarding proportion. | N=39 | Carers of people with moderate to severe AD:  Formal carers (N=19).  Informal carers (N=20).  (Number of people with AD that were being interviewed about =20) | Nursing home: 85 (nr, nr)  Day care: 81 (nr,nr) | 25% | nr | nr |
| [53] | AES-C | Chinese | Structural validity; Internal consistency; Reliability (interrater & test-retest). Hypothesis testing (convergent, divergent & known groups). | Outpatients - confirmed all community dwelling via correspondence with author | N=92 | Major depressive disorder:  Current Depression (CD; N=31)  Remitted Depression (RD; N=30)  Healthy Controls (Ctrl; N=31) | CD=66.13 (8.24)  RD=67.83 (6.20);  Ctrl=68.90 (6.20); | CD=45.16%  RD=33.33%  Ctrl=48.39% | nr | 2 means for each group reflect 2 different clinicians’ ratings:  CD=42.32 (10.45);  40.32 (11.92)  RD=32.17 (8.27) ;  30.33 (7.46)  Ctrl=27.87 (7.55);  28.55 (9.24) |
| [54] | AES-C; AES-I; AES-S | nr | Structural validity;  Hypothesis Testing (convergent & divergent). | Community-dwelling (95.8%) and nursing home residents (4.2%). | N=121 | Dementia:  AD (55.2%);  MD (AD-DLB, 14.3%; AD-VaD, 5.7%);  DLB (9.5%);  VaD (5.7%),  FtD (4.8%);  ‘other dementia’ (4.8%). | 73.7 (9.4) | 47.1% | nr | nr |
| [55] | AES-C; AES-I; AES-S | English^ | Structural validity; Internal consistency; Hypotheses testing (divergent & known groups). | Outpatient and community sample – confirmed all community dwelling via correspondence with author | N=75 | MCI (N=57);  Cognitively normal (Ctrl N=18) | MCI: 74.5 (8.6, 53 to 86) Ctrl: 75.4 (6.0, 63 to 84)  Total: 74.7 (8.0, 53 to 86) | MCI: 70.2% Ctrl: 22.2% Total: 58.7% | MCI: 27.3 (1.9, 23 to 30) Ctrl: 29.4 (0.8, 28 to 30)  Total: 27.8 (1.9, 23 to 30) | AES-C:  MCI: 60.9±7.7 (39–72) Ctrl: 68.4±4.3 (55–72) Total: 62.7±7.7 (39–72) |
| AES-I:  MCI: 61.1 (8.0, 42 to 72) Ctrl: 68.3 (4.5, 58 to 72)  Total: 62.8 (7.9, 42 to 72) |
| AES-S:  MCI: 63.3 (8.0, 40 to 72) Ctrl: 67.2 (4.2, 56 to 72)  Total: 64.3 (7.4, 40 to 72) |
| [33] | AES-C; AES-I; AES-S | English^ | Development (item elicitation and pilot); Structural validity; Internal consistency; Reliability (interrater &test-retest); Hypothesis testing (divergent & known groups). | Community-dwelling | N=123  (N=40 for pilot)  (n/a for item elicitation) | Mixed sample:  Healthy controls (Ctrl, N=31);  Probable AD (N=21);  Major Depression (Dep; N=30);  Left Hemisphere Stroke (LHS, N=19);  Right Hemisphere Stroke (RHS =22). | Ctrl: 68.3 (5.7,nr)  AD: 70.8 (7.6,nr)  Dep: 71.6 (5.7,nr)  LHS: 66.2 (6.6,nr)  RHS: 70.1 (5.0,nr)  Total: 69.53 (6.03)\* 55 to 85) | Ctrl: 45.16%  AD: 47.62%  Dep: 10.00%  LHS: 57.89%  RHS: 54.55%  Total: 40.65% | Ctrl: 29.1 (1.1, nr)  AD: 19.1 (6.5, nr)  Dep: 28.0 (1.7, nr)  LHS: 25.0 (4.6, nr)  RHS: 26.9 (2.3, nr) | AES-C:  reported separately for the 2 clinician ratings: Ctrl: 26 (6.2, nr); 25.8 (5.8, nr) AD: 44.4 (11.1, nr); 45.2 (11.7, nr);  Dep: 40.5 (9.7, nr); 36.6 (8.3, nr)  LHS: 31.9 (9.6, nr); 32.0 (11.7, nr) RHS: 34.7 (7.3, nr); 35.4 (9.6, nr) |
| AES-I:  Ctrl: 26.3 (7.5, nr) AD: 49.1 (9.9, nr) Dep: 41.7 (15.0, nr)  LHS: 28.1 (6.9, nr) RHS: 35.4 (10.9, nr) |
| AES-S:  Ctrl: 28.1 (6.4, nr) AD: 35.5 (8.1, nr) Dep: 38.7 (9.8, nr)  LHS: 32.2 (8.6, nr) RHS: 31.6 (6.7, nr) |
| [71] | AES-I;  AES-I-16 | German | Internal consistency; Hypothesis Testing (divergent). | Community-dwelling. | N=100  (AES-I N=80.) | Dementia | 83.19 (8.32, 59 to 100, N=99) | 29% | 16.35 (7.60, 0 to 29, N=65) | AES-I: 31.74 (10.43, 8 to 48) |
| [56] | AES-I; AES-S | Swedish | Structural validity; Internal consistency; Measurement error; Hypothesis Testing (divergent). | Outpatients and community sample – No confirmation from correspondence whether the outpatients were community-dwelling. | N=511  Complete AES-I N=367.  Complete AES-S N=496. | Neurodegenerative disease and cognitive impairment:  MCS (N=222. AES-I N=192. AES-S N=209) with subgroups of subjective cognitive decline (SCD, N=97) and MCI (N=125). Parkinson's Symptoms (PS, N=88. AES-I N=76. AES-S N=88), with subgroups of PD (PD, N=71); Parkinson’s Disease Dementia or Dementia with Lewy Bodies (PDD-DLB, N=17).  Ctrl (N=201. AES-I N=135; AES-S N=199) | MCS: 70 (6) MCI: 71 (6) PD: 67 (9) PDD-DLB: 74 (6) Ctrl: 75 (5)  Total: 72 (7) | MCS: 44.3%\* MCI: 52%\* PD: 56.3%\* PDD-DLB: 76.5%\*  Ctrl: 37.8%\*  Total: 46.4%\* | *median (Q1 to Q3)*  MCS: 29 (27to29) MCI: 27 (26to28) PD: 29 (27to30) PDD-DLB: 23 (20to24)  Ctrl: 29 (28to30)  Total: 29 (27to29) | AES-I  MCS: 36.2 (10.6, nr)  PS: 52.3 (11.4, nr))  Ctrl: 28.7 (8.2, nr)  Total: 36.6 (12.9, nr) |
| AES-S  MCS: 32.6 (8.8 , nr)  PS: 53.3 (10.6, nr)  Ctrl: 28.0 (5.7, nr)  Total: 34.2 (11.9, nr) |
| [89] | AES I; AES-S | Italian | Hypothesis Testing (divergent). | Outpatients – No confirmation from correspondence whether community-dwelling. | N=48 | Parkinson’s Disease (PD) | 72.21 (9.01, nr) | 64.58%\* | 22.83 (4.71,nr) | AES-I: 45.14 (13.09, nr) |
| AES-S: 49.85 (10.37, nr) |
| [57] | AES-S | German^ | Structural validity; Internal consistency; Hypothesis testing (convergent & divergent). | Author confirmed all community via correspondence. | N=665 | Parkinson’s Disease  Sub-sample of PD excluding comorbidities of dementia or depression (PDexclDd; N=339) | PD: 67.3 (7.90,nr) PDexclDd: 66.52 (7.96,nr) | PD: 67.9%  PDexclDd: 66.52% | PD: 27.94 (2.23) PDexclDd: 28.47 (1.58) | PD: 30.63 (9.49) PDexclDd: 27.96 (7.59) |
| [78] | AES-12PD | German | Internal consistency; Hypothesis testing (convergent & divergent) | Data taken from a study that has been confirmed community via correspondence. | N=339 | Parkinson’s Disease.  (Sample split for analyses: Sample 1: N=170; Sample 2: N=169)  Subsample of PDDd: N=42 | Sample 1: 68 (nr, nr)  Sample 2: 68 (nr, nr) | Sample 1: 70.00%  Sample 2: 70.41% | *median (Q1 to Q3)*  Samples 1&2: 29 (nr, nr) | *median (Q1 to Q3)*  AES:  Samples 1&2: 27.0 (nr)  AES-12PD:  Sample 1: 17.0 (nr)  Sample 2: 18.0 (nr) |
| [34] | AI | French^ | Development (item elicitation) | n/a no participants. | n/a | n/a | n/a | n/a | n/a | n/a |
| Internal consistency; Reliability (test-retest & interrater), hypothesis testing (convergent, divergent & known groups) | Author advised outpatients via correspondence nut unable to confirm whether community dwelling. | N=115.  (Test-retest N=14). | People with neurodegenerative disease or cognitive Impairment:  AD (N=60);  PD without dementia (N=12),  MCI (N=24)  Ctrl (N=19).  Test-retest: AD only. | AD: 74.90 (7.11, nr)  PD: 64.1 (11.9, nr)  MCI: 71.67 (5.92, nr)  Ctrl: 70.68 (8.21, nr)  Total: 72.40 (7.52)\* | AD: 45.00  PD: 58.33  MCI: 29.17  Ctrl: 42.11 | AD: 22.55 (3.98, nr)  PD: 27.2 (3.5, nr)  MCI: 28.21 (1.06, nr)  Ctrl: 29 (nr, nr) | AI-I  AD: 9.20 (10.4, nr)  PD: 8.00 (6.0, nr)  MCI: 4.21 (8.6, nr)  Ctrl: 1.05 (2.0, nr) |
| AI-S  AD: 3.74 (5.9, nr)  PDexlD: 9.10 (8.3, nr)  MCI: 2.47 (3.8, nr)  Ctrl: 1.51 (2.9, nr) |
| [80] | AI-C | Portuguese | Internal consistency; Reliability (interrater); Hypothesis testing (convergent). | nr, but confirmed all community via correspondence | N=175. | Mixed sample:  AD (N=55)  MCI (N=35)  Dep (N=32)  PD (N=30)  Ctrl (N=23) | AD: 78.4 (nr, 61 to 95)  MCI: 69.1 (nr, 60 to 86)  Dep: 69.7 (nr, 55 to 88)  PD: 66.5 (nr, 42 to 84);  Ctrl: 67.3 (nr, 52 to 88)  Total: 71.45\* | Total: 34.3% | AD: 16.8 (nr, 0 to 27)  PD: 26.9 (nr, 18 to 20)  Dep: 24.3 (nr, 16 to 30)  MCI: 25.4 (nr, 22 to 27)  Ctrl: 29.1 (nr, 28 to 30)  Total: 23.28\* | AI scores nr. Apathy 'diagnosis' according to Robert et al criteria:  AD: 63.6%  PD: 20%  Dep: 68.8%  MCI: 0%  Ctrl: 0% |
| [72] | AI-C | French | Internal consistency; Hypothesis testing (convergent). | Outpatients – No confirmation from correspondence whether community-dwelling. | N=40 | Cognitive Impairment  AD (N=17); MCI (N=12); MD (N=8); VaD (N=2); DLB (N=1) | 77.5 (8.01, nr) | 45%\* | 20 (6.73, nr) | nr |
| [67] | AMI | English | Internal consistency; Hypothesis testing (convergent) | Outpatients – No confirmation from correspondence whether community-dwelling. | N=149 | PD (N=102)  Ctrl (N=147) | PD: 67.7 (8.1,nr)  Ctrl: 66.1 (8.5, nr)  All at least 18 to 80 | PD: 77.5%  Ctrl: 70.75% | ACE-III:  PD: 89.4 (9.0, nr)  All at least over 50  Ctrl: nr | PD: 35.29% apathetic in at least one AMI subscale |
| [90] | AS-I | Portuguese | Content validity | Outpatients – No confirmation from correspondence whether community-dwelling. | N=11 | Dementia:  AD (N=8);  FtD (N=3); | AD: 78.3 (4.7)  FtD: 55 (8.7)  Total: 71.95 (5.59)\* | AD: 50.00%\*  FtD: 33.33%\*  Total:45.45%\* | nr for this sample.  Total: 20.64 (3.85)\* | 22.8 (8.4, 12 to 39) |
| Hypothesis testing (convergent & divergent) | Population random sample – No confirmation from correspondence whether community-dwelling. | N=20 | Probable or Possible AD | 84.1 (5.8) | 30% | 17.4 (SD=4.7) | 23.6 (10.6; 9 to 40) |
| [59] | AS-S (14/13 item) | English^ | Structural validity; Internal consistency. | nr, but confirmed all community via correspondence | N=226 | Parkinson’s Disease, without dementia. | 65.02 (8.84, nr) | 66.70% | (N=7) 29.14 (0.69, nr) | 10.99 (6.26, nr) |
| [36] | AS-S | English^ | Development | n/a, no participants | n/a | n/a | n/a | n/a | n/a | n/a |
| Internal consistency; Reliability (interrater & test-retest); Hypothesis testing (known groups) | nr. Author unable to access the information. | N=50  (Reliability studies: N=11) | Parkinson's disease, grouped into sub-samples based on apathy and depression scores:  PD, no apathy, no depression (PD; N=16)  PD, with apathy, no depression, (PDa; N=6)  PD, no apathy, with depression, (PDd; N=13)  PD, with depression and apathy (PDa&d; N=15) | PD: 67 (9, nr)  PDa: 69 (7, nr)  PDd: 62 (12, nr)  PDa&d: 69 (8, nr) Total: 66.54 (9.26)\* | PD: 50%  PDa: 66%  PDd: 57%  Pa&d: 73%  Total: 62%\* | PD: 28.7 (1.1, nr)  PDa: 28.3 (1.2, nr)  PDd: 26.3 (4.6, nr)  PDa&d: 25.4 (4.5, nr)  Total: 27.04 (3.06)\* | PD: 7.3 (2.8, nr)  PDa: 17.1 (4.0, nr)  PDd: 10.0 (2.0, nr)  PDa&d: 19.5 (3.3, nr) Total =12.84 (2.87)\* |
| [58] | AS-S  AS-HC | Japanese | Structural validity; Internal consistency; Hypothesis testing (divergent). | “Home-care” recipients. Assumed community-dwelling | N=122 | Parkinson's Disease | 70.9 (7.8, nr) | 49.2% | nr | AS-S: 26.6 (8.12, nr) AS-S-11: 21.3 (6.88, nr) |
| [60] | AS-S | Norwegian | Structural validity; Internal consistency; Hypothesis testing (divergent). | nr. No confirmation from correspondence whether community-dwelling. | N=194 | Parkinson’s Disease | 67.9 (9.0, nr) | 59.3% | 27.8 (2.3, nr) | 15.5 (4.6, 4 to 29) (median =15.0). |
| [77] | AS - S | Spanish^ | Internal consistency; Reliability (test-retest); Measurement error; Hypothesis testing (divergent & known-groups) | Outpatients – No confirmation from correspondence whether community-dwelling. | N=211  (test-retest: N=71) | Parkinson’s Disease | 67.5 (10.2, nr) | 65.5%\* | Short Portable Mental Status Questionnaire of Pfeiffer: 1.3 (1.6, nr). | 12.7 (7.1, nr) |
| [61] | AS-S | English^ | Structural validity; Internal consistency. | Outpatients. Confirmed community-dwelling via correspondence | N=233 | Parkinson’s Disease and healthy controls  PD (N=157)  Ctrl (N=76) | PD: 67.64 (8.27, nr)  Ctrl: 66.95 (8.73, nr) | PD: 68.15%\* Ctrl: 44.74%\* | Mattis dementia rating scale:  PD: 138.48 (3.88,nr)  Ctrl: 140.46 (3.24,nr) | PD: 11.59 (5.36,nr)  HC: 9.21 (4.67,nr) |
| [91] | AS-S | Spanish | Internal consistency; Hypothesis testing (convergent; divergent) | nr. Unknown to corresponding author as data not collected. | N=60 | Advanced Parkinson’s Disease | 68.02 (7.43; 50 to 81) | 60.70% | nr | 11.55 (6.49, 1 to 24) |
| [42] | BMDS | English^ | Development (item elicitation) | n/a, no participants | n/a | n/a | n/a | n/a | n/a | n/a |
| Reliability (test-retest). | nr, but scale designed to assess people living in the community | N=38 (test-retest reliability N=18) | Dementia | 76 (nr, 59 to 87) | 23.68% | nr | 24.95 (9.30, nr) |
| [43] | BSSD | English | Development (item elicitation and pilot) | Item elicitation: n/a no participants  Pilot: nr | nr | nr | nr | nr | nr | nr |
| Internal consistency; Reliability (interrater & test-retest); Hypothesis Testing (divergent & known groups) | Outpatients – No confirmation from correspondence whether community-dwelling. | N=106  (hypothesis testing: N=83 to 97; reliability: N=20 to 21) | Alzheimer’s Disease | 72.1 (9.8, 45 to 93) | 35% male | Modified MMSE: 26.2 (13.8, 0 to 52) | Global apathy / indifference =31.1% absent; 50.0% minimal to mild; 18.8% moderate to severe. raw scores nr. |
| [37] | DAIR | English^ | Development (item elicitation and pilot); | nr | nr | Mixed sample: People with AD, their carers and clinical researchers. | nr | nr | nr | nr |
| Structural validity; Internal consistency; Hypothesis testing (convergent & divergent) | nr  Designed to assess people living in environments whose daily activities are not structured, suggesting community-dwelling. No confirmation from correspondence whether community-dwelling. | N=100 | Alzheimer’s Disease | 75.00 (8.48; 52 to 92) | 50% | 18.55 (7.20; 3 to 29)  (Unobtainable for 16%) | 1.19 (0.69, 0 to 3) |
| [38] | DAS | English (assumed) | Development study (item elicitation) | n/a no participants involved in item elicitation | n/a | n/a | n/a | n/a | n/a | n/a |
| [75] | DAS | English^ | Internal consistency; Hypothesis testing (convergent; divergent) | Outpatients - all confirmed community via correspondence. | DAS-S  N=68 | Parkinson’s Disease without dementia and healthy controls:  PD (N=34)  Ctrl (N=34) | PD: 68.2 (9.2, nr)  Ctrl: 66.1 (9.2, nr) | 44.12% | nr | PD: 25.8 (8.7, nr) Ctrl: 21.2 (7.0, nr) |
| DAS-I  N=60 | (sub-sample of those above)  PD (N=30)  Ctrl (N=30) | nr for this sub-sample | nr for this sub-sample | nr | PD: 25.1 (12.8, nr) Ctrl: 19.7 (9.5, nr) |
| [74] | DAS | English^ | Internal Consistency; Hypothesis testing (convergent & divergent) | Community-dwelling | N=157\* | DAS-I  Alzheimer’s Disease and controls  AD (N=102)  Ctrl (N=55) | AD: 78.2 (8.5, nr) 82.4% aged 65 and over. Ctrl: 75.0 (6.1, nr) | AD: 51.0%\*  Ctrl: 50.9%\* | AD (N=80): 22.0 (5.3, nr)  Ctrl: nr | nr, but AES:  AD: 51.7 (11.5, nr)  Ctrl: 28.8 (5.2, nr) |
| DAS-S  AD (N=55, sub-sample of those above)  Ctrl (same as above, n=55) | AD: 77.5 (7.9, nr)  Ctrl: 75.0 (6.1, nr) | AD: 50.9%\*  Ctrl: 50.9%\* | nr | nr, but AES:  AD: 38.9 (9.0, nr) |
| [62] | DAS-S | Italian | Structural validity, Internal consistency, Hypothesis testing (convergent, divergent & known groups) | Outpatients - all confirmed community via correspondence. | N=207 | Parkinson’s Disease and controls  PD (N=107)  Ctrl (N=100) | PD: 66.02 (9.01,nr)  Ctrl: 64.52 (8.79,nr) | PD: 60.75%\* | PD: 27.63 (2.09,nr) | PD: 25.25 (12.76,nr) (Median (skewness)=23 (1.254))  Ctrl: 21.29 (8.35,nr) |
| [63] | bDAS | English | Structural validity | AD: Community-dwelling  ALS: nr | N=204 | Neurodegenerative Disease  AD (N=102)  ALS (N=102) | AD: 78.2 (8.5, nr)  ALS: 63.8 (11.0, nr)  Total: 71.0 (12.1, nr) | AD: 51%\* ALS: 70%\*  Total: 60%\* | AD: (N=80): 22.0 (5.3, nr)  ALS: nr  Total: nr | nr for bDAS  AES:  AD: 51.7 (11.5, nr)  ALS: 33.2 (10.8, nr)  Total: 42.4 (14.4, nr) |
| [76] | bDAS | English^ | Internal consistency; Reliability (test-retest). | All confirmed community via correspondence. | N=53 (reliability N=43) | ALS | 68.0 (7.5, nr) | 83.01%\* | ECAS cognitive score: 107.0 (14.1,nr) | nr for total score  DAS-I subscales:  Executive: 6.1 (4.8, nr)  Emotional: 8.9 (4.2, nr)  Initiation: 12.1 (5.5, nr)  b-DAS  Executive: 2.0 (2.0, nr)  Emotional: 2.9 (1.9, nr)  Initiation: 4.3 (2.6, nr) |
| [81] | DEX | Japanese | Reliability (test-retest); Hypothesis testing (convergent & divergent) | Outpatients. | N=122 (reliability N=44) | Alzheimer’s Disease | 72.0 (7.7, nr) | 37.70%\* | 20.8 (2.0, nr) | nr |
| [68] | FrSBe-I | English^ | Content validity (cognitive interview) | Outpatients - all confirmed community via correspondence. | N=10 | People attending neuropsychological evaluation. 90% had memory complaints. Diagnoses nr. | nr | nr | nr | nr |
| Structural validity; Internal consistency; hypothesis testing (groups & divergent); | Outpatients - all confirmed community via correspondence. | N=494 | Mixed sample:  Dementia: AD (19.3%\*), VaD (4.9%); Dementia not otherwise specified (4.1%); MD (4.5%); FTD (4.1%); DLB (1.8%).  PD (16.6%).  MCI (12.5%).  Cognitive disorder not otherwise specified (CDNOS, 8.8%).  Frontal stroke (7.2%).  Head injury (2.1%).  Other neurological disorder (<1%). | 69.92 (13.96, 19 to 95) | 47.04%\* | nr | Original FrSBe-apathy:  PD=33.29 (12.71); AD =37.24 (10.18); Frontal impairment =38.18 (10.35)  Revised FrSBe-apathy:  PD=27.24 (10.13); AD =29.71 (7.83); Frontal impairment =30.21 (8.08) |
| [64] | FrSBe-I | English^ | Structural validity; Internal consistency | Outpatients - all confirmed community via correspondence. | N=304 | Older adults with memory complaints:  Dementia (N=166)  MCI (N=63)  No definitive diagnosis (NDD; N=28)  Ctrl (N=47) | 79.12 (8.05; 52 to 99) | 28.29%\* | nr | 86.12 (24.39) |
| [46,47] | GDS-30 | English^ | Development (Item elicitation and pilot study) | Item elicitation: n/a no participants  Pilot: Community dwellers (N=20) and inpatients (N=51). | N=71 | Healthy older adults (Ctrl: N=20)  Depressed older pts (Dep: N=51) | nr.  All over 55. | nr | nr | nr |
| [92] | GDS-3A | Dutch^ | Hypothesis testing (convergent validity) | Community-dwelling | Study 1  N =427 | Older adults with mild cognitive deficits | 81.3 (4.6, nr)  All at least 75 and over | 39.8%\* | *median (Q1 to Q3)*  26 (25 to 27) | GDS-3a score:0 =52.8%; 1=30.7%; 2=12.2%; 3=4.4%  AS: 11.3 (4.7) |
| Study 2  N=1118 | Older adults with depressive symptoms | 81.8 (4.9, nr)  All at least 75 and over | 38.9%\* | *median (Q1 to Q3)*  28 (27 to29) | GDS-3a: 0 =64.2%; 1 =25.6%; 2 =9.3%; 3 =0.89% AS: 7.5 (4.6, nr) |
| [69] | GDS-6A | English^ | Internal consistency, Hypothesis testing (divergent & known groups) | Community-dwelling | N=140 | Mixed sample:  Dementia: AD (29.3%); VaD (29.3%); MD (13.6%)  Cognitive disorder not specified or MCI (CNS-MCI, 17.1%)  Other (6.4%); None (2.1%)  (2.2% nr) | 78.2 (7.23, nr)  All at least 65 or over | 35.0%\* | 24.86 (3.35, nr) | GDS-6a: 1.66 (1.39, nr) |
| [82] | GIP-a-s  GIP-a-d | Dutch | Reliability (test-retest); Measurement error. | All confirmed community via correspondence. | N=109 Complete and analysed: N=56. | Mixed sample: Dementia: AD (82%); VaD (13%); Other dementia (3%);  Other (affective disorder or other cognitive disorder, 2%) | *median (Q1 to Q3, range)*  80 (75.5 to84, 53 to 96) | 42.2%\* | *median (Q1 to Q3, range)*  Cognitive Screening test: 13.3 (10.4 to16, 3.5 to 20) Amsterdam Dementia Screening test 3: 0 ( -2 to 1, -5 to 4)  Amsterdam Dementia Screening test 5: 1 (-1 to3, -5 to 8). | N=56: GIP-a-s: 2.2 (2.3, 0 to 9)  GIP-a-d: 2.8 (3.5, 0 to 15) |
| [49] | IMD | Italian^ | Development (item elicitation) | n/a no participants | n/a | n/a | n/a | n/a | n/a | n/a |
| Hypothesis testing (divergent) | Sample 1:  Some Community-dwelling and some institutionalised.  Author unable to confirm proportion. | N=236 | nr, but at least some healthy older adults.  Mild to moderate functional impairment (52.5%). Severe functional impairment (24.8%). | 74.2 (6.8, nr) | 40.6%\* | 19.4 (4.3, nr) | nr |
| Sample 2:  nr.  Author unable to confirm. | N=203 | Dementia | 74.1 (5.56; 63 to 83) | 33.99%\* | 19.7 (2.61, 15 to 23) | 5.4 (3.15) |
| [50,104] | KBCI | English^ | Development (item elicitation) | nr | nr | People with TBI, their carers, and TBI rehabilitation specialists. | nr | nr | nr | nr |
| Development (item refinement) | panel1: nr. panel 2 & 3: n/a. | N=14 | Panel 1: carers for people with TBI (N=4)  Panel 2: clinical psychologists (N=3)  Panel 3: clinical neuropsychologists (N=7) | nr | nr | nr | nr |
| [93] | KBCI-a | English^ | Hypothesis testing (divergent) | Outpatients. No reply from author. | N=97 | Mixed sample:  Ctrl (31%)  MCI (18%)  Probable AD (7%)  Other (depression, CDNOS, PD, DLB, and possible AD) | 72.34 (9.05, nr) | nr | 26.89 (2.63, nr) | nr |
| [39] | LARS | French; English | Development | n/a – no participants involved. | n/a | n/a | n/a | n/a | n/a | n/a |
| [83] | LARS - C | Spanish | Reliability (interrater & test-retest); Hypothesis Testing (convergent) | Community-dwelling (“non-institutionalised”) | N=151  (test-retest N=16, interrater N=21) | Dementia (Dem, N=101) and healthy controls  AD (N=43)  FtD (N=41)  Primary Progressive Aphasia (N=17)  Ctrl (N=50) | Dem: 74.3 (7.7, nr) Ctrl: 72.0 (9.7, nr) | Dem: 45.5%\* Ctrl: 38%\* | Dem: 21.59 (6.21, nr) Ctrl: 28.72 (1.42, nr) | Dem: -0.16 (18.50, nr)  Ctrl: -29.54 (5.44, nr) |
| [70] | LARS-I | French^ | Internal consistency; Reliability (interrater & test-retest); Hypothesis Testing (convergent) | Correspondence with author confirmed all community | N=60  (interrater N=34, test-retest N=29) | Parkinson’s Disease:  PD without dementia (PDexclD, N=43)  PD with dementia (PDD, N=17) | PDexclD: 64.74 (9.29, nr)  PDD: 69.53 (9.06, nr)  Total: 66.10 (9.23)\* | PDexclD: 67.44%\* PDD: 35.29%\* | nr | -16.18 (11.99, nr) |
| [65] | LARS - C | Spanish | Content validity; Structural validity; Internal consistency ; Reliability (interrater & test-retest); Hypothesis testing (convergent & divergent) | nr. No confirmation from correspondence whether community-dwelling. | N=200 (content validity and reliability N=30) | Parkinson’s Disease and healthy controls  PD (N=130)  Ctrl (N=70) | PD: 71.6 (8.1, nr) Ctrl: 69.4 (8.7, nr) | PD: 60.0%\*  Ctrl: 55.7%\* | MEC:  PD: 30.7 (3.8, nr)  Ctrl: 33.3 (1.7, nr) | PD: -14.5 (9.1, nr)  Ctrl: -25.0 (5.5, nr) |
| [94] | NPI | Korean | Hypothesis Testing (known groups). | Assessment setting suggests outpatients. No confirmation from correspondence whether community-dwelling. | N=141  (test-retest N=29) | Dementia (N=92) and healthy controls:  AD (N=43)  VaD (N=32)  FtD (N=11)  Other dementia (N=6)  Ctrl (N=49) | Dem: 67.5 (9.7, 38 to 85) Ctrl: 66.9 (8.4, 51 to 82) | Dem: 47.8%\*  Ctrl: 34.7%\* | Dem: 17.5 (6.8, 0 to 29) Ctrl: 26.3 (2.3,19 to 30) | NPI-apathy total nr.  Dem:  Prevalence: 77.2%.  Frequency: 2.52 (1.67; 0 to 4)  Severity: 1.75 (1.18; 0 to 3)  Ctrl:  Prevalence =6.1%.  Frequency =0.06 (0.24; 0 to 1)  Severity =0.06 (0.24; 0 to 1) |
| [51] | NPI | English^ | Development (item elicitation and Delphi study of comprehensiveness) | Item elicitation: n/a no participants  Delphi study: n/a professionals | N=10 | Geriatric psychiatrists, behavioural neurologists, and neuropsychologists | n/a | n/a | n/a | n/a |
| Reliability (interrater & test-retest) | Community-dwelling | N=80  (interrater N=45, test-retest N=20) | Dementia (Dem) and healthy controls:  AD (N=20)  VaD (N=9)  Other dementia (N=11)  Ctrl (N=40) | 75.7 (56 to 90) | Dem: 55.00%\*  Control: 50.00%\* | Dem: 19.2 (0 to 29)  Control: 28.4 (25 to 30) | NPI-apathy total nr.  Frequency: 2.83 (1.55; 0 to 4) Severity: 1.35 (0.83; 0 to 3) |
| [85] | NPI | Icelandic | Reliability (test-retest); Hypothesis testing (known groups). | Community-dwelling | N=38  (test-retest N=6) | Dementia: AD (N=19)  VaD (N=19) | 78.84 (6.66; 59 to 89) | 47% | 19.26 (5.95; 1 to 29) | nr for total sample. Reported separately for two different severity groups (N in each group nr).  Less severe dementia: 4.69 (3.72, nr)  More severe dementia: 7.45 (4.45, nr) |
| [73] | NPI | Farsi | Internal consistency; Reliability (interrater & test-retest); Hypothesis testing (convergent, divergent & known groups) | 51% living with family, suggesting at least majority community dwellers. No confirmation from correspondence | N=100.  (interrater N=50, test-retest reliability N=30, hypothesis testing N=50) | Dementia and healthy controls.  Dem (N=100)  Ctrl (N=49) | Dem: 74.5 (8.3, 60 to 90)  Ctrl: 74.3yrs (8.5) | Dem: 47%  Ctrl: 51% | nr for total sample.  Hypothesis testing (N=50):  Dem: 11.3 (7.5, nr)  Ctrl: 29.4 (1.0, nr) | NPI-apathy total nr.  Prevalence: 74% Frequency 2.5 (1.7, nr) Severity 1.6 (1.1, nr) |
| [79] | NPI | Spanish | Internal consistency; Reliability (interrater); Hypothesis testing (convergent) | Outpatients – No confirmation from correspondence whether community-dwelling. | Total N=63.  (interrater N=39) | Mixed sample:  Dem (N=44)  Dep (N=6)  Ctrl (N=13) | 72.76 (9.67; 35 to 85) | 49.21%\* | nr | NPI-apathy total nr.  Prevalence: 56% |
| [95] | NPI | Greek | Hypothesis testing (convergent) | Outpatients. Author correspondence confirmed all community. | N=29 | Dementia | 71.05 (5; 60 to 84) | 60% | 12.4 (6.0; 0 to 24) | 5.8 (4.4, nr) |
| [86] | NPI | Chinese | Reliability | Community dwelling | N=91 | Dementia and healthy controls.  Dementia (Dem, N=62\*): AD (N=41), VaD (N=16), Other (N=5)  Ctrl (N=29) | Dem: 76.4 (7.0; 54 to 88). Ctrl: 74.9 (4.7; 68 to 86) | Dem: 22.58%\* Ctrl: 72.41%\* | Dem: 12.7 (5.9; 0 to 25.) Ctrl: 27.5 (2.2; 23 to 30.) | nr |
| [84] | NPI | Brazilian Portuguese | Reliability (interrater & test-retest) | Outpatients. Author correspondence confirmed all community | N=36 | Alzheimer’s Disease | 78.78 (7.48) | 22%\* | 7.06 (6.92) | NPI-apathy total nr.  Severity: 5.31 (4.91) Frequency: 1 =33%, 2 =3%, 3 =64%. |
| [96] | NPI | Dutch | Hypothesis testing (divergent validity) | 83.33% community-dwelling | N=24 | Mixed sample:  Dementia: AD (N=19), FtD (N=1), MD (N=1)  Stroke (N=2)  Amnestic disorder (N=1) | 74.3 (10.4, nr) | 33.33%\* | 21.5 (4.6; 12 to 29)." | nr |
| [66] | NPI-A | English^ | Structural Validity; Internal consistency. | Outpatients. Author was unable to confirm whether community-dwelling. | N=124 | Dementia:  AD (N=62)  VaD (N=43)  MD of AD+VaD (N=19) | 79.8 (6.1; 61 to 91) | 21.77%\* | 22.6 (3.5; 13 to 29) | 8.89 (8.5, nr) |
| [87] | NPI-C | English^, French^, Greek^, Italian^, Hungarian^, Portuguese^, Spanish^ | Content validity (further item elicitation and Delphi study) | Item elicitation: n/a no participants  Delphi study: n/a professionals | Delphi study: N=8 | Experts in dementia research | n/a | n/a | n/a | n/a |
| Reliability (interrater); Hypothesis Testing (convergent) | 79.5% community-dwelling | N=128 | Alzheimer’s Disease | 75.7 (9.0; 54 to 94) | nr | 17.6 (7.0; 0 to 28). | NPI-C-apathy total nr.  AES (N=113): 33.1 (11.3; 0 to 51) |
| [88] | NPI-C | Portuguese | Reliability (interrater); Hypothesis Testing (convergent) | Author confirmed all community via correspondence | N=156 | Dementia | 76.7 (nr, nr) | 26.28%\* | 17.2 (nr, nr) | NPI-C-apathy total nr.  AI: 5.9 (nr, nr) |
| [52] | UPDRS | English | Development (item elicitation and review of comprehensibility) | n/a no participants involved | n/a | n/a | n/a | n/a | n/a | n/a |
| [100] | UPDRS | Spanish^ | Hypothesis Testing (convergent) | Outpatients – No confirmation from correspondence whether community-dwelling. | N=168  (convergent validity N=164) | Parkinson's Disease | 65.9 (9.8, nr) | 57% | 24.4 (5.4, nr) | nr |
| [99] | UPDRS | Norwegian^ | Hypothesis Testing (convergent) | nr. Participants were assessed in outpatient clinics, at home and in nursing homes. No confirmation from correspondence regarding proportion of community-dwellers. | N=89  (convergent N=58) | Parkinson’ Disease  (41.4% with cognitive impairment) | 74.2 (8.8, nr) | 44.8% | 23.0 (7.2, nr) | UPDRS-apathy item nr.  17% had apathy according to diagnostic criteria. |
| [98] | UPDRS | English^ | Hypothesis Testing (convergent) | Outpatients.  Confirmed all community via correspondence with authors | N=301 | Parkinson’s Disease | 67.8 (10.6; 30 to 90) | 63% | nr | 1.14 (1.1; 0 to 4)  AS =13.7 (6.9) range =0 to 31. AS≥14: 50% |
| [105,106] | mds-UPDRS | English | Development (Item elicitation [including adaptation of items from UPDRS to create mds-UPDRS], Pilot study) | nr | nr | Item elicitation: nr.  Pilot study:  Part 1: Patients (PD, N=80), carers (N=nr) and professionals (N=nr)  Part 2: Patients (N=32) and professionals (N=14) | nr | nr | nr | nr |
| [97] | mds-UPDRS | Hungarian | Hypothesis testing (convergent) | nr. Correspondence with author confirmed majority community. | N=584 | Parkinson’s Disease  PD with neurocognitive disorder (N=310)  PD with depression (N=217)  Apathy status: No apathy (N=477), Apathy (N=107) | *median (Q1 to Q3)*  No apathy: 67 (61 to 73.  Apathy: 68 (61 to 75) | No apathy: 60.2% Apathy: 52.3% | *median (Q1 to Q3)*  No apathy: 28, (27 to 29) Apathy: 27 (24 to 28) | *median (Q1 to Q3)*  LARS:  No apathy: -26 (-30 to -21)  Apathy: -15 (-22 to 5) |

Note: Where the study had used secondary data, the primary data sources were sought to gain the necessary information where it was not available in the article in question.

^ Assumed based on location of study and/ or nationality of participants.

\*Calculated by authors

Abbreviations: AD-RD, Alzheimer's Disease and Related Dementias Mood Scale; ACE, Addenbrooke’s Cognitive Examination; AD, Alzheimer’s Disease; AES-12PD, Apathy Evaluation Scale 12-item Parkinson’s Disease; AES-C, Apathy Evaluation Scale Clinician; AES-I, Apathy Evaluation Scale Informant; AES-S, Apathy Evaluation Scale Self; AI, Apathy Inventory; AI-C, Apathy Inventory Clinician; AI-I, Apathy Inventory Informant; ALS, Amyotrophic Lateral Sclerosis; AMI, Apathy Motivation Index; AS-S, Apathy Scale Self; AS-I, Apathy Scale Informant; bDAS, brief Dementia Apathy Scale; BMDS, Behavioural and Mood Disturbance Scale; BSSD, Behavioral Syndromes Scale for Dementia; CD, Current Depression; CDNOS, Cognitive Disorder Not Otherwise Specified; Ctrl, Healthy Controls; DAIR, Dementia Apathy Interview Rating; DAS, Dementia Apathy Scale; DAS-I, Dementia Apathy Scale Informant; DAS-S, Dementia Apathy Scale Self; Dem, Dementia; Dep, Depression; DEX, Dysexecutive Questionnaire; DLB, Dementia with Lewy Bodies; FrSBe-I, Frontal Systems Behavior Scale Informant; FtD, Frontotemporal Dementia; GDS, Geriatric Depression Scale; GIP, Behavioral Rating Scale for Psychogeriatric Inpatients ; IMD, Index of Mental Decline; KBCI, Key Behaviors Change Inventory; LARS, Lille Apathy Rating Scale; LARS-C, Lille Apathy Rating Scale Clinician; LARS-I, Lille Apathy Rating Scale Informant; LHS, Left Hemisphere Stroke; MCI, Mild Cognitive Impairment; MCS, Mild Cognitive Symptoms; MD, Mixed Dementia; mds-UPDRS, Movement disorder Society Unified Parkinson’s Disease Rating Scale; NPI, Neuropsychiatric Inventory; NPI-A, Neuropsychiatric Inventory Alternative; NPI-C, Neuropsychiatric Inventory Clinician; nr, not reported; PD, Parkinson’s Disease; PDa&d, Parkinson’s Disease with apathy and depression; PDa, Parkinson’s Disease with apathy; PDD, Parkinson’s Disease Dementia; PDd, Parkinson’s Disease with depression; PDDd, Parkinson’s Disease with dementia and depression; PDexclD, Parkinson’s Disease without dementia; PDexclDd, Parkinson’s Disease without dementia or depression; PS, Parkinsonian Symptoms; RD, Remitted Depression; RHS, Right Hemisphere Stroke; SCD, Subjective Cognitive Decline; UPDRS, Unified Parkinson’s Disease Rating Scale; VaD, Vascular Dementia.

## Table S.3. Risk of bias and results of development and content validity studies

| Reference | Measure | Met criteria? *(Y/N)* | Description | Relevance | | Comprehensiveness | | Comprehensibility | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Methodological quality | Result (quality rating) | Methodological quality | Result (quality rating) | Methodological quality | Result (quality rating) |
| [40,41] | AD-RD | Y | Development study: qualitative interviews for concept elicitation and expert review to refine the measure. | Apathy subscale: Inadequate | Construct of apathy is not clear. Items were all based on their mention by at least two carers (informal or formal) in qualitative interviews about how people with dementia express their mood. No justification was provided for the response options or recall period. (1?) | Doubtful | Patients or carers were not asked specifically about the measure. Expert review lead to reduction of items to avoid repetition. However, it was unclear what professionals were asked. (1?). | Doubtful | Patients or carers were not asked specifically about the measure. Expert review lead to modified instructions. However, it was unclear what professionals were asked (1?). |
| [33] | AES | Y | Development study and pilot study. | Inadequate | Construct of apathy is clear. Items were developed from the literature, professionals, and authors’ observations and opinions of people with apathy, but participants not involved in eliciting items and observations not reported on. (1?). |  |  | Doubtful | Unclear what participants were asked. 14 items were removed from the preliminary item pool due to poor comprehensibility. (1?). |
| [34] | AI | Y | Development study. | Inadequate | Construct of apathy is clear. Items were developed from the literature and diagnostic criteria, but participants not involved in eliciting items. (1?). |  |  |  |  |
| [35] | AMI | N | Development study. | Inadequate | Construct of apathy is clear. Items were developed from the relevant items of the LARS and by professionals. Participants were not involved in eliciting items. (1?). |  |  |  |  |
| [36] | AS | Y | Development study (Adaptation of AES to make AS.) | Inadequate | Construct of apathy is clear. Participants not involved in eliciting items. Most relevant items of AES were selected by 2 professionals (S. Starkstein personal, communication, October 01, 2018). (1?). | Doubtful | Pilot study conducted with participants with neurological disorders, but not published, so unable to rate. New items were included by 2 professionals (S. Starkstein personal, communication, October 01, 2018). (1?). | Inadequate | Pilot study conducted with participants with neurological disorders, but not published, so unable to rate. Some items were modified by 2 professionals (S. Starkstein personal, communication, October 01, 2018). (1?). |
| [90] | AS-I | Y | Content validity study. |  |  |  |  | Doubtful | Unclear what participants were asked. Participants showed good understanding and no modifications were required (1?). |
| [42] | BMDS | Y | Development study. | Inadequate | Constructs of behaviour and mood, and apathy were not clear. Items were developed from the literature and author opinion, but participants not involved in eliciting items. (1?). |  |  |  |  |
| [43] | BSSD | Y | Development study and pilot study. | Inadequate | Items were developed from professionals and previous measures, but participants not involved in eliciting items. (1?). | Doubtful | Multiple pilot studies conducted to refine scale, but methods and results not reported. (1?). | Doubtful | Multiple pilot studies conducted to refine scale, but methods and results not reported. (1?). |
| [37] | DAIR | Y | Development study and pilot study. | Doubtful | Construct of apathy is clear. Items refer to apathy, and were developed with participation from people with dementia and carers. No justification was provided for the response options or recall period. (1+/-). | Doubtful | Unclear what participants were asked. (1?). | Doubtful | Unclear what participants were asked. (1?). |
| [38] | DAS | Y | Development study. | Inadequate | Items were developed from existing scales and experts, but participants not involved in eliciting items. (1?). |  |  |  |  |
| [68] | FrSBe-11a | Y | Content validity: cognitive interviewing study |  |  |  |  | Doubtful | 27% items had no discrepancies, with 82% of items having acceptable discrepancy\*. However, participants do not appear to have been asked about the comprehensibility of instructions or response options. (1?) |
| [68] | FrSBe-14a | Y | Content validity: cognitive interviewing study |  |  |  |  | Doubtful | 21% items had no discrepancies, with  86% of items having acceptable discrepancy\*. However, participants do not appear to have been asked about the comprehensibility of instructions or response options. (1?) |
| [46,47] | GDS | N | Development and pilot study (as a measure of depression) | Inadequate | Items were developed from professionals, but participants not involved in eliciting items. (1?). |  |  | Doubtful | Reported that patients accepted the measure, but methods by which this was ascertained were unclear. (1?) |
| [49] | IMD | Y | Development | Inadequate | Items were developed from existing measures and professionals, but participants not involved in eliciting items. (1?). |  |  |  |  |
| [50,104] | KBCI | N | Development and pilot | Doubtful | Construct of apathy clear. Items were developed from the literature and interviews with patients, carers and professionals. Methods not clear. No justification for response options and recall period not clear. Patients and carers were later asked to rate the importance of items, and the majority were rated very or extremely important, but exact ratings not reported. (1+/-). | Doubtful | Patients and carers did not suggest any additional items. However, items were later removed after another phase in the development, so comprehensiveness may have changed. Method not clear. (1?) | Doubtful. | Patients and carers were asked about comprehensibility and no changes were suggested. Professionals were asked about comprehensibility and 15 items were re-worded. Methods and focus not clear (e.g. whether they were asked about each item, response options and recall period) (2?) |
| [39] | LARS | N | Development | Inadequate | Items were developed from Marin’s concept of apathy and authors’ clinical experience, but no systematic process and participants not involved in eliciting items. (1?). |  |  |  |  |
| [65] | LARS | Y | Pilot study | Doubtful | Participants asked about relevance, but results not reported. Methods and focus not clear (e.g. whether they were asked about each item, response options and recall period) (1?) |  |  | Doubtful | Participants asked about comprehensibility and format. Methods and focus not clear (e.g. whether they were asked about comprehensibility of instructions and response options as well as items) (1?) |
| [51] | NPI | N | Development and Delphi study | Inadequate | Items developed from the literature, but participants not involved in eliciting items. (1?). | Doubtful | Delphi panel of 10 professionals. Assessed “whether the essential elements of the behavior were captured” in each domain by rating screening and sub questions from 1 (well assessed) to 4 (poorly assessed). Apathy: screening questions mean score = 1.3; sub-questions mean score = 1.4. No assessment of comprehensiveness by participants. (1?) |  |  |
| [87] | NPI-C | Y | Content validity (adaptation) | Doubtful | New items added from symptoms listed by alternative measures. Items were selected that were consistent with diagnostic criteria 2009. Participants not involved in eliciting new items. (1?) | Doubtful | Delphi panel of 8 professionals. Unclear what was asked. (1?) | Doubtful | Delphi panel of 8 professionals. Unclear what was asked. (1?) |
| [52] | UPDRS | N | Development study | Inadequate | Expert group elicited items from existing measures, but participants not involved in eliciting items. (1?). |  |  | Inadequate | Authors reviewed comprehensiveness of preliminary items. Changes were made and final version does not appear to have been reviewed. (1?) |
| [105,106] | mds-UPDRS | N | Development (Adaptation of UPDRS but involved new item elicitation and pilot study) | Apathy subscale: Inadequate | Expert group elicited items from literature, existing measures, clinical experience and participant survey, though methods not described in sufficient detail. Justification provided for response options but not recall period. (1?). |  |  | Doubtful | Comprehensiveness of preliminary items was reviewed by participants and professionals in a qualitative, then quantitative study. Items, instructions and response options were assessed. Unsure if recall period discussed. Changes were made in the first round and then again in the second round. (1?) |

Note: Studies only listed if they assessed content validity in some way or were a study describing the development of a measure. Some studies have multiple citations as multiple articles or similar (e.g. PhD thesis) were published on the same study. Blank cells indicate this measurement property was not investigated by the study.

Quality of measurement property: Number of studies in parenthesis followed by rating: +, Sufficient; +/-, Inconsistent; -, Insufficient; ?, Indeterminate.

\* Acceptable discrepancy was defined by the authors of the study as less than 30% of participants interpreting the items meaning in the way it was intended [68].

Abbreviations: AD-RD, Alzheimer's Disease and Related Dementias Mood Scale; AES Apathy Evaluation Scale; AMI, Apathy Motivation Index; AI, Apathy Inventory; AS, Apathy Scale; AS-I, Apathy Scale Informant; BMDS, Behavioural and Mood Disturbance Scale; BSSD, Behavioral Syndromes Scale for Dementia; DAIR, Dementia Apathy Interview Rating; DAS, Dimensional Apathy Scale; FrSBe-11a, Frontal Systems Behavior Scale 11 item apathy subscale; FrSBe-14a, Frontal Systems Behavior Scale 14 item apathy subscale; GDS, Geriatric Depression Scale; IMD, Index of Mental Decline; KBCI, Key Behaviors Change Inventory; LARS, Lille Apathy Rating Scale; mds-UPDRS, Movement Disorder Society Unified Parkinson’s Disease Rating Scale; NPI, Neuropsychiatric Inventory; NPI-C, Neuropsychiatric Inventory Clinician; UPDRS, Unified Parkinson’s Disease Rating Scale.

Unable to obtain development articles for: Dysexecutive Questionnaire (DEX), FrSBe and Behavioral Rating Scale for Psychogeriatric Inpatients (GIP).

## Table S.4. Reviewer rating of content validity

| Measure | Relevance | | Comprehensiveness (quality rating) | Comprehensibility (quality rating) | Overall validity | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Older adults (quality rating) | Dementia & MCI (quality rating) | Older adults | Dementia & MCI |
| AD-RD | Unable to obtain the full list of items and instructions. |  |  |  |  |  |
| AES | 94% relevant to apathy. 100% relevant to older adults. 94% relevant to research context. Response options appropriate. Suggested recall period too long, but personalised recall period also possible. (1+). | AES-I & AES-S: 94% relevant to apathy. 100% relevant to people with dementia. 94% relevant to research context. Response options appropriate. Suggested recall period too long, but personalised recall period also possible. (1+).  AES-C: 94% relevant to apathy. 78% relevant to people with dementia, as some items based on where some items are rated based on patient free-recall. 94% relevant to research context. Response options appropriate. Suggested recall period too long, but personalised recall period also possible. (1+/- ). | 3 domains of apathy included. (1+). | AES-I & AES-S: 94% appropriately worded. 72% match response options. (1+/-).  AES-C: has additional guidance around this so AES-C response options deemed appropriate. (1+). | Sufficient  (AES-I & AES-S: 2+, 1+/-;  AES-C: 3+) | Sufficient (2+, 1+/-) |
| AI | 100% items relevant to apathy, older adults and the research context. Response options appropriate for AI-C and AI-I, but not for AI-S. Recall period referencing onset of disease not appropriate for older adults, but personalised recall period possible.  (Using the given recall period: 1+/-. Using the personalised recall period: 1+.) | 100% items relevant to apathy, people with dementia and the research context. Response options appropriate for AI-C and AI-I, but not for AI-S. Recall period of since onset of disease too long for people with dementia, but personalised recall period possible.  (Using the given recall period: 1+/-. Using the personalised recall period: 1+.) | 3 domains of apathy included. (1+). | 0% of items appropriately worded. (1-) | Inconsistent (Given recall period: 1+, 1-, 1+/-; Personalised recall period: 2+, 1-) | Inconsistent (Given recall period: 1+, 1-, 1+/-; Personalised recall period: 2+, 1-) |
| AMI | 78% relevant to apathy. 100% relevant to older adults. 100% relevant to research context. Response options and recall period appropriate. (1+/-). | 78% relevant to apathy and to older adults. 100% relevant to research context. Response options and recall period appropriate. (1+/-). | 3 domains of apathy included. (1+). | 100% of items appropriately worded. 100% match response options. (1+). | Sufficient (2+, 1+/-) | Sufficient (2+, 1+/-) |
| AS | 93% relevant to apathy. 93% relevant to older adults.100% relevant to research context. Response options appropriate. Recall period too long. (1+). | 93% relevant to apathy. 100% relevant to people with dementia and the research context. Response options appropriate. Recall period too long. (1+) | 3 domains of apathy included. (1+). | 93% of items appropriately worded. 57% match response options (1+/-) | Sufficient (2+, 1+/-) | Sufficient (2+, 1+/-) |
| BMDS | 55% relevant to apathy. 100% relevant to older adults and the research context. Response options appropriate. Recall period uncertain. (1+/-). | 55% relevant to apathy. 100% relevant to people with dementia and the research context. Response options appropriate. Recall period uncertain. (1+/-). | Emotional dimension missing. (1-). | 100% of items appropriately worded, but combination with response options produces double negatives. (1+/-). | Inconsistent (1-, 2+/-) | Inconsistent (1-, 2+/-) |
| BSSD | 71% relevant to apathy. 100% relevant to older adults and research context. 14% response options appropriate. Recall period appropriate. (1+/-) | 71% relevant to apathy. 100% relevant to people with dementia and research context. 14% response options appropriate. Recall period appropriate. (1+/-) | 3 domains of apathy included. (1+). | 86% of items (questions directed at informants) appropriately worded. 100% match response options. (1+). | Sufficient (2+, 1+/-) | Sufficient (2+, 1+/-) |
| DAIR | 94% items relevant to apathy. 0% relevant for healthy older adults due to mandatory follow-up question relating to “illness”. Response options appropriate. Recall period too long. (1+/-). | 94% items relevant to apathy. 100% relevant for people with dementia. Response options appropriate. Recall period too long. (1+). | 3 domains of apathy included. (1+). | 100% items appropriately worded. 81% match the response options. (1+/-). | Inconsistent (1+, 2+/-). | Sufficient (2+, 1+/-). |
| DAS | DAS: 79% items relevant to apathy.  bDAS: 67% items relevant to apathy  Both versions: 100% relevant to older adults. Response options appropriate. Recall period too long. (1+/-). | DAS: 79% items relevant to apathy.  bDAS: 67% items relevant to apathy  Both versions: 100% relevant to people with dementia. Response options appropriate. Recall period too long. (1+/-). | 3 domains of apathy included. (1+). | 100% of items appropriately worded. 100% match response options. (1+). | Sufficient (2+, 1+/-). | Sufficient (2+, 1+/-). |
| DEX | 63% items relevant to apathy. 100% relevant to older adults and research context. Complete response options not available. Recall period appropriate. (1+/-).\* | 63% items relevant to apathy. 100% relevant to people with dementia and research context. Complete response options not available. Recall period appropriate. (1+/-).\* | 3 domains of apathy included. (1+).\* | Full wording not available, but 75% of items appear appropriately worded. Complete response options not known. (1?). | Inconsistent (1+, 1+/-, 1?)\* | Inconsistent (1+/-, 1+/-, 1?)\* |
| FrSBe | FrSBe-6a: 83% relevant to apathy. 100% relevant to older adults.  FrSBe-11a: 82% relevant to apathy. 91% relevant to older adults.  FrSBe-14a: 86% relevant to apathy. 93% relevant to older adults.  And all versions: 100% relevant to research context. Response options not available. Recall period not appropriate for older adults. (1+/-).\* | FrSBe-6a: 83% relevant to apathy. 100% relevant to older adults  FrSBe-11a: 82% relevant to apathy. 91% relevant to people with mild dementia.  FrSBe-14a: 86% relevant to apathy. 93% relevant to mild dementia.  And all versions: 100% relevant to research context. Response options not available. Recall period not appropriate for people with dementia. (1+/-).\* | All versions: 3 domains of apathy included. (1+).\* | 6a: Full wording not available, but items suggests that 67% appropriately worded. Response options not available. (1?).  11a: Full wording not available, but items suggests that 91% appropriately worded. Response options not available. (1?).  14a: Full wording not available, but items suggests that 86% appropriately worded. Response options not available. (1?). | Inconsistent (1+, 1+/-, 1?)\* | Inconsistent (1+, 1+/-, 1?)\* |
| GDS-3a | 67% of items are relevant to apathy. All items relevant to older adults and the research context. Dichotomous response options not appropriate. Recall period appropriate. (1+/-). | 67% of items are relevant to apathy. All items relevant to people with dementia and the research context. Dichotomous response options not appropriate. Recall period appropriate. (1+/-). | Emotional dimension of apathy is missing (1-). | 100% appropriately worded and match response options. (1+). | Inconsistent (1+, 1-, 1+/-) | Inconsistent (1+, 1-, 1+/-) |
| GDS-6a | 50% of items are relevant to apathy. All items relevant to older adults and the research context. Dichotomous response options not appropriate. Recall period appropriate. (1+/-). | 50% of items are relevant to apathy. All items relevant to older adults and the research context. Dichotomous response options not appropriate. Recall period appropriate. (1+/-). | 3 domains of apathy included. (1+). | 100% appropriately worded and match response options. (1+). | Sufficient (2+, 1+/-) | Sufficient (2+, 1+/-) |
| GIP-9a | 44% of items relevant to apathy. 89% relevant to older adults in the community. 100% relevant to research context. Recall period appropriate. Response options not available. (1+/-).\* | 44% of items relevant to apathy. 89% of items relevant to people with dementia in the community. 100% relevant to research context. . Recall period appropriate. Response options not available. (1+/-).\* | Emotional dimension of apathy is missing. (1-).\* | Full wording and official English translation of items not available, but authors translation suggest 89% appropriately worded. Response options not available. (1?).\* | Inconsistent (1-, 1+/- 1?) | Inconsistent (1-, 1+/- 1?) |
| IMD | 100% of items relevant to apathy, older adults and the research context. Response options and recall period not available. (1?). | 100% of items relevant to apathy, people with dementia and the research context. Response options and recall period not available. (1?). | 3 domains of apathy included. (1+). | Full wording not available, but items suggest 33% appropriately worded. Response options not available. (1?). | Indeterminate (1+, 2?) | Indeterminate (1+, 2?) |
| KBCI-10a | 90% of items relevant to apathy. 80% of items relevant to older adults. All items relevant to research context. Response options appropriate. Recall period not available. (1+/-). | 90% of items relevant to apathy. 80% of items relevant to people with dementia. All items relevant to research context. Response options appropriate. Recall period not available. (1+/-). | 3 domains of apathy included. (1+). | 80% of items appropriately worded. 100% match response options. (1+/-). | Inconsistent (1+, 2+/-) | Inconsistent (1+, 2+/-) |
| LARS | 94% of items relevant to apathy. 100% relevant to older adults. Response options appropriate. Recall period too long. (1+). | 94% of items relevant to apathy. 94% relevant to people with dementia. Response options appropriate. Recall period too long. (1+). | 3 domains of apathy included. (1+). | 87% appropriately worded. 100% match response options. (1+). | Sufficient (3+) | Sufficient (3+) |
| NPI (original) | 100% of items relevant to apathy, older adults and the research context. Response options appropriate. Suggested recall period too long, but personalised recall period also possible. (1+). | 100% of items relevant to apathy, people with dementia and the research context. Response options appropriate. Suggested recall period too long, but personalised recall period also possible. (1+). | Emotional dimension of apathy is missing from the screening questions. No dimensions are rated separately. (1-). | Assessments of frequency and severity are based on multiple symptoms, so could be considered a double barrelled question and therefore not appropriately worded. However carers are advised to rate the worst one. 100% match the response options. (1+). | Inconsistent (2+, 1-) | Inconsistent (2+, 1-) |
| NPI-A | Unable to obtain full instructions and guidance. |  |  |  |  |  |
| NPI-C | 100% of items relevant to apathy, older adults and the research context. Response options appropriate. Recall period too long. (1+). | 100% of items relevant to apathy, people with dementia and the research context. Response options appropriate. Recall period too long. (1+). | 3 domains of apathy included. (1+). | Assessments of frequency and severity are based on multiple symptoms, so could be considered a double barrelled question and therefore not appropriately worded. However carers are advised to rate the worst one. 100% match the response options. (1+). | Sufficient (3+) | Sufficient (3+) |
| UPDRS | 100% relevant to apathy, older adults and research context. (Note: only 1 item). Response options appropriate. Recall period not clear. (1+). | 100% relevant to apathy, people with dementia and research context. (Note: only 1 item). Response options appropriate. Recall period not clear. (1+). | Emotional domain of apathy missing. Cognitive and Behavioural elements included but not rated separately. (1-). | Item wording is not given, or could not be obtained; only the heading is provided, so it is unclear if it matches the response options. (1?). | Inconsistent (1+, 1-, 1?) | Inconsistent (1+, 1-, 1?) |
| mds-UPDRS | 100% relevant to apathy, older adults and the research context. (Note: only 1 item). Response options and recall period appropriate. (1+). | 100% relevant to apathy, people with dementia and the research context. (Note: only 1 item). Response options and recall period appropriate. (1+). | Emotional domain of apathy missing. Cognitive and Behavioural elements included but not rated separately. (1-). | 100% appropriate worded and match response options. (1+). | Inconsistent (2+, 1-) | Inconsistent (2+, 1-) |

\*based on list of apathy items presented by another publication (DEX [81]; FrsBE [64,68]; GIP [107])

Quality of measurement property: Number of studies in parenthesis followed by rating: +, Sufficient; +/-, Inconsistent; -, Insufficient; ?, Indeterminate.

Abbreviations: AD-RD, Alzheimer's Disease and Related Dementias Mood Scale; AES Apathy Evaluation Scale; AES-C, Apathy Evaluation Scale Clinician; AES-I, Apathy Evaluation Scale Informant; AES-S, Apathy Evaluation Scale Self; AI, Apathy Inventory; AI-C, Apathy Inventory Clinician; AI-I, Apathy Inventory Informant; AI-S, Apathy Inventory Self; AMI, Apathy Motivation Index; AS, Apathy Scale; b-DAS, brief-Dimensional Apathy Scale; BMDS, Behavioural and Mood Disturbance Scale; BSSD, Behavioral Syndromes Scale for Dementia; DAIR, Dementia Apathy Interview Rating; DAS, Dimensional Apathy Scale; DEX, Dysexecutive Questionnaire; FrSBe, Frontal Systems Behavior Scale; FrSBe-6a, Frontal Systems Behavior Scale 6-item apathy subscale; FrSBe-11a, Frontal Systems Behavior Scale 11-item apathy subscale; FrSBe-14a, Frontal Systems Behavior Scale 14-item apathy subscale; GDS-3a, Geriatric Depression Scale 3 item apathy subscale; GDS-6a, Geriatric Depression Scale 6 item apathy subscale; GIP-9a, Behavioral Rating Scale for Psychogeriatric Inpatients 9 item apathy subscale; IMD, Index of Mental Decline; KBCI-10a, Key Behaviors Change Inventory 10 item apathy subscale; LARS, Lille Apathy Rating Scale; MDS-UPDRS, Movement Disorder Society-Sponsored Revision of the Unified Parkinson’s Disease Rating Scale; NPI, Neuropsychiatric Inventory; NPI-A, Neuropsychiatric Inventory Alternative; NPI-C, Neuropsychiatric Inventory Clinician; UPDRS, Unified Parkinson’s Disease Rating Scale

## Table S.5. Risk of bias and results of studies of remaining measurement properties

| Reference | Measure | Structural validity | | Internal consistency | | Reliability | | Measurement error | | | Hypothesis testing | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Methodological quality | Result (% variance explained) [quality rating] | Methodological quality | Result (quality rating) | Methodological quality | Result (quality rating) | Methodological quality | Result (quality rating) | Methodological quality | | Result (quality rating) |
| [40] | AD-RD |  |  |  |  | 1 Doubtful. | r=.72 (1+). |  |  |  | |  |
| [78] | AES-12PD |  |  | 3 Very good. | α=.90 to .92 (3+) |  |  |  |  | 3 Adequate. 1 Very Good. | | 3 met hypothesis (3+). 1 did not meet hypothesis (1-). |
| [53] | AES-C | 1 Doubtful. | 3 factors (57.06%): Apathy (40.02%); Novelty Seeking (9.35%); Insight & social (7.68%). [1+] | 1 Very good. | α=.90. (1+). | 2 Doubtful. | r=.88 to .86 (2+). |  |  | 2 Inadequate. 2 Very Good. | | 4 met hypothesis (4+). |
| [54] | AES-C | 1 Doubtful. | 2 factors (51.1%): Apathy (42.4%); Interest (8.7%). [1+] |  |  |  |  |  |  | 1 Inadequate. 1 Doubtful. 2 Adequate. | | 2 met hypothesis (2+). 2 did not meet hypothesis (2-). |
| [55] | AES-C | 1 Inadequate. | 3 factors (84.17^): Interest & Motivation (39.72%^); Task Completion (29.67%^); Insight (14.78%^). [1-] | 1 Very good. | α=.93. (1+). |  |  |  |  | 1 Inadequate. 1 Doubtful. 1 Very Good. | | 3 met hypothesis (3+). |
| [33] | AES-C | 1 Inadequate. | 3 factors: Apathy (32-53%); Novelty Seeking (5-10%); Insight & dependency (7-8%). [1?] | 1 Very good. | α=.90. (1+). | 1 Doubtful. 1 Adequate. | r=.88 (1+).  ICC= .94 (+). |  |  | 3 Inadequate. 1 Doubtful. 1 Adequate. 4 Very Good. | | 5 met hypothesis (5+). 1 did not meet hypothesis (1-). 3 insufficient information (3?). |
| [54] | AES-I | 1 Doubtful. | 2 factors (54.4%): Interest (45.1%); Apathy (9.3%). [1+] |  |  |  |  |  |  | 1 Inadequate. 1 Doubtful. 2 Adequate. | | 2 met hypothesis (2+). 2 did not meet hypothesis (2-). |
| [55] | AES-I |  |  | 1 Very good. | α=.89. (1+). |  |  |  |  | 1 Inadequate. 1 Doubtful. 1 Very Good. | | 3 met hypothesis (3+). |
| [33] | AES-I | 1 Inadequate. | 3 factors: Apathy (32-53%); Novelty Seeking (5-10%); Insight & dependency (7-8%). [1?] | 1 Very good. | α=.94. (1+). | 1 Doubtful. | r=.94 (1+). |  |  | 3 Inadequate. 1 Doubtful. 1 Adequate. 4 Very Good. | | 4 met hypothesis (4+). 2 did not meet hypothesis (2-). 3 insufficient information (3?). |
| [56] | AES-I | 1 Doubtful. 1 Adequate. | 2 factors (62.56%^): Factor 1 (56.2%); Factor 2 (6.36%). [1+].  1 factor (62.8%). [1+]. | 1 Very good. | α=.95. (1+). |  |  | n/a | SEM=2.9. (1?). | 2 Very Good. | | 2 met hypothesis (2+). |
| [89] | AES-I |  |  |  |  |  |  |  |  | 1 Adequate. 2 Very good. | | 1 met hypothesis (1+). 2 did not meet hypothesis (2-). |
| [71] | AES-I |  |  | 1 Very good. | α=.88. (1+). |  |  |  |  |  | |  |
| [71] | AES-I-16 |  |  | 1 Very good. | α=.90. (1+). |  |  |  |  | 1 Adequate. 1 Very Good | | 1 met hypothesis (1+).  1 did not meet hypothesis (1-). |
| [55] | AES-S |  |  | 1 Very good. | α=.90. (1+). |  |  |  |  | 1 Inadequate. 1 Doubtful. 1 Very Good. | | 1 met hypothesis (1+). 1 did not meet hypothesis (1-). 1 insufficient information (1?) |
| [54] | AES-S | 1 Doubtful. | 2 factors (43.3%^): Apathy (36.4%); Other (6.9%) [1+] |  |  |  |  |  |  | 1 Inadequate. 1 Doubtful. 2 Adequate. | | 2 met hypothesis (2+). 2 did not meet hypothesis (2-). |
| [33] | AES-S | 1 Inadequate. | 3 factors: Apathy (32-53%); Novelty Seeking (5-10%); Insight & dependency (7-8%). [1?] | 1 Very good. | α=.86. (1+). | 1 Doubtful. | r=.76 (1+). |  |  | 3 Inadequate. 1 Doubtful. 1 Adequate. 4 Very Good. | | 5 met hypothesis (5+). 1 did not meet hypothesis (1-). 3 insufficient information (3?) |
| [56] | AES-S | 1 Doubtful. 1 Adequate. | 2 factors (61.69%^): Factor 1 (55.37%); Factor 2 (6.32%). [1+]  1 factor (61.2%). [1+]. | 1 Very good. | α=.95. (1+). |  |  | n/a | SEM=2.7. (1?). | 2 Very Good. | | 2 met hypothesis (2+). |
| [89] | AES-S |  |  |  |  |  |  |  |  | 1 Adequate. 2 Very good. | | 1 met hypothesis (1+). 2 did not meet hypothesis (2-). |
| [57] | AES-S | 2 Doubtful. | 3 factors (58%): Apathy (38.27%); Friendship (10.86%); Other (8.88%) [1+].  3 factors (59.54%; variance explained per factor not reported.) [1?] | 2 Very good. | α=.90 to .92. (2+). |  |  |  |  | 2 Doubtful. 4 Adequate. 4 Very Good. | | 5 met hypothesis (5+). 3 did not meet hypothesis (3-). |
| [80] | AI-C |  |  |  |  | 1 Doubtful. | ICC=.97 (1+). |  |  | 1 Inadequate. | | 1 met hypothesis (1+). |
| [72] | AI-C |  |  | 1 Doubtful. | α=.83. (1?). |  |  |  |  |  | |  |
| [34] | AI-I |  |  | 1 Doubtful. | α=.84. (1?). | 1 Doubtful. 1 Inadequate. | Kappa= .96 to .99 (2+). |  |  | 1 Adequate. 3 Very Good. | | 3 met hypothesis (3+). 1 did not meet hypothesis (1-). |
| [72] | AI-I |  |  | 1 Doubtful. | α=.83. (1?). |  |  |  |  |  | |  |
| [34] | AI-S |  |  |  |  |  |  |  |  | 3 Very Good. | | 1 met hypothesis. (1+). 2 did not (2-) |
| [72] | AI-S |  |  | 1 Doubtful. | α=.61. (1?). |  |  |  |  |  | |  |
| [67] | AMI |  |  | \* | α=.86 |  |  |  |  | 2 Adequate. | | 2 did not meet hypothesis (2-). |
| [58] | AS-HC | 1 Very Good. | 1 factor CFI=1.00, RMSEA=0.00. [1+] | 1 Very Good | α=.94. (1+). |  |  |  |  | 1 Very Good. | | 1 did not meet hypothesis (1-). |
| [90] | AS-I |  |  |  |  |  |  |  |  | 1 Inadequate. 1 Doubtful. 1 Very Good. | | 1 met hypothesis (1+). 1 did not meet hypothesis (1-). 1 insufficient information available (1?). |
| [59] | AS-S | 1 Doubtful. | 13-item: 3 factors (55.61%). Variance explained per factor not reported. [1?] | 2 Doubtful. | 14 item version: α=.82.  13 item version: α=.85. (2?). |  |  |  |  |  | |  |
| [36] | AS-S |  |  | 1 Doubtful. | α=.76. (1?). | 2 Doubtful. | r=.81 to .90. (2+). |  |  | 1 Doubtful. | | 1 met hypothesis (1+). |
| [58] | AS-S | 1 Very Good. | 1 factor. CFI=1.00, RMSEA=0.00. [1+]. |  |  |  |  |  |  |  | |  |
| [60] | AS-S | 2 Adequate. | 14-item: 2 factors (57.7%): Cognitive-Behavioural (24.2%); Apathy and insight (15.05%). [1-].  13-item: 2 factors (41.7%) Variance explained per factor not reported. [1?] | 2 Doubtful. | 14 item: α=.69.  13 item: α=.74. (2?). |  |  |  |  | 1 Adequate. 2 Very Good. | | 3 met hypothesis (3+). |
| [77] | AS-S |  |  | 1 Inadequate. | Guttman’s λ = .89. (1?). | 1 Inadequate. | ICC=.78 (1+). | n/a | SEM = 2.34. (1?). | 1 Doubtful. 2 Very Good. | | 1 met hypothesis (1+). 1 did not meet hypothesis (1-). 1 insufficient information (1?). |
| [61] | AS-S | 1 Very Good. 1 Adequate. | AS-S: 3 factors (nr). [1+/-].  11 item: 2 factors: 54.1% of variance explained. [1-]. | 11-item: 1 Inadequate. | 11 item: α=.77 (1?) |  |  |  |  |  | |  |
| [91] | AS-S |  |  | 1 Doubtful. | α=.78. (1?). |  |  |  |  | 1 Inadequate. 2 Doubtful. 1 Very good. | | 3 met hypothesis (3+). 1 did not meet hypothesis (1-). |
| [42] | BMDS |  |  |  |  | 1 Inadequate. | r=.90. (1+) |  |  |  | |  |
| [43] | BSSD |  |  | 1 Doubtful. | α=.82 to .83 (1?) | 1 Inadequate. 3 Doubtful. | ICC=.65 to .85. (2+, 2-). |  |  | 2 Inadequate. 1 Doubtful. 1 Very Good. | | 1 met hypothesis (1+). 1 did not meet hypothesis (1-). 2 insufficient information (2?). |
| [37] | DAIR | 1 Adequate. | 1 factor (38%) [1+] | 1 Very Good. | α=.89. (1+). | 1 Inadequate. | r=.85 (1+) | 1 Doubtful. | 100% agreement (1+). | 2 Inadequate. 2 Very Good. | | 3 met hypothesis (3+). 1 did not meet hypothesis (1+). |
| [75] | DAS-I |  |  | \* | α=.92 |  |  |  |  | 2 Adequate. | | 1 met hypothesis (1+). 1 did not meet hypothesis (1-). |
| [74] | DAS-I |  |  | \* | α=.93 |  |  |  |  | 2 Adequate. | | 2 met hypothesis (2+). |
| [75] | DAS-S |  |  | \* | α=.84 |  |  |  |  | 2 Adequate. | | 1 met hypothesis (1+). 1 did not meet hypothesis (1-). |
| [74] | DAS-S |  |  | \* | α=.85 |  |  |  |  | 2 Adequate. | | 2 met hypothesis (2+). |
| [62] | DAS-S | \* | 3 factors (45.87^) Organisation & planning (28.21%); Initiation (9.76%); Emotional (7.90%). | \* | α=.87 |  |  |  |  | 4 Adequate. 2 Very Good. | | 4 met hypothesis (4+). 2 did not meet hypothesis (2-). |
| [63] | bDAS | \* | Item Hi=.40 to .76. No other fit measures reported. |  |  |  |  |  |  |  | |  |
| [76] | bDAS |  |  | \* | α=.81. | 1 Inadequate. | ICC=.84 (1+). |  |  |  | |  |
| [81] | DEX |  |  |  |  | 1 Doubtful. | ICC=.93 (1+). |  |  | 1 Inadequate. 1 Adequate. 2 Very Good. | | 2 met hypothesis (2+). 2 did not meet hypothesis (2+) |
| [64] | FrSBe-6a |  |  | 1 Doubtful. | α=.88. (1?). |  |  |  |  |  | |  |
| [68] | FrSBe-11a |  |  | 1 Doubtful. | α=.83. (1?). |  |  |  |  |  | |  |
| [68] | FrSBe-14a |  |  | 1 Doubtful. | α=.88. (1?). |  |  |  |  | 6 Doubtful. | | 5 met hypothesis (5+). 1 did not meet hypothesis (1-). |
| [64] | FrSBe-14a | 1 Inadequate. | 1 Factor specified: 12 out of 14 items had loadings >.40. (nr). [1?]. | 1 Doubtful. | α=.80. (1?). |  |  |  |  |  | |  |
| [92] | GDS-3a |  |  |  |  |  |  |  |  | 2 Adequate. | | 2 did not meet hypothesis. (2+). |
| [69] | GDS-6a |  |  | 1 Doubtful. | α=.51 (1?). |  |  |  |  | 1 Doubtful. 2 Adequate. | | 3 met hypothesis (3+). |
| [82] | GIP-apathy subscale |  |  |  |  | 1 Doubtful. | ICC=.72 (1+). | n/a | SEM=1.22. (1?) |  | |  |
| [82] | GIP-apathy domain |  |  |  |  | 1 Doubtful. | ICC=.83 (1+.) | n/a | SEM=1.38. (1?) |  | |  |
| [49] | IMD |  |  |  |  |  |  |  |  | 1 Inadequate. 3 Doubtful. | | 3 met hypothesis (3+). 1 insufficient information (1?). |
| [93] | KBCI |  |  |  |  |  |  |  |  | 1 Inadequate. 1 Doubtful. 5 Adequate. | | 6 met hypothesis (6+). 1 did not meet hypothesis (1-). |
| [83] | LARS-C |  |  |  |  | 2 Doubtful. | ICC=.94 to .99 (2+). |  |  | 2 Inadequate. 2 Doubtful. 5 Very Good. | | 7 met hypothesis (7+). 1 did not meet hypothesis (1-). 1 insufficient information (1?). |
| [65] | LARS-C | \* | 4 factors (67.5%): intellectual curiosity (nr); emotion (nr); action-initiation (nr); self awareness (nr). | \* | α=.81. (\*). | 2 Doubtful. | ICC= .97. (1+). Kappa = .93 (1+). |  |  | 1 Inadequate. 2 Adequate. | | 2 met hypothesis (2+). 1 did not meet hypothesis (1-). |
| [70] | LARS-I |  |  | \* | α=.87. (\*). | 2 Doubtful. | ICC =.99. (1+).  . (1+)ICC =.99. (1+). |  |  | 2 Adequate. | | 2 met hypothesis (2+). |
| [84] | NPI |  |  |  |  | 1 Doubtful. 1 Inadequate. | ICC = .67 (1-).  rs= .53 (1-). |  |  |  | |  |
| [94] | NPI |  |  |  |  |  |  |  |  | 2 Doubtful. | | 2 insufficient information (2?). |
| [51] | NPI |  |  |  |  |  |  |  |  | 1 Doubtful. | | 1 insufficient information (1?). |
| [85] | NPI |  |  |  |  | 1 Inadequate | r=.96 (1+). |  |  | 1 Doubtful. | | 1 insufficient information (1?). |
| [73] | NPI |  |  | 1 Doubtful. | α=.82 (1?) | 1 Doubtful. 1 Inadequate. | ICC=.87. (1+). r=.76 (1+). |  |  | 1 Inadequate. 1 Doubtful. 1 Very Good. | | 1 met hypothesis (1+). 1 did not meet hypothesis (1-). 1 insufficient information (1?). |
| [95] | NPI |  |  |  |  |  |  |  |  | 1 Inadequate. | | 1 did not meet hypothesis (1-). |
| [79] | NPI |  |  | 1 Doubtful. | α=.83 (1?) | 1 Doubtful. | Kendell CC= 1.00 (1+). |  |  | 1 Inadequate. | | 1 did not meet hypothesis (1-). |
| [96] | NPI |  |  |  |  |  |  |  |  | 1 Very Good. | | 1 did not meet hypothesis (1-). |
| [86] | NPI |  |  |  |  | 1 Doubtful. | ICC=.99 (1+). |  |  |  | |  |
| [66] | NPI-A | 1 Adequate. | 1 factor (66%). [1+]. | 1 Very Good. | α=.91 (1+) |  |  |  |  |  | |  |
| [87] | NPI-C |  |  |  |  | 1 Doubtful. | Item ICC= .74 to .89 (1+). |  |  | 1 Adequate. | | 1 did not meet hypothesis (1-). |
| [88] | NPI-C |  |  |  |  | 1 Doubtful. | ICC=.87 (1+). |  |  | 1 Adequate. | | 1 met hypothesis (1+). |
| [97] | mds-UPDRS |  |  |  |  |  |  |  |  | 1 Very Good. | | 1 met hypothesis (1+). |
| [98] | UPDRS |  |  |  |  |  |  |  |  | 1 Adequate. | | 1 met hypothesis (1+). |
| [99] | UPDRS |  |  |  |  |  |  |  |  | 1 Inadequate. 2 Very Good. | | 3 did not meet hypothesis (3-). |
| [100] | UPDRS |  |  |  |  |  |  |  |  | 1 Very Good. | | 1 met hypothesis (1+). |

Blank cells indicate this measurement property was not investigated.

\*Was assessed by the study, but methodological quality rating nor quality rating of result conducted, as the measure is based on a formative model.

^ Value calculated by review team based on information provided in the article.

Quality of measurement property: Number of studies in parenthesis followed by rating: +, Sufficient; +/-, Inconsistent; -, Insufficient; ?, Indeterminate.

Abbreviations: +, Sufficient; -, Insufficient; ?, Indeterminate; AD-RD, Alzheimer's Disease and Related Dementias Mood Scale; AES-12PD, Apathy Evaluation Scale for Parkinson Disease; AES-C, Apathy Evaluation Scale Clinician; AES-I, Apathy Evaluation Scale Informant; AES-I-16, Apathy Evaluation Scale Informant 16 item version; AES-S, Apathy Evaluation Scale Self; AI-C, Apathy Inventory Clinician; AI-I, Apathy Inventory Informant; AI-S, Apathy Inventory Self; AMI, Apathy Motivation Index; AS-HC, Apathy Scale Home Care; AS-I, Apathy Scale Informant; AS-S, Apathy Scale Self; b-DAS, brief-Dimensional Apathy Scale; BMDS, Behavioural and Mood Disturbance Scale; BSSD, Behavioral Syndromes Scale for Dementia; DAIR, Dementia Apathy Interview Rating; DAS-I, Dimensional Apathy Scale Informant; DAS-S, Dimensional Apathy Scale Self; DEX, Dysexecutive Questionnaire; FrSBe-6a, Frontal Systems Behavior Scale 6-item apathy subscale; FrSBe-11a, Frontal Systems Behavior Scale 11-item apathy subscale; FrSBe-14a, Frontal Systems Behavior Scale 14-item apathy subscale; GDS-3a, Geriatric Depression Scale 3 item apathy subscale; GDS-6a, Geriatric Depression Scale 6 item apathy subscale GIP, Behavioral Rating Scale for Psychogeriatric Inpatients; IMD, Index of Mental Decline; KBCI, Key Behaviors Change Inventory; LARS-C, Lille Apathy Rating Scale Clinician; LARS-I, Lille Apathy Rating Scale Informant; mds-UPDRS, Movement Disorder Society-Sponsored Revision of the Unified Parkinson’s Disease Rating Scale; NPI, Neuropsychiatric Inventory; NPI-A, Neuropsychiatric Inventory Alternative; NPI-C, Neuropsychiatric Inventory Clinician; nr, not reported; UPDRS, Unified Parkinson’s Disease Rating Scale

Where there is no rating available for the researcher, this means it was not possible to obtain sufficient information regarding the measure to assess its content validity. Ratings of content validity are for both people with dementia or MCI and older adults unless otherwise specified.