Validation of the Cohen-Mansfield Agitation Inventory Observational (CMAI-O) tool*

Alys W. Griffiths,[1] Christopher P. Albertyn,[2] Natasha L. Burnley,[1] Byron Creese,[3] Rebecca Walwyn,[4] Ivana Holloway,[4] Jana Safarikova,[5] and Claire A. Surr[1]

[1]Centre for Dementia Research, Leeds Beckett University, Leeds, UK
[2]Wolfson Centre for Aged Related Diseases, King's College London, London, UK
[3]University of Exeter Medical School, University of Exeter, Exeter, UK
[4]Leeds Institute for Clinical Research, University of Leeds, Leeds, UK
[5]Oxford Health NHS Foundation Trust, Oxford, UK

Abstract

Objectives: Behaviours associated with agitation are common in people living with dementia. The Cohen-Mansfield Agitation Inventory (CMAI) is a 29-item scale widely used to assess agitation completed by a proxy (family carer or staff member). However, proxy informants introduce possible reporting bias when blinding to the treatment arm is not possible, and potential accuracy issues due to irregular contact between the proxy and the person with dementia over the reporting period. An observational measure completed by a blinded researcher may address these issues, but no agitation measures with comparable items exist.

Design: Development and validation of an observational version of the CMAI (CMAI-O), to assess its validity as an alternative or complementary measure of agitation.

Setting: Fifty care homes in England.

Participants: Residents ($N = 726$) with dementia.

Measurements: Two observational measures (CMAI-O and PAS) were completed by an independent researcher. Measures of agitation, functional status, and neuropsychiatric symptoms were completed with staff proxies.

Results: The CMAI-O showed adequate internal consistency ($\alpha = .61$), criterion validity with the PAS ($r = .79$, $p = < .001$), incremental validity in predicting quality of life beyond the Functional Assessment Staging of Alzheimer's disease ($\beta = 1.83$, $p < .001$ at baseline) and discriminant validity from the Neuropsychiatric Inventory Apathy subscale ($r = .004$, $p = .902$).

Conclusions: The CMAI-O is a promising research tool for independently measuring agitation in people with dementia in care homes. Its use alongside the CMAI could provide a more robust understanding of agitation amongst residents with dementia.

Key words: dementia, agitation, observation, care homes, proxy bias

Behaviours associated with agitation are common amongst people living with dementia, particularly in care home settings with prevalence estimates of clinically significant symptoms ranging from 40 to 85% in various countries including the UK, Norway, and Holland (Livingston *et al.*, 2017; Testad *et al.*, 2007; Zuidema *et al.*, 2007). The aetiology of agitation is not clear but is likely to be multifactorial

with care approaches, the physical and social environment, medical comorbidities such as pain, genetics and the progression of dementia pathology all being contributing factors (Lanctôt *et al.*, 2017). Consistent with its multifactorial aetiology, there are many behaviours typically associated with agitation including repetitive mannerisms, hoarding, screaming, hitting, wandering, verbal aggression, and general restlessness, which can be extremely distressing to the person with dementia, their carers and others around them. Recent consensus criteria for agitation state that symptoms of agitation should: (1) occur in the context of cognitive impairment of dementia; (2) be consistent with emotional distress; (3) manifest as excessive motor activity, verbal or physical

aggression; and (4) not be solely attributable to another disorder (Cummings *et al.*, 2015). While these criteria provide a useful and much needed framework for the assessment of agitation, particularly with respect to inclusion in clinical trials, there is still a need to refine and validate assessment tools to accurately evaluate agitation as a clinical outcome.

The Neuropsychiatric Inventory (NPI; Cummings *et al.*, 1994) and the Cohen-Mansfield Agitation Inventory (CMAI; Cohen-Mansfield and Billig, 1986) are both commonly used to evaluate agitation. The CMAI consists of 29 items forming four subscales; physically aggressive behaviour (e.g. hitting others), physically non-aggressive behaviour (e.g. pacing), verbally aggressive (e.g. swearing) and verbally non-aggressive behaviours (e.g. repetitive sentences). The CMAI incorporates both the frequency and severity of behaviours associated with agitation and allows the quantification of agitated behaviours into a continuous measure, which is sensitive to change. The combination of these factors has led to its widespread use in clinical trials of pharmacological (e.g. Porsteinsson *et al.*, 2014) and psychosocial interventions (e.g. Ballard *et al.*, 2018).

Proxy informant interviews to complete outcome measures have clear benefits for populations of people with dementia, who are likely to have significant cognitive impairment or communication difficulties that often make direct interviews difficult (Moyle *et al.*, 2007). Notable drawbacks include informant recall/ knowledge (due to factors such as the amount of contact between the proxy and research participant during the reporting period) and — in the case of psychosocial interventions — lack of informant blinding to treatment, both of which can affect measurement accuracy and increase the chance of unwanted reporting error. In clinical trial settings, where there is an increasing number of clinical trials relating to agitation (e.g. Creese *et al.*, 2018), and where close attention is being paid to the accuracy of outcome measurements and minimising placebo response rates, these drawbacks illustrate the need to explore alternative and complementary methods for outcome measurements.

Observational tools are one avenue for addressing the issues outlined above, as independent observers (i.e. members of the research team who are not part of the intervention delivery) are able to remain blinded to treatment and are not subject to recall issues. There are a number of observational measures of agitation in people with dementia (Curyto *et al.*, 2008, Zeller and Rhoades, 2010). However, some measure a broader range of behaviours that include but are not restricted to agitation (Auer *et al.*, 1996; Beck *et al.*, 1997; McCann *et al.*, 1997; Morgan and Stewart, 1998; Van Haitsma *et al.*, 1997), some are not appropriately validated (Camberg *et al.*, 1999; Cohen-Mansfield

et al., 1989; Yudofsky *et al.*, 1997) or are not validated specifically for use with people with dementia. Others include only a narrow range of agitated behaviours, such as aggression (e.g. Almvik *et al.*, 2000; Perlman and Hirdes, 2008), and others are not available (Whall, 1999).

There are then few observational agitation measures appropriately validated for use with people with dementia, which include a range of potential agitated behaviours. These include the Agitated Behaviour Scale (ABS; Bogner *et al.*, 1999; Corrigan, 1989) and the Pittsburgh Agitation Scale (PAS; Rosen *et al.* 1994). However, these too have limitations. The ABS, for example, was originally developed for use in people with traumatic brain injury, and some items are relevant only to a hospital rather than care home environment (e.g. pulling at tubes, restraints etc). The scale also does not capture the frequency of occurrence of a particularly behaviour over an extended period of time due to the 20-minute observation periods used. The PAS measures the severity of four groups of agitated behaviours and is completed by trained researchers or clinical staff based on observations of participant behaviour over a shift or similar extended period of time. The PAS lacks detail in key areas including the breadth of symptoms assessed and the frequency with which symptoms occur. These additional pieces of information are vital for any assessment scale to capture heterogeneous range of symptoms that can constitute an agitated clinical syndrome. Likewise, due to the different approaches to measurement of agitation, none of the measures are appropriate for use as a direct comparator measure to the CMAI to detect potential reporting error by proxy reporters.

We sought to address these issues by developing and testing an observational version of the CMAI and including it in a randomized controlled trial where the primary outcome measure was resident agitation measured by the CMAI. This paper presents data that evaluates the psychometric properties and reliability of the CMAI-O and its criterion, incremental, and discriminant validity in comparison to existing measures. It also discusses its potential as an alternative or complementary measure of agitation in people with dementia. This was achieved through secondary analysis of the [name redacted] DCM EPIC trial dataset (see Surr *et al.*, 2016 for trial protocol).

Method

Participants

Participants [number (N) = 726] were recruited from 50 randomly selected care homes (mean = 15 residents per care home) in three regions across

Table 1. Overview of measures and use within the study

MEASURE	NUMBER OF ITEMS	COMPLETED BY	REPORTING PERIOD	FORM OF VALIDITY/ RELIABILITY USED TO TEST
CMAI	29	Researcher with family/staff proxy	Previous 2 weeks	Criterion
CMAI-O	29	Independent observer	Five hours on a single day	N/A
PAS	4	Independent observer	Five hours on a single day	Criterion
NPI-NH	12 (with sub scales)	Researcher with staff proxy	Previous 2 weeks	Criterion
QUALID	11	Researcher with family/staff proxy	Previous 7 days	Criterion
FAST	7 levels (sub-levels for levels 6 and 7)	Researcher with staff proxy	Not specified	Incremental

England (Yorkshire, Oxfordshire and London) for a randomized controlled trial. Residents were eligible to participate if they lived in the care home permanently and had a formal diagnosis of dementia or scored ≥ 4 on the Functional Assessment Staging of Alzheimer's disease (FAST; Reisberg, 1988). Residents were ineligible to participate if they had been formally admitted to an end of life care pathway or were cared for in bed. For inclusion in the present work, participants were required to have the CMAI-O (completed by an independent observer) completed for at least 2 hours, and at least one measure completed by a staff proxy; therefore, the sample differed at each time point due to loss to follow-up and missing data.as assessed by their key worker status and/or the judgement of the home manager) for at least several months. Additionally, this person must have been in regular contact with the resident during the previous 2 weeks. Researchers provided information on the completion of measures, including the period that the staff proxy should consider, and the use of rating scales.

Where possible, residents provided consent to participate. Where potential participants were deemed not to have capacity, a personal consultee (relative or friend) was approached to provide advice on the person's wishes. For potential participants whose personal consultee did not respond to contact or who did not have anyone to act as a personal consultee, a nominated consultee (care staff member) was approached to provide this advice.

Ethical considerations

Ethical approval for the clinical trial was obtained from the Bradford Leeds NHS REC committee and subsequent ethical approval for this sub-study was obtained from the Leeds Beckett University research ethics committee. Where possible, residents provided informed written consent to participate. Where potential participants were deemed not to have capacity, in line with the Mental Capacity Act

(2005) and following standard guidance (Medical Research Council, 2007), a personal consultee (relative or friend) was approached to provide advice on the person's wishes. For potential participants whose personal consultee did not respond to contact or who did not have anyone to act as a personal consultee, a nominated consultee (care staff member) was approached to provide this advice. For reasons of privacy and dignity, CMAI-O and PAS observations took place only in the public areas of the care home (e.g. lounge, dining room, corridors), and no private or personal care, or observations in bedrooms or bathrooms were conducted.

Measures

Six measures were completed at three time points; baseline, 6 months post care home randomisation (first follow-up) and 16 months post randomisation (second follow-up; see Table 1 for an overview). The independent observer, who was blinded to intervention allocation, conducted observations (of both the CMAI-O and PAS) between approximately 10am and 12pm, and 2pm and 5pm, on a single day, to reach a total of 5 hours observation. All other measures, completed by a researcher with the staff proxy, were completed within 4 weeks of the observational measures.

Researchers were able to act as the independent observer if they had not previously visited the care home to consent participants or complete any data collection. Prior to this, inter-rater reliability with another researcher of at least 80% agreement on the CMAI-O over a 1-hour period was required.

COHEN-MANSFIELD AGITATION INVENTORY
The CMAI (Cohen-Mansfield and Billig, 1986) measures 29 behaviours typically associated with agitation or aggression. Proxy reporters identify the frequency of 29 behaviours during the past 2 weeks, on a seven-point Likert scale ranging from "never" to "several times an hour", with higher scores indicating more agitation. The measure has moderate

concurrent validity with NPI-NH Agitation sub-scale ($r = .52$; Wood *et al.*, 2000) and converge with an observational scale of agitation, the Agitated Behaviours Mapping Instrument ($r = .32$-$.39$; Cohen-Mansfield and Libin, 2004). The CMAI showed internal consistency of $\alpha = .85$ at baseline.

OBSERVATIONAL COHEN-MANSFIELD AGITATION INVENTORY (CMAI-O)

The CMAI-O includes the same 29 items as the CMAI, but completed through observations of the participant by a trained researcher. The CMAI-O is scored on a four-point Likert scale of 1 = "never", 2 = "less than once per hour", 3 = "once per hour" and 4 = "several times an hour", with higher scores indicating more agitation. In the present study, data were collected across 5 hours split into two time points during a single day. Up to 15 observations (individuals) were conducted simultaneously. The CMAI-O showed internal consistency of $\alpha = .61$ at baseline.

PITTSBURGH AGITATION SCALE (PAS)

The PAS (Rosen *et al.*, 1994) is an observational rating of agitation and was completed by the independent observer concurrently with the CMAI-O. This measures intensity of agitation on four domains; aberrant vocalization (e.g. screaming), motor agitation (e.g. wandering), aggressiveness, and resisting care, each scored on a five-point Likert scale from 0 (not present) to 4 (high intensity of agitated behaviour). The measures also asks whether any interventions were used by care staff to support or manage agitated behaviours e.g. seclusion or restraint. PAS observations should be conducted for between 1 and 8 hours, and in the present study, data were collected across 5 hours in a single day. The measure has internal consistency of $\alpha = .80$ (Rosen *et al.*, 1994) and in the current study, had internal consistency of $\alpha = .63$ at baseline.

THE NEUROPSYCHIATRIC INVENTORY – NURSING HOME (NPI-NH)

The NPI-NH (Cummings, 1997; Cummings *et al.*, 1994) identified psychopathology for people with dementia in care homes. It assesses the presence and occupational disruptiveness of 12 behaviours, such as agitation/aggression, anxiety, and disinhibition. The frequency (within the past 2 weeks) and severity of each behaviour are rated on a four-point (1–4) and three-point (1–3) Likert scale, respectively. A score is then calculated for each behaviour by multiplying the severity and frequency score (1 – 12). A total score is then calculated by adding the scores of all 12 subscales. A composite score comprising agitation/aggression, irritability/lability, disinhibition and aberrant motor behaviour accounts for 60%

Table 2. Loss to follow up (of 726 registered residents at baseline)

	6 MONTHS	16 MONTHS
CMAI questionnaire (staff proxy completed)		
Received	587 (80.9%)	408 (56.2%)
Not received	139 (19.1%)	318 (43.8%)
Reasons not received		
Died	115 (15.8%)	264 (36.4%)
Moved out	21 (2.9%)	51 (7%)
Withdrew	1 (0.1%)	1 (0.1%)
Other	9 (1.2%)	3 (0.4%)

of the variance of a total CMAI score (Wood *et al.*, 2000). The measure showed internal consistency of $\alpha = .63$ at baseline.

QUALITY OF LIFE IN LATE-STAGE DEMENTIA

The Quality of Life in Late-Stage Dementia (QUALID) scale (Weiner *et al.*, 2000) measures 11 observable behaviours thought to indicate quality of life such as 'enjoys interacting or being with others'. Each behaviour is scored on a five-point Likert scale, lower scores indicate higher quality of life. The measure is based on observations made over the previous 7 days. Scores previously correlated with the NPI ($r = .40$, $p = .001$; Weiner *et al.*, 2000), although not the CMAI ($r = .19$, $p > .05$; Benhabib *et al.*, 2013). The NPI-NH showed internal consistency of $\alpha = .74$ at baseline.

FUNCTIONAL ASSESSMENT STAGING OF ALZHEIMER'S DISEASE

The Functional Assessment Staging of Alzheimer's Disease (FAST; Reisberg, 1988) records how severely dementia affects individuals' daily functioning. Scores range from 1 (no dementia) to 7 (severe dementia), with levels 6 and 7 having sub-levels, focusing on deficits associated with personal care and communication. It was completed by a researcher through interview with a staff proxy.

MISSING DATA

The levels of missing data were relatively high by the second follow up (43.8%), due to loss to follow up across 16-months of the trial (see Table 2). During data analysis, missing data were identified and recoded before being incorporated or excluded from all analyses. For the incremental validity and discriminant validity analyses all missing data recorded from each scale were excluded from the analysis. When performing correlation analyses between the PAS and CMAI-O, multiple imputation was used (Rubin, 1987). This technique creates complete data sets by generating several possible values for

Table 3. Completion rates for Cohen Mansfield Agitation Inventory-Observational and reasons for non-completion (of all 726 registered residents)

CMAI-O		BASELINE		6 MONTHS		16 MONTHS	
		AM	PM	AM	PM	AM	PM
Completed	**Yes**	451 (62.1%)	470 (64.7%)	369 (50.8%)	357 (49.2%)	231 (31.8%)	221 (30.4%)
	No	246 (33.9%)	226 (31.1%)	221 (30.4%)	233 (32.1%)	183 (25.2%)	193 (26.6%)
	Missing (includes individuals who died, withdrew and left care home)	29 (4.0%)	30 (4.1%)	136 (18.7%)	136 (18.7%)	312 (43.0%)	312 (43.0%)
If not completed, reasons (% out of not completed):							
Resident not in communal area		240 (97.6%)	221 (97.8%)	210 (95.0%)	217 (93.1%)	180 (98.4%)	187 (96.9%)
Resident did not wish to be observed		1 (0.4%)	–	–	–	–	–
In hospital		2 (0.8%)	2 (0.9%)	7 (3.2%)	7 (3.0%)	3 (1.6%)	3 (1.6%)
Missing		2 (0.8%)	1 (0.4%)	2 (0.9%)	6 (2.6%)	–	3 (1.6%)
Other		1 (0.4%)	2 (0.9%)	2 (0.9%)	3 (1.3%)	–	–

any missing values. Analyses are conducted across all of these data sets and outputs provide estimates for each data set about the results that would have been expected if there had been no missing values in the original data set. In the present study, five data imputations were created.

Furthermore, the CMAI-O was only completed for participants who were in a communal area during data collection (either AM, PM, or both). The number of participants with completed CMAI-O at each time point and reasons why people were not in a communal area of the care home were recorded (see Table 3).

Data Analysis

Data manipulation, analyses and graphical visualisations were performed in IBM SPSS 24. Prior to performing inferential statistical analyses, each dataset was assessed for parametric assumptions. A skewed distribution was consistently observed across all data, suggesting that scores were predominantly weighted at the lower end of each scale. This was expected as the majority of residents within care homes score lowly on neuropsychiatric outcome measures; whilst only a minority of residents typically score highly. In order to account for this non-normal data, parametric bootstrapping was performed on all analyses (see Davison and Hinkley, 1997). Data were deemed to meet all other parametric assumptions.

Several forms of validity were tested; criterion validity, incremental validity and discriminant validity. Criterion validity refers to the degree of correlation between a proposed novel measure and a pre-existing, validated assessment that targets the same phenomenon. It is important that a new measure demonstrates criterion validity as it assures researchers and practitioners that any results derived can be assumed to be consistent with alternative methods of assessment. Incremental validity is used to determine whether a new measure increases the predictive ability beyond that provided by an existing tool. It is important to demonstrate this form of validity as it demonstrates that a measure elicits additional information relative to other measures. Discriminant validity illustrates whether scores on a scale are truly independent of scores on another scale that measures distinctly different phenomena. Previous evidence has found no correlation between agitation and apathy in a sample of nursing home residents living with dementia (Mouriz-Corbelle et al., 2017). Therefore, we hypothesised that the CMAI-O would be unrelated to the NPI Apathy subscale if discriminant validity is high.

Results

Sample

The baseline sample consisted of 726 participants, with an average age of 85 (range: 57–102). The majority of participants were female (536; 73.8%) and identified as White British (702; 96.7%). Participant dementia severity, as rated by the FAST, spanned mild cognitive impairment (6; 1%), mild dementia (95; 13.6%), moderate dementia (74; 10.6%), moderate/severe dementia (380 ;54.5%) and severe dementia (142; 20.4%), with data missing for 29 participants (3.9%). The average score on the CMAI was 47.35 points (standard deviation =18.40) and average score on the NPI was 14.18 (SD = 14.69). Of these participants, 451 (62%) had a completed CMAI-O for AM, and 470 (65%) for PM.

At 6 months, the sample consisted of 587 participants. The average score on the CMAI was 45.35 (standard deviation = 16.32). Of these participants, 369 (62%) had a completed CMAI-O for AM and 357 (61%) for PM. At 16 months, the sample consisted of 408 participants. The average score on the CMAI was 44.38 (standard deviation = 16.72). Of these participants, 231 (56%) had a completed CMAI-O for AM and 221 (54%) for PM.

Preliminary analyses

The prevalence of each symptom at baseline, measured using the CMAI and CMAI-O can be seen in Table 4. Within this table, for the CMAI, only items that were present 'several times a week' or more were included, in order to be comparable with the CMAI-O. This is due to the CMAI-O being completed during one single day; therefore, items that are displayed once per week or less were unlikely to be displayed during a single day observation. Generally, levels of agitation were lower for the CMAI-O, with the exception of 'performing repetitious mannerisms', which was seen more commonly.

Correlations between the CMAI-O and the CMAI, PAS and NPI, at each time point were also conducted (see Table 5). Pearson correlations for the CMAI-O and the CMAI as scored by the staff proxy were significant: baseline $r = .44$, 6 months $r = .23$ and 16 months $r = .28$ (all $p < .001$). The NPI Agitation subscale had a weak but statistically significant correlation with the CMAI-O at baseline and 6 months ($r = .38$, $p = < .001$ and $r = .12$, $p = .021$), however no significant correlation was found between the two measures at 16 months ($r = .09$, $p = .199$). Finally, both the AM and the PM scores for the PAS showed a significant correlation with total CMAI-O scores across all time points, with a significance of $p = < .001$.

To understand whether scores on the CMAI-O were related to scores on another scale measuring

Table 4. Proportion of residents experiencing symptoms associated with agitation at baseline (% of those with both Cohen Mansfield Agitation Inventory-Observational and Cohen Mansfield Agitation Inventory completed)

SYMPTOM	CMAI-O	CMAI
Pacing, aimless wandering	15	32
Inappropriate dress or disrobing	2	11
Spitting (including at meals)	1	7
Cursing/verbal aggression	8	14
Constant unwarranted request for attention/help	4	15
Repetitive sentences/questions	9	27
Hitting (including self)	1	8
Kicking	1	6
Grabbing onto people	2	10
Pushing	2	6
Throwing	1	3
Strange noises (weird laughter or crying)	11	13
Screaming	1	10
Biting	1	2
Scratching	1	6
Trying to get to a different place	3	10
Intentional falling	1	1
Complaining	10	14
Negativism	4	11
Eating/drinking inappropriate substances	1	2
Hurting self or others	1	1
Handling things inappropriately	4	4
Hiding things	0	5
Hoarding things	1	4
Tearing things/destroying property	1	2
Performing repetitious mannerisms	22	15
Making verbal sexual advances	1	2
Making physical sexual advances	1	1
General restlessness	16	24

Table 5. Correlation between measures completed at each time point

	MAI-O		
	BASELINE	6 MONTHS	16 MONTHS
CMAI	.436[*]	.226[*]	.276[*]
NPI (Agitation)	.382[*]	.121[*]	.085
PAS (AM)	.583[*]	.825[*]	.770[*]
PAS (PM)	.638[*]	.479[*]	.540[*]

[*] **significant at p < 0.01** **CMAI-O – Cohen-Mansfield Agitation Inventory – Observational; CMAI – Cohen Mansfield Agitation Inventory; NPI – Neuropsychiatric Inventory; PAS – Pittsburgh Agitation Scale

agitation, correlations were calculated between this and the PAS. There is no existing evidence of validation of the CMAI and the PAS. Given the PAS is the only existing observational measure of agitation it was used to compare to the CMAI-O,

to establish criterion validity. A partial correlation was completed with bootstrapping to account for non-normally distributed data, controlling for the data collection time point (baseline/6-month follow up/16-month follow up) and the time of day (AM or PM) when both were completed. When controlling for time point and the time of day completed, a significant partial correlation was found ($r = .80$, $p = <.001$). Additionally, we compared CMAI-O scores to the NPI Agitation subscale. Whilst not an observational measure, it represents a psychometric tool validated to measure agitation in nursing home residents, thus a significant correlation between this and the CMAI-O would support criterion validity. A partial correlation analysis of the CMAI-O and the NPI Agitation subscale, controlling for data collection time point, found a small but significant relationship between the measures ($r = .24$, $p = <.001$).

To assess whether the CMAI-O increased the predictive ability beyond that provided by an existing method of assessment, we hypothesised that agitation as measured by the CMAI-O would predict quality of life as measured by the QUALID, above the prediction from scores on the FAST. It has been suggested in previous studies that functional status significantly impacts the quality of life of a person living with dementia (e.g. Andersen *et al.*, 2004) with those with more severe dementia experiencing poorer quality of life.

In step one of a hierarchal multiple regression, scores on the FAST were included as a predictor of quality of life. The FAST score significantly predicted the quality of life score across all time points, baseline ($\beta = .21$, $p < .001$), 6 months ($\beta = .202$, $p = .001$) and 16 months ($\beta = .198$, $p = .006$). In step two of the analysis, when the CMAI-O score was also included as a predictor, an R^2 change of .095 ($R^2 = .139$, $F(2, 390) = 66.87$, $p < .001$) was observed at baseline, an R^2 change of .027 ($R^2 = .068$, $F(2, 286) = 33.7$, $p < .001$) at 6-months and an R^2 change of .042 ($R^2 = .082$, $F(2, 191) = 32.12$, $p < .001$) at 16 months. The FAST remained significant at each of these time points also [6 months ($\beta = 3.45$, $p = .001$), 16 months ($\beta = 3.09$, $p = .002$)]. These analyses demonstrated that the observational CMAI had incremental predictive value in the measurement of quality of life beyond levels that could be predicted by participants' stage of dementia, as determined by functional status (FAST).

To establish whether the CMAI-O failed to correlate with a measure that should be conceptually unrelated, bootstrapped correlation analyses between the CMAI-O and all NPI subscales were conducted, controlling for the data collection time point. We hypothesized that the CMAI-O would

show no relationship with the NPI Apathy subscale. No relationship was found between the CMAI-O and NPI Apathy subscales, when controlling for time point ($r = .004$, $p = .902$). This suggests that agitation as measured by the CMAI-O is completely unrelated to apathy, in line with previous research.

Discussion

The present study describes the development and validation of the CMAI-O. This tool assesses agitation over a shorter period than the original CMAI, can be completed by an independent observer rather than relying on proxy reports and therefore could form part of an agitation assessment battery for research, including clinical trials. The psychometric properties of the CMAI-O were examined. Scores on the CMAI-O correlated with scores on the CMAI, PAS and NPI-NH, and the tool had incremental predictive value when measuring quality of life beyond levels predicted by functional status. No relationship was found between the CMAI-O and NPI Apathy subscale, suggesting that agitation as measured by the CMAI-O is completely unrelated to apathy. The internal consistency of the CMAI-O, whilst being adequate, suggested that there may be issues with the items within the scale. Further examination of this is required, to establish whether this was caused by general low levels of agitation within our sample, or issues within the measure. Within the current study, we sought to assess the psychometric properties of an alternative method of administration, therefore we have not suggested refinement of the CMAI-O items.

To identify whether the CMAI-O identified behaviours typically associated with agitation comparisons between the CMAI-O and CMAI were also conducted. Scores on the CMAI-O were consistently slightly lower than those on the CMAI. This may be due to the restricted locations and time periods of observation. For example, observations took place in the communal areas therefore agitated behaviours in other areas were not recorded, in particular during personal care. Previously, personal care activities including bathing and toileting, particularly when initiated by a caregiver rather than the resident themselves, have been found to significantly increase agitation behaviours (Cohen-Mansfield *et al.*, 1992).

There are several limitations with the present research. One difficulty when using observational methods for assessing agitation is that many behaviours do not occur often, therefore long observations are sometimes required to detect behaviours. This is especially true for the less common behaviours like aggression, which are some of the most important in terms of impact on others (Cohen-Mansfield, 1996). A standardised period of 5 hours

within a single day in this study provided data that was adequately correlated with alternative proxy measures and, therefore, is the recommended minimum observation period for use of the CMAI-O. Secondly, we recruited participants from 50 care homes in three areas of England, which may not represent average care homes across the UK. For example, those under admissions bans due to breaches of statutory regulations were not eligible to participate and those approached but not interested in taking part in research did not consent to take part. The levels of agitation seen in the present sample were lower than previous samples (e.g. Zuidema *et al.*, 2007). This may reflect improved understanding of how to effectively manage behaviours associated with agitation by care home staff over recent years. Understanding these as a reflection of unmet needs rather than symptoms of dementia is becoming a more common approach and recent research has found lower levels of agitation amongst people living with dementia in care homes (Livingston *et al.*, 2017). Furthermore, the average CMAI score in the present study was similar to that of participants in another recent psychosocial clinical trial (Ballard *et al.*, 2016). However, comparing levels of agitation in this study to those reported in earlier care home studies identified lower levels in this sample generally. For example, 16% of individuals experienced general restlessness, in comparison to 44% of individuals in the Netherlands (Zuidema *et al.*, 2007), although our sample was comparable to the 22% of individuals who experienced general restlessness within a Norwegian sample (Testad *et al.*, 2007). However, it may be that given the relationship between agitation and poor quality care, individuals living in care homes where there are breaches of statutory regulations, such as unsafe staffing levels, might experience greater levels of agitation, and such individuals were not recruited in the present study as their care homes were ineligible to participate. Additionally, we only observed participants in communal areas, which may have impacted on levels of agitation. This may be particularly relevant as some of the measures completed by staff proxy informants included specific questions about personal care, which are not observed due to issues around privacy and dignity.

By developing the CMAI-O we have addressed the disadvantage of the CMAI relating to potential reporting biases from proxy reporters. However, observational measures of agitation also have limitations. Levels of agitation observed by the CMAI-O were generally lower than those reported by care staff members in the CMAI, although scores on the two were significantly correlated. It may be that staff proxies over-report the prevalence of these behaviours on the CMAI, particularly if they find the behaviours difficult to support. Alternatively the CMAI-O may miss some agitation due to its use in this study only in communal areas during the daytime. Research indicates higher levels of agitation may be seen in the evenings and during personal care (e.g. Sloane *et al.*, 2004). However, the CMAI-O does have the benefit of using continual observation over two sustained periods, which may offer a more accurate picture of agitation than measures that use short time-sampling observations. Additionally, observer fatigue is a concern with longer observation periods and one which should be considered. Therefore, further research on optimal observation length and time(s) of day when using the CMAI-O is required, as well as exploration of ethical and practical issues around the potential for observations to be conducted in other areas where care may be delivered. Researchers have rightly questioned the accuracy of recall of agitated behaviours by staff members working in care homes, due to the impact of caregiver burden. Care staff reporting a lower burden have been found to provide higher quality of life ratings than those experiencing more caregiver burden (Graske *et al.*, 2014). One possible explanation for this is that care staff experiencing more burden also expect residents to be experiencing this burden (Graske *et al.*, 2014). Alternatively, it may be that those experiencing more burden are less able to see positive aspects of residents' lives. However, there is currently very little research examining this amongst formal care staff. A systematic review of the relationship between family caregiver distress and quality of life ratings concluded that family caregivers who reported experiencing more stress associated with caregiving provided more negative reports of their relative's quality of life (Neumann *et al.*, 2000). However, more recently no relationship has been found between family caregiver distress and their proxy quality of life ratings (Sheehan *et al.*, 2012). Furthermore, there is scope for influence by interpretation or perceptual biases of individual proxies. For example, a resident whose physical mobility has improved following an exercise intervention and who is spending more time walking may be noted as 'wandering' (Brett *et al.*, 2017) by care home staff. This may be due to a perception that residents who are not sitting are more problematic to care for, or that 'wandering' is a common problem in dementia and therefore all walking behaviours must be wandering. Perceiving or interpreting this behaviour as wandering would indicate an increase in agitation when completing the CMAI, when the behaviour is instead walking that reflects an increase in physical ability and activity. Use of the CMAI-O, by trained researchers who receive instruction on

how to observe and interpret resident behaviours may allow a more objective understanding of the behaviour of individuals, through concentrated observation of them over a substantial period of time. This may provide a more accurate measure of agitation than untrained staff recall of behaviours over the previous 2 weeks. Future research should consider the influence of proxy reporters and their subjective experiences on research outcomes.

Furthermore, a single researcher can complete the CMAI-O for multiple participants over several time points, whereas it is particularly difficult to recruit staff proxies for multiple time points, based on turnover and shift patterns. In addition the use of staff proxies to report on agitation of a 2-week period raises concerns about accuracy reporting, since they work shifts and even when on duty do not always spend long periods of time with individual residents. Therefore, assumptions about frequency of agitated behaviours may be made on based on intensity over a short period or reports of disruptiveness from colleagues. In one recent study, to overcome this, staff members completed a tally score of the CMAI at the end of each shift for a 2-week period (Brett *et al.*, 2017). The CMAI-O addresses this potential reporting bias, as a researcher captures all behaviours across a single day, leading to decreased likelihood of over or under-reporting of agitated behaviours. However, observation on a single day does limit ability to observe less common but potentially impactful behaviours and therefore, further research is needed to examine the test-retest reliability of the CMAI-O and establish the optimal periods over which to use the measure to most accurately capture the full extent and range of agitated behaviours. Future research should also extend the validation of the CMAI-O to confirm the factor structure of the measure amongst individuals experiencing higher levels of agitation and establish whether it is suitable for use in other settings, such as hospital wards, or with other populations such as individuals living with severe and enduring mental health problems, where agitation is a concern.

The CMAI-O offers a novel approach for evaluating the presence of behaviours associated with agitation amongst people with dementia in care homes (both residential and nursing) within research and practice. The CMAI-O may be particularly useful to include in research where the CMAI is the primary outcome, since it may provide a measure of potential performance or reporting bias from staff members who cannot be blinded to treatment arm. Moreover, as an observational measure of agitation, the CMAI-O offers a more suitable observational measure than the PAS or ABS because its items are directly aligned to the CMAI and thus the results are complementary to those obtained on the CMAI. As such, future researchers will be able to confidently provide composite CMAI scores using information from both proxy-informant and researcher observations.

Conclusion

In conclusion, this initial validation of the CMAI-O has demonstrated that the tool is appropriate for use with people living with dementia in care homes. The tool provides a complementary or alternative to measure to staff proxy completed measures of agitation. This may be particularly valuable when evaluating care home interventions. It is hoped that the development of the CMAI-O leads to increased use of independent assessment to understand behaviours associated with agitation amongst people living with dementia in care homes and its use as an outcome measure to evaluate change in agitation.

Conflicts of interest

The authors declare no conflicts of interest.

Description of authors' roles

AWG: design of the study, data acquisition, data analysis, interpretation of data and drafting of this paper.

CPA: data acquisition, data analysis, interpretation of data and drafting of this paper.

NLB: data acquisition, data analysis, interpretation of data and drafting of this paper.

BC: data acquisition, data analysis, interpretation of data and drafting of this paper.

RW: data analysis, interpretation of data and drafting of this paper.

IH: data analysis, interpretation of data and drafting of this paper.

JS: data acquisition and drafting of this paper.

CAS: conception and design of the study, data acquisition, interpretation of data and drafting of this paper.

Funding

Acknowledgements

References

Almvik, R., Woods, P. and Rasmussen, K. (2000). The Brøset violence checklist: sensitivity, specificity, and interrater reliability. *Journal of Interpersonal Violence*, 15, 1284–1296.

Andersen, C. K., Wittrup-Jensen, K. U., Lolk, A., Andersen, K. and Kragh-Sorensen, P. (2004). Ability to perform activities of daily living is the main factor affecting quality of life in patients with dementia. *Health and Quality of Life Outcomes*, 2, 52.

Auer, S. R., Monteiro, I. M. and Reisberg, B. (1996), The empirical behavioral pathology in Alzheimer's disease (E-BEHAVE-AD) rating scale, *International Psychogeriatrics*, 8, 247–266.

Ballard, C. *et al.* (2018) Impact of person-centred care training and person-centred activities on quality of life, agitation, and antipsychotic use in people with dementia living in nursing homes: a cluster-randomised controlled trial. *Plos Med*, 15, e1002500

Ballard, C. *et al.* (2016). Impact of antipsychotic review and nonpharmacological intervention on antipsychotic use, neuropsychiatric symptoms, and mortality in people with dementia living in nursing homes: a factorial cluster-randomized controlled trial by the well-being and health for people with dementia (WHELD) program. *American Journal of Psychiatry*, 173, 252–262.

Beck, C., Heithoff, K., Baldwin, B., Cuffel, B., O'Sullivan, P. and Chumbler, N. R. (1997). Assessing disruptive behavior in older adults: the Disruptive Behavior Scale. *Aging & Mental Health*, 1, 71–80.

Benhabib, H., Lanctot, K. L., Eryavec, G. M., Li, A. and Herrmann, N. (2013). Responsiveness of the QUALID to improved neuropsychiatric symptoms in patients with Alzheimer's disease. *Canadian Geriatrics Journal*, 16, 180–185. doi: 10.5770/cgj.16.78.

Bogner, J. A., Corrigan, J. D., Stange, M. and Rabold, D. (1999). Reliability of the agitated behavior scale. *The Journal of Head Trauma Rehabilitation*, 14, 91–96.

Brett, L., Traynor, V., Meedya, S. and Stapley, P. (2017). Impressions of using the Cohen- Mansfield Agitation Inventory as an outcome measure: lessons learnt for future clinical researchers. *Dementia*. Advanced online publication.

Camberg, L. *et al.* (1999). Evaluation of simulated presence: a personalized approach to enhance well- being in persons with Alzheimer's disease. *Journal of the American Geriatrics Society*, 47, 446–452.

Cohen-Mansfield, J. (1996). New ways to approach manifestations of Alzheimer's disease and reduce burdens. *International Psychogeriatrics*, 8, 91–94.

Cohen-Mansfield, J. and Billig, N. (1986) Agitated behaviours in the elderly. I. A conceptual review. *Journal of the American Geriatrics Society*, 34, 711–721.

Cohen-Mansfield, J. and Libin, A. (2004). Assessment of agitation in elderly patients with dementia: correlations between informant rating and direct observation. *International Journal of Geriatric Psychiatry*, 19, 881–891.

Cohen-Mansfield, J., Marx, M. S. and Werner, P. (1992). Observational data on time use and behavior problems in the nursing home. *Journal of Applied Gerontology*, 11, 111–121.

Cohen-Mansfield, J., Werner, P. and Marx, M. S. (1989). An observational study of agitation in agitated nursing home residents. *International Psychogeriatrics*, 1, 153–165.

Corrigan, J. D. (1989) Development of a scale for assessment of agitation following traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, 11, 261–277.

Creese, B., Vasconcelos Da Silva, M., Johar, I. and Ballard, C. (2018). The modern role of antipsychotics for the treatment of agitation and psychosis in Alzheimer's disease. *Expert Review of Neurotherapeutics*, 18, 461–467.

Cummings, J. L. (1997). The neuropsychiatric inventory: assessing psychopathology in dementia patients. *Neurology*, 48, S10–S16.

Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D. A. and Gombein, J. (1994). The Neuropsychiatric Inventory. Comprehensive assessment of psychopathology in dementia. *Neurology*, 44, 2308.

Cummings, J. *et al.* (2015). Agitation in cognitive disorders: International Psychogeriatric Association provisional consensus clinical and research definition. *International Psychogeriatrics*, 27, 7–17.

Curyto, K. J., Van Haitsma, K. and Vriesman, D. K. (2008). Direct observation of behavior: A review of current measures for use with older adults with dementia. *Research in Gerontological Nursing*, 1, 52–76.

Davison, A. C. and Hinkley, D. V. (1997). Bootstrap methods and their application (Vol. 1). Cambridge, UK: Cambridge University Press.

Graske, J., Meyer, S. and Wolf-Ostermann, K. (2014). Quality of life ratings in dementia care – a cross-sectional study to identify factors associated with proxy-ratings. *Health and Quality of Life Outcomes*, 12, 177.

Lanctôt, K. L. *et al.* (2017). Neuropsychiatric signs and symptoms of Alzheimer's disease: New treatment paradigms. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 3, 440–449.

Livingston, G. *et al.* (2017). Prevalence of and associations with agitation in residents with dementia living in care homes: MARQUE cross-sectional study. *British Journal of Psychiatry Open*, 3, 171–178.

McCann, J. J., Gilley, D. W., Hebert, L. E., Beckett, L. A. and Evans, D. A. (1997). Concordance between direct observation and staff rating of behavior in nursing home residents with Alzheimer's disease. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 52, 63–72.

Medical Research Council (2007). Medical research involving adults who cannot consent. https://mrc.ukri.org/ documents/pdf/medical-research-involving-adults-who-cannot-consent/

Mental Capacity Act (2005). London: HMSO.

Morgan, D. G. and Stewart, N. J. (1998). High versus low density special care units: impact on the behaviour of elderly residents with dementia. *Canadian Journal on Aging*, 17, 143–216.

Mouriz-Corbelle, R., Facal, D., Caamano-Ponte, X. and Dosil-Diaz, C. (2017). Apathy and agitation in a sample of dementia patients living in a nursing home. *Alzheimer's & Dementia*, P1325, P4–167.

Moyle, W., McAllister, M., Venturato, L. and Adams, T. (2007) Quality of life and dementia: the voice of the person with dementia. *Dementia*, 6(2), 175–191.

Neumann, P. J., Araki, S. S. and Gutterman, E. M. (2000). The use of proxy respondents in studies of older adults: Lessons, challenges, and opportunities. *Journal of the American Geriatrics Society*, 48, 1646–1654.

Perlman, C. M. and Hirdes, J. P. (2008). The aggressive behavior scale: a new scale to measure aggression based on the minimum data set. *Journal of the American Geriatrics Society*, 56, 2298–2303.

Porsteinsson, A. P. *et al*. (2014). Effect of Citalopram on Agitation in Alzheimer Disease: the CitAD randomized clinical trial. *JAMA*, 311, 682–691.

Reisberg, B. (1988). Functional assessment staging (FAST). *Psychopharmacology Bulletin*, 24, 653–659.

Rosen, J. *et al*. (1994). The Pittsburgh Agitation Scale: a user-friendly instrument for rating agitation in dementia patients. *American Journal of Geriatric Psychiatry*, 2, 52–59.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York, NY: Wiley.

Sheehan, B. D. *et al*. (2012). Patient and proxy measurement of quality of life among general hospital in-patients with dementia. *Aging & Mental Health*, 16(5), 603–607.

Sloane, P. D. *et al*. (2004). Effect of person-centered showering and the towel bath on bathing-associated aggression, agitation, and discomfort in nursing home residents with dementia: a randomized, controlled trial. *Journal of the American Geriatrics Society*, 52, 1795–1804.

Surr, C. A. *et al*. (2016). Evaluating the effectiveness and cost effectiveness of Dementia Care Mapping™ to enable person-centred care for people with dementia and their carers (DCM-EPIC) in care homes: study protocol for a randomised controlled trial. *Trials*, 17, 300.

Testad, I., Aasland, A.M. and Aarsland, D. (2007). Prevalence and correlates of disruptive behavior in patients in Norwegian nursing homes. *International Journal of Geriatric Psychiatry*, 22, 916–921.

Van Haitsma, K., Lawton, M. P., Kleban, M. H., Klapper, J. and Corn, J. (1997). Methodological aspects of the study of streams of behavior in elders with dementing illness. *Alzheimer Disease and Associated Disorders*, 11, 228–238.

Weiner, M. F., Martin-Cook, K., Svetlik, D. A., Saine, K., Foster, B. and Fontaine, C. S. (2000). The Quality of Life in Late-Stage Dementia (QUALID) scale. *Journal of the American Medical Directors Association*, 1, 114–116.

Whall, A. L. (1999). The measurement of need-driven dementia-compromised behavior: Achieving higher levels of inter rater reliability. *Journal of Gerontological Nursing*, 25, 33–37.

Wood, S., Cummings, J. L., Hsu, M-A., Barclay, T. and Wheatley, M.V. (2000). The use of the Neuropsychiatric Inventory in nursing home residents. *American Journal of Geriatric Psychiatry*, 8, 75–83.

Yudofsky, S. C., Kopecky, H. J., Kunik, M., Silver, J. M. and Endicott, J. (1997). The Overt Agitation Severity Scale for the objective rating of agitation. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 9, 541–548.

Zeller, S. L. and Rhoades, R. W. (2010). Systematic reviews of assessment measures and pharmacologic treatments for agitation. *Clinical Therapeutics*, 32, 403–425.

Zuidema, S. U., Derksen, E., Verhey, F. R. J. and Koopmans, R. T. C. M. (2007). Prevalence of neuropsychiatric symptoms in a large sample of Dutch nursing home patients with dementia. *International Journal of Geriatric Psychiatry*, 22, 632–638.