Supplementary Material 5. Outcome categories and measures within studies assessing effectiveness of interventions for Plwd

Study, design	Outcome category	Outcome measure	Definition/scale items if reported or identified
Improving staff information, knowledge and skills			
Mador (2004) **RCT**	Behaviour (level of agitation)	Pittsburgh Agitation Scale (PAS)	Levels of patient agitation were measured using the Pittsburgh Agitation Scale (PAS) (Rosen et al., 1994). The ward nurse caring for the patient scored the highest level of agitation, on a four-point scale, seen during their shift for each of four behaviours; vocal and motor agitation, aggressiveness and resisting care. There is a minimum score of 0 and maximum of 16.
Miller (2004) **BA**	Discomfort	Modified 6-item Discomfort Screen - Dementia Alzheimer's Type (DS-DAT)	Discomfort as a negative emotional and physical state, because of age-associated chronic health conditions, postoperative and treatment associated pain, immobility-related problems, and fatigue. DS-DAT was developed on the basis of interviews with experienced nurses who cared for persons with dementia. The scale has nine items that are made operational by 40 defining characteristics. Items are scored 0, 1, 2, or 3 depending on the number of defining characteristics present and their intensity and duration during a 5-minute observation period. Unclear which 6 items are used in modified version
Beernaert (2017) **cRCT**	Comfort (around dying, assessed by nurses and assessed by family carers)	End-of-Life in Dementia–Comfort Assessment in Dying (CAD-EOLD; range 14–42)	Answers to items of this scale indicated the condition of the care recipient during the dying process. The possible scores of the resulting scale range from 14 to 42 with a higher score indicating *better* comfort level. 14 items (scores 1–3): discomfort, pain, restlessness, shortness of breath, choking, gurgling, difficulty swallowing, fear, anxiety, crying, moaning, serenity, peace, and calm
	Discomfort (symptom management assessed by nurses and by family)	Modified version of the End-of-Life in Dementia-Symptom Management (SM-EOLD) scale (range 0-27)	The scale quantifies the frequency a resident experiences the following nine symptoms and signs during the previous 90 days: pain, shortness of breath, depression, fear, anxiety, agitation, calm, skin breakdown, and resistance to care. Modified version- Nine items (scores 0-3). Higher score suggesting *less* symptom control
	Discomfort (symptoms and care needs in the last 3	Palliative Care Outcome scale (range 0-40)	Ten items (scores 0–4). Higher score suggesting *more* symptoms and care needs.

	days of life assessed by nurses)		The Palliative Care Outcome Scale (POS) was developed using components of other measures which appear to have worked well, and includes aspects about pain and symptom control, patient and family psychosocial needs, and communication and information. These are assessed over the previous three days
Activity-based interventions for people with dementia			
DiNapoli (2016) **RCT**	Quality of life	The Dementia Quality of Life instrument (DQoL)	The Dementia Quality of Life instrument is 29-item direct-interview instrument used to assess QoL in mild to moderate dementia participants. It has two distinct factors: negative affect (11 items) and General DQoL (comprising the remaining 18 items). To assess QOL, this study used an average score of the items that comprise the General DQoL (a=0.81). A global item of QOL (i.e., ‘Overall, how would you rate your quality of life?’) was also used to determine participant inclusion. The General DQoL and global item scores range from 1 (never or bad) to 5 (very often or excellent), with higher scores indicating greater QoL
	Behaviour (cognitive, psychiatric, and behavioural symptoms; BPSD)	Neurobehavioral Rating Scale-Revised	The NRS-R is a 29-item direct-interview instrument and each item is scored on a scale of 0 (Absent) to 3 (Severe). An overall score was calculated by summing the responses, with higher scores indicating greater presence of behavioural and psychological symptoms (a= 0.83). The NRS-R has been shown to be a reliable assessment measure of BPSD.
Gitlin (2016) **TS**	Patient engagement	1.following treatment sessions, OTs completed 10 questions about patients 2.Direct observation to evaluate behavioural and affective reactions	First, following treatment sessions, OTs completed 10 questions (e.g., patient “appears bored or disinterested,” “shows signs of interest,” “expresses or shows feelings of pleasure,” “conveys that she/he wanted to stop”). Each item was scored from not at all (0) to extremely (4) for a possible range of 0–40. Higher scores indicated greater engagement. Second, direct observation was used to evaluate patient behavioural and affective reactions. Video recordings were conducted in a private room on the hospital unit designated for the study. To establish a baseline, each enrolled patient was observed within 3 days of admission and prior to OT prescribed activities, using a standard activity such as CNAs providing patients with a magazine for up to 15 min. Then, patients were observed in up to six TAP-H sessions involving the introduction or use of activities (Sessions 3–8) over 3 weeks or time-to-discharge. Video-recordings were subsequently coded and analysed across three behavioural domains (affect, verbal, and nonverbal behaviours) by up to five

			raters. Codes for each behavioural domain were identified from previously developed protocols. Affect codes were based upon the Apparent Affect Rating Scale and involved two positive emotional states (general alertness and pleasure) and three negative emotional states (anxiety, anger, and sadness). Verbal and nonverbal behaviour codes were also based upon protocols from previous observational studies with this population. Verbal behaviours were coded into 11 categories: positive, repetitive, nonsensical, aggressive, delusions, expressions of abandonment, somatic, sexual, self-harm, rejection of care or activity, and depressive. Nonverbal behaviours involved 12 categories: positive gestures, positive touch, motoric disturbance, facial disturbance, aggression, eyes closed, somatic, self-harm, sexual, rejection of care or activity, distracted, or depressive behaviours. Behaviours not observed were coded as “Can’t see.”
Weber (2009) **TS**	Patient engagement (therapeutic progress of group)	17-item tool (GES)	The Group Evaluation Scale (GES) was used to assess group progress in the elderly. It is rated on a 5-point scale that evaluates satisfaction, perception of group’s purpose, relevance, affective climate, cohesion, quality of interaction, personal involvement and adhesion to group treatment
	Behaviour (neuropsychiatric symptoms)	The Neuropsychiatric Inventory (NPI)	The NPI assesses 12 neuropsychiatric symptoms commonly observed in dementia: delusions, hallucinations, agitation, depression, anxiety, euphoria, apathy, disinhibition, irritability, aberrant motor behaviour, night-time behaviour disturbances, as well as appetite and eating disorders. The severity (1–3) and frequency (1–4) of each symptom are rated in an interview with the caregiver on the basis of scripted questions, generating a composite score for each of the 12 domains (product of frequency and severity subscores, range 0–12) as well as a total NPI score (sum of all composite scores for each domain 0–144). Content and concurrent validity, as well as inter-rater and test–retest reliability are established, and the NPI has proven sensitive to treatment efficacy.
Cheong (2016) **BA**	Patient engagement	Menorah Park Engagement Scale	The Menorah Park Engagement Scale was used to assess engagement. It captures four types of engagement: constructive engagement, passive engagement, self-engagement, and non-engagement. Constructive engagement referred to any motor or verbal behaviour exhibited in response to the activity the patient was participating in. Passive engagement pertained to listening and/or observing in response to the activity presented. Self-engagement was defined as any purposeless behaviour involving the patient’s engagement with himself or herself during the activity. Non-engagement related to staring into space or another direction away from the

			activity for >10 s or sleeping during an activity. The duration of each type of engagement was recorded in 5-min periods
	Emotional state/mood	Observed Emotion Rating Scale	The Lawton Observed Emotion Rating Scale was used to measure affect. The five types of emotion recorded included: pleasure, general alertness, anger, anxiety/fear, and sadness. Observers used standardized guidelines on facial and vocal expressions to code the different categories of emotion.
Daykin (2017) **BA**	Emotional state/mood	Arts Observational Scale (ArtsObs)	Observational data were collected using ArtsObs. This is a structured assessment tool that allows observers to record the impact of activities on participants' mood, distraction and relaxation. These impacts are scored on a scale of 1 (no benefit or negative impact) to 3 (very positive impacts for individuals and the ward atmosphere). The tool also allows recording of individual participants' happiness scores at the start and after the activity on a scale of 0 (negative, angry response) through to 7 (happy and excited)
	Behaviour (agitation)	Arts Observational Scale (ArtsObs)	observational data were collected using the Arts Observational Scale (ArtsObs). This is a structured assessment tool that allows observers to record the impact of activities on participants' mood, distraction and relaxation. These impacts are scored on a scale of 1 (no benefit or negative impact) to 3 (very positive impacts for individuals and the ward atmosphere). The tool also allows recording of individual participants' happiness scores at the start and after the activity on a scale of 0 (negative, angry response) through to 7 (happy and excited)
Windle (2018) **PC/BA**	Quality of life (self-report, proxy)	Dementia Quality of Life (DEMQOL)	The DEMQOL assesses five domains of QoL, including positive and negative emotions, memory, loneliness, and daily activities. It uses both self-reporting (29-items) and ratings by family carer or staff member as proxy (DEMQOL-Proxy; 32- items). Higher scores indicate better QoL. Respondents are asked to indicate how much they experienced each item during the last week ("a lot"=1; "quite a bit"=2; "a little"=3; "not at all"=4)
	Patient engagement (communication and social behaviour)	Communication Scale (HCS) (range 0-48);	Communication and social behaviour was assessed with the Holden Communication Scale completed by formal or informal carers. The scale consists of 12-items (range 0–48) covering a range of social behaviour and communication variables, including conversation, awareness, pleasure, humour, and responsiveness. Higher scores indicate *more difficulty* in communication
	Wellbeing	Greater Cincinnati Chapter Well-Being Observation Tool (GCCWBOT)	The Greater Cincinnati Chapter Well-Being Observation Tool (GCCWBOT) was developed specifically to observe the effects of a visual arts activity on the well-being of people with dementia. The original instrument addresses seven domains of well-being with different numbers of indicators (interest, sustained attention, pleasure, negative affect, sadness, self-esteem, and normalcy). In this

			study, we the tool is used with five further indicators included: "interest in own work" (interest domain), "negative comments" (self-esteem domain), with an additional domain to capture disengagement represented by "neutral passivity," "staring into space," and "sleeping behaviour"
Special care units			
Goldberg (2013) RCT	Quality of life (self-report, proxy)	Dementia Quality of Life (DEMQOL); EuroQol (EQ-5D)	The 28-item DEMQOL and 31-item DEMQOL-Proxy provide a method for evaluating health-related QoL for people with dementia. The measures are appropriate for use in mild/moderate dementia (MMSE ≥ 10) and DEMQOL-Proxy also shows promise in severe dementia. As DEMQOL and DEMQOL-Proxy give different but complementary perspectives on quality of life in dementia, the use of both measures together is recommended. EQ-5D is a standardized instrument developed by the EuroQol Group as a measure of health-related quality of life that can be used in a wide range of health conditions and treatments. The EQ-5D consists of a descriptive system and the EQ VAS. The descriptive system comprises five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The EQ VAS records the patient's self-rated health on a vertical visual analogue scale. This can be used as a quantitative measure of health outcome that reflects the patient's own judgement. The scores on these five dimensions can be presented as a health profile or can be converted to a single summary index number
	Mood/engagement; (positive mood/engagement, negative mood/engagement, active state, social interaction); Staff interactions meeting patient psychological and emotional needs; (personal enhancers, personal detractors)	Quantified mood, engagement and interaction scores measured by direct observation (proportion of 5-minute periods in which feature occurred)	Quantified mood and engagement scores, and staff interactions that met or disregarded patients' emotional and psychological needs ("personal enhancers" and "personal detractors") were recorded, according to strict definitions. Direct observation was based on Dementia Care Mapping tool. No additional details provided in paper or identified through other resources; Personal enhancers are actions that meet emotional or psychological needs, detractors are those that disregard them
	Behaviour (BPSD)	The Neuropsychiatric Inventory (NPI)	No description in paper. See above Weber-2009 for description of NPI.

Tay (2018) **CBA**	Quality of life (proxy?)	EuroQol (EQ-5D)	See description above for Goldberg-2013
	Well-being score; Ill-being score	Bradford Well-being and Ill-being profiling through observations on a set of unspecified indicators	Well-Being (WB) and Ill-Being (IB) profiling provides an index of the relative state of well-being and ill-being experienced by patients through independent observations on a set of indicators. A higher WB score indicates a better state of wellbeing, whereas a higher IB score indicates greater ill-being
	Behaviour (level of agitation)	Pittsburgh Agitation Scale (PAS)	The Pittsburgh Agitation Scale (PAS) [27] comprises four behaviour groups: (1) aberrant vocalization; (2) motor agitation; (3) aggressiveness; and (4) resistance to care. An intensity score, on a scale of 0 (not present) to 4 (highly present), is assigned to each behaviour group. The ratings across the four behaviour groups are summed to give the total score, with higher score indicating greater agitation
Skea (1996) **BA**	Quality of life	Quality of Interactions Schedule (QUIS)	Quality of life defined as quantity and quality of staff-resident interactions. Quality of Interactions Schedule (QUIS) was developed as a measure of quality of care in residential settings. QUIS is a non-participant time-sampling process where interactions are estimated from a series of 15-minute observations carried out over the working day and coded as: positive social (PC), positive care (PC), neutral (Ne), negative protective (NP) or negative restrictive (NR)
	Behaviour (‘aggressivity’ item of behaviour rating scale)	Adaptive Behaviour Rating Scale (ABRS)	Self-care, mobility, communication and other social functioning were measured using the Adaptive Behaviour Rating Scale (ABRS) (Woods and Britton, 1985; Ward et al., 1991), completed by the resident’s key worker. No total score reported, only the individual items
Volicer (1994) **PC**	Discomfort (monthly scores)	Discomfort Scale for Patients with advanced Dementia of Alzheimer’s type (DS-DAT; range 0-27)	Discomfort was defined as a negative emotional and/or physical state, subject to variation in magnitude in response to internal or environmental conditions. The scale has nine items that are made operational by 40 defining characteristics. Items are scored 0, 1, 2, or 3 depending on the number of defining characteristics present and their intensity and duration during a 5-minute observation period
Support for carers			
Catic (2013) **BA**	Comfort (based on symptom management)	Symptom Management at the End-of-Life in Dementia scale (SM-EOLD; range 0-45)	Answers to items of this scale indicated the condition of the care recipient during the last 90 days of his/her life. Higher scores indicate *greater* comfort (i.e. symptom control)