Assessment of Behavioural and Psychological Symptoms Associated with Dementia

David Conn, Lilian Thorpe

ABSTRACT: Neuropsychiatric symptoms (mood, psychotic, and behavioural) are very common in dementia and do not necessarily correlate well with other measures of cognition. However, these symptoms are of great importance, as they are a major source of excess disability, patient distress and caregiver burden and have great impact on the level of care required, and the associated costs. This paper is a review of the most useful outcome measures for behaviour and mood symptoms. Investigators who require a comprehensive instrument to measure neuropsychiatric symptoms in studies of patients with dementia should consider using the Neuropsychiatric Inventory (NPI), the Behavior Rating Scale for Dementia of the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD-BRSD) or, possibly, the Behavioral Pathology in Alzheimer’s Disease Scale (BEHAVE-AD). The Cornell Scale for Depression in Dementia and the Dementia Mood Assessment Scale (DMAS) are recommended for evaluating depressive symptoms and the Cohen-Mansfield Agitation Inventory (CMAI) is very useful for evaluating the full range of agitation symptoms.

RÉSUMÉ: Évaluation des symptômes comportementaux et psychologiques associés à la démence. Les symptômes neuropsychiatriques (dysthymiques, psychotiques et comportementaux) sont très fréquents dans la démence et leur corrélation aux autres mesures de la cognition n’est pas toujours bonne. Cependant, ces symptômes sont très importants parce qu’ils sont une source majeure d’invalidité supplémentaire et de détresse pour le patient, un fardeau pour les soignants et parce qu’ils ont un impact considérable sur le niveau de soins requis et les coûts qui y sont associés. Cet article revoit les mesures d’impact les plus utiles pour l’évaluation des symptômes comportementaux et thymiques. Les chercheurs qui ont besoin d’instruments détaillés pour évaluer les symptômes neuropsychiatriques en cours des études portant sur des patients déments devraient faire appel au Neuropsychiatric Inventory (NPI), au Behavior Rating Scale for Dementia of the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD-BRSD), ou encore au Behavioral Pathology in Alzheimer’s Disease Scale (BEHAVE-AD). Le Cornell Scale for Depression in Dementia et le Dementia Mood Assessment Scale (DMAS) sont recommandés pour l’évaluation des symptômes dépressifs et le Cohen-Mansfield Agitation Inventory (CMAI) est très utile pour l’évaluation de toute la gamme des symptômes d’agitation.

Neuropsychiatric symptoms (mood, psychotic, and behavioural) are very common in dementia and do not necessarily correlate well with other measures of cognitive functioning. However, these symptoms are of great importance, as they are a major source of excess disability, patient distress, caregiver burden and contribute greatly to the level of care required, and associated costs. It is therefore important that the assessment of neuropsychiatric symptoms is included as a core component in trials of antidementia therapies. Numerous instruments are available for assessing neuropsychiatric symptoms in patients suffering from dementia. It is likely that the lack of a consistent approach to the measurement of these problems contributes to the variability of rates reported in observational studies.

Antidementia therapies may not have a uniform impact on all areas, and it is conceivable that an improvement in one domain is associated with worsening in another. For example, there is evidence from some trials that cholinesterase inhibitors may prevent the emergence of neuropsychiatric symptoms in patients with dementia and may also have direct benefits for certain behaviours, such as apathy and psychotic symptoms. On the other hand, some independently designed studies have shown no significant improvement in various neuropsychiatric symptoms.
In addition, worsening in some domains, such as sleep, may occur. It is important to note that trials of antidementia drugs have almost exclusively measured behaviour as a secondary outcome measure, with post-hoc analyses of the behavioural domains. Aggregate measures, such as global outcome scales, may be less useful than measures addressing specific clinical domains. This is also supported by population-based phenomenological surveys, such as the Cache County Study, that have explored different symptom clusters. Authors reporting on this large population study of adults over 65 (n=5,092) suggested that symptoms clustered into three main syndromes: an affective syndrome, a psychotic syndrome, and other neuropsychiatric disturbances. Further work has also gone into defining diagnostic criteria for such syndromes, such as that by Olin et al who described diagnostic criteria for depression of Alzheimer’s disease, and by Schneider & Dagerman who described clustering of symptoms to suggest the validity of a syndrome of psychosis of Alzheimer’s disease.

Thus, instruments that assess neuropsychiatric symptoms in dementia include those that are designed to identify a variety of common neuropsychiatric symptoms not necessarily part of another comorbid diagnostic category, as well as those that are designed to identify specific syndromes or comorbid diagnoses, such as depression or psychosis. Although it is tempting to include instruments detailed enough to make valid diagnoses of these comorbid disorders, this increases the complexity and length of the dementia trial protocol, so this is not usually done unless the trial is geared particularly to this problem.

There are a number of important issues related to the interpretation and evaluation of neuropsychiatric symptoms. One of the most significant issues relates to the source of the information. Patients with significant cognitive impairment are likely to be somewhat unreliable historians. This is most notable for data requiring intact memory (as required in the documentation of previous behaviours), but also for data requiring relatively intact and sophisticated comprehension and communication. Therefore, the ideal respondent, especially in moderate to late dementia trials, is not the patient him/herself but a close contact with intact cognitive functioning. This is usually a family caregiver for community-dwelling patients, although he/she may lack the skills needed for sophisticated observation. In addition, caregivers’ reports may be influenced by their own emotional states or their relationship with the patient. Nevertheless, other non-family informants may be of limited helpfulness because of fewer and shorter periods of available observation, such as, in the case of data provision by visiting home care aides or nurses.

There has been debate with regard to the usefulness of measuring frequency versus severity. Reisberg et al pointed out that because the time spent by caregiver informants with patients varies greatly, frequency may be insensitive compared to the magnitude of the disturbance. It is also important to note that magnitude may be of greater clinical relevance. Tariot et al argue that frequency is preferable, as severity is more difficult to anchor. Some scales rely on a summation of frequency multiplied by severity. A further complication is that neuropsychiatric symptoms fluctuate, are not present in all patients, and do not necessarily progress at a uniform rate. Many scales have been described in the literature, but, for this review, the choice of scales is primarily based on neuropsychiatric symptom ratings commonly used in clinical trials. The Table lists the number of items, range of scores for each item and maximum score for each scale described in the paper.

### Mood Scales

Mood symptoms have an important and heterogeneous relationship to dementia. They may present as isolated symptoms and either predate the recognition of dementia, or appear in the course of an established dementing process. Mood symptoms may also be more pervasive and be part of a subsyndromal, yet clinically significant, depressive disorder. Alternatively, symptoms may be part of a major depressive disorder. Symptoms of mood disorders have considerable overlap with core symptoms of dementia, such as apathy, sleep disturbance, weight loss, and emotional dyscontrol. Therefore, instruments used to assess mood changes in dementia must avoid either over or under identification of mood syndromes. Self-reports are less reliable as dementia progresses, making the integration of caregiver information most important.

Many global dementia-rating instruments have included some mood symptoms as items contributing to the total score; however, mood items within these instruments are often scored in such a way that it is difficult to assess their magnitude, clinical importance, or progression, or to make a judgment of whether the mood symptom is an isolated symptom of dementia or part of a core mood disorder. On the other hand, more detailed mood instruments developed to describe and quantify mood disorders in the non-demented population [Hamilton Depression Rating Scale (1960); Montgomery-Asberg Depression Rating Scale (1979); Zung Scale (1965); Beck Depression Inventory (1961); Geriatric Depression Scale (1983)] are not valid in later stages of dementia when language and comprehension have declined significantly. Therefore, in studies where the assessment of mood in dementia is of major importance, it is best to use an instrument that has been developed specifically for this purpose. Two such instruments are the Dementia Mood Assessment Scale and the Cornell Scale for Depression in Dementia. Of course, even these are screening tools, and the gold-standard diagnosis of depressive disorder in dementia is still a clinical one. The process is complex, and requires a high level of experience, knowledge, acumen, and, usually, informed second-party information.

### The Cornell Scale for Depression in Dementia

The Cornell Scale is a well known, mostly caregiver-rated scale that is particularly suited to differentiating between cognitive and mood symptoms, and is sensitive to treatment effects over a wide range of depression severity. The scale has 19 items that are based on the week prior to the interview and which are rated as absent, mild or intermittent, and severe. Symptoms are clustered into five main categories: mood related signs, behavioural disturbance, physical signs, cyclic functions, and ideational disturbance. Published inter-rater reliability kappa is 0.67, internal consistency is reasonable (0.84), and it has been found to be valid, based on comparison with the Hamilton Depression Rating Scale and Research Diagnostic Criteria. The time required for this instrument is about 20 minutes with the caregiver and 10 minutes with the patient.
Table: Scoring of Scales

<table>
<thead>
<tr>
<th></th>
<th>Number of items</th>
<th>Score for each item</th>
<th>Maximum Score*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. NPI</td>
<td>12</td>
<td>0-12</td>
<td>144</td>
</tr>
<tr>
<td>2. CERAD-BRSD</td>
<td>46</td>
<td>0-4</td>
<td>148*</td>
</tr>
<tr>
<td>3. BEHAVE-AD</td>
<td>25</td>
<td>0-3</td>
<td>75</td>
</tr>
<tr>
<td>4. ADAS-Noncog</td>
<td>10</td>
<td>0-5</td>
<td>50</td>
</tr>
<tr>
<td>5. BPRS</td>
<td>16</td>
<td>1-7</td>
<td>112</td>
</tr>
<tr>
<td>6. Cornell Scale</td>
<td>19</td>
<td>0-2</td>
<td>38</td>
</tr>
<tr>
<td>7. DMAS</td>
<td>24</td>
<td>0-6</td>
<td>102b</td>
</tr>
<tr>
<td>8. CMAI-long form</td>
<td>29</td>
<td>1-7</td>
<td>203</td>
</tr>
</tbody>
</table>

* 37 items are rated 0-4. 8 items are scored yes/no and one asks for additional behaviours. b 17 items rate mood. c Higher scores indicate greater psychopathology on all scales.

Dementia Mood Assessment Scale

The Dementia Mood Assessment Scale (DMAS) was developed with an inpatient population by Sunderland and Minichello and was modeled on the Hamilton Depression Scale (HDS).15,19 Items were removed from the HDS that were felt to be subjective and vulnerable to error in demented patients. The final version of the DMAS has 24 items, with the first 17 items designed to assess the severity of depression and the last 7 items designed to measure the severity of dementia. Trained interviewers score items with input from nursing staff and/or family caregivers. Validity was found to be good, as measured against the Geriatric Depression Scale and the Montgomery-Asberg Depression Rating Scale. Inter-rater reliability and intra-class correlations were high. As this scale was developed with inpatients, and less information is available on it, it is used less often than the Cornell Scale.

Global Behaviour Scales

Neuropsychiatric Inventory (NPI)

This scale was published in 1994, and measures a wide range of neuropsychiatric disturbances.20 The initial scale evaluated ten items, which included delusions, hallucinations, dysphoria, anxiety, agitation, euphoria, apathy, irritability, disinhibition, and aberrant motor behaviour. Two items were later added (i.e., nighttime behaviour and changes in appetite and eating behaviours). It is based on a structured interview conducted by a clinician with a caregiver. Screening questions are asked, initially, and followed-up, if necessary, by subquestions. The caregiver rates the frequency on a 4-point scale and severity on a 3-point scale. On a test-retest reliability study, correlation coefficients were .79 for overall frequency and .86 for overall severity.21 Studies have reported good inter-rater reliability.20,22 There is high internal consistency, and criterion (concurrent) validity was also assessed by correlating questions measuring similar behaviours on the NPI and the BEHAVE-AD.21 Most item pairs showed good to moderate correlations. Several studies have evaluated the nursing home version of the NPI. A recent Canadian study suggested that five factors accounted for 63% of the variance.23 These factors were agitation, mood, psychosis, sleep/motor activity, and elevated behaviour. Convergent and discriminate validity of the five factors by correlating them with other behavioural measures was considered satisfactory. A study by Ivenson et al. looked at measuring change using the nursing home version of the NPI.24 Estimates of reliable change on the individual subscales range from 1.9 points on the euphoria/elation subscale to 5.13 points on the anxiety subscale. A change in the total score of plus or minus 22 points was required to exceed the possible range of measurement error at the 0.8 confidence interval.24

A number of versions of the NPI have been translated into other languages and these have also been studied. These studies have generally supported the use of the NPI and include translations into Greek;25 Spanish (NPI-Q);26 Dutch;27,28 Korean;29 and Chinese.30 A brief version of the NPI (NPI-Q) that is intended for use in routine clinical practice has also been studied.31 It has been cross-validated with the NPI and is considered to be a brief, reliable, informant-based assessment of neuropsychiatric symptoms and associated caregiver distress. A review by Forester and Oxman recommends the use of the NPI-Q as being the most appropriate for use in primary care.32

A review by Perrault of outcome measurement instruments in Alzheimer's disease suggested that more work was required to provide evidence of the NPI's responsiveness to change.2 Concerns were also raised about the gaps in score distribution (i.e., there are no multiples of 5, 7, and 11).

CERAD-BRSD

The Behaviour Rating Scale for Dementia of the Consortium to Establish a Registry for Alzheimer's Disease is composed of 46 questions, 37 of which are rated on a 5-point scale.33 The ratings measure behavioural symptoms over the previous month. Items were taken from previously existing scales and were designed to be administered to caregivers. The BRSD has six subscales: depressive symptoms, psychotic symptoms, inertia, vegetative symptoms, irritability/aggression, and behavioural disregulation. A study comparing the BRSD with the CMAI, the Revised Memory and Behavioural Problems Checklist (RMBPC), and the Agitated Behaviour and Dementia Scale (ABD), as compared to the Clinical Global Impression of Change (CGIC), was carried out by Weiner.34 The four specific behavioural/agitation subscales had excellent cross-sectional and longitudinal correlations with each other, suggesting high validity. Changes on the CGIC did not correlate well with changes scores on the other instruments.35 Test/retest reliability was moderately high and inter-rater reliability was high on individual items.8,35 The authors did suggest that estimates of reliability may have been inflated due to the lack of inclusion of subjects with extensive psychopathology. Construct validity was supported by the high correlation (r=.76) between the total BRSD and total CMAI scores in assessments of 206 patients with Alzheimer's disease.36 Validity was also supported from a factor analysis study.37 Perrault et al suggested that there was enough preliminary evidence on the reliability and validity of the BRSD to suggest that it can be used provisionally as an outcome measure in Alzheimer's disease drug trials.2
**Behavioural Pathology in AD Scale (BEHAVE-AD)**

This scale is one of the earliest rating scales to be used in the dementia field. The items were taken from a chart review of 57 outpatients with Alzheimer’s disease.38 It includes the assessment of symptoms and a global rating of caregiver distress. Twenty-five behaviours in seven clusters are rated. These areas include paranoid and delusional ideation, hallucinations, activity disturbances, aggressiveness, diurnal rhythm disturbances, affective disturbance, and anxieties and phobias. Caregivers rate the symptoms over the preceding two weeks on a 0 to 3 scale. The caregiver also determines a global assessment of caregiver distress on a scale of 0 to 3. The BEHAVE-AD is considered to be a valid and reliable scale, which takes approximately 20 minutes to administer.39 It assesses psychosis as well as agitation, aggression, and affective changes. The BEHAVE-AD has been used in recent pharmacological studies of the effectiveness of treatments for BPSD. A variation of the BEHAVE-AD, termed the E-BEHAVE-AD, which relies on the direct observation of observable behaviours and not phenomenology that would require a professionally trained examiner to elicit. Finally, a Chinese version of the BEHAVE-AD has been studied, and the findings suggest that the instrument is a valid tool for behavioural disturbances in patients with AD.40

**Other Instruments**

The ADAS-noncog is a subscale of the ADAS, which measures behaviour and mood symptoms.53,44 It rates patients on 10 items grouped into six clusters. The assessment is based on clinicians’ observations, plus an interview with a caregiver. Perrault et al note that the validity of the ADAS-noncog in Alzheimer’s disease drug trials is questionable.3 Three of its items are outside of the domain of behaviour and mood (tremors, concentration/distractibility, and appetite changes). In addition, aggressiveness and anxiety are not included in this scale. At this time, the ADAS-noncog is not considered to be one of the instruments of choice.

Despite the fact that the Brief Psychiatric Rating Scale (BPRS) was originally designed to assess drug treatment effects in a general adult psychiatric population, it has also been used in studies of the elderly.45,46 There are 18 items, each of which is rated on a scale of 1 to 7 according to severity. Schneider et al note that the BPRS has proven to be very useful and sensitive to changes in many AD clinical trials;6 however, Perrault et al believe that the scale has important shortcomings in terms of content validity and report that it includes items that are irrelevant, unspecific, or confounded by cognition.2 They also note that direct patient interview may lead to inadequate assessment and conclude that the BPRS has limited utility as an outcome measurement scale in Alzheimer’s disease drug trials.

**Scale for Observable Behaviour (Agitation)**

**Cohen-Mansfield Agitation Inventory (CMAI)**

The Cohen-Mansfield Agitation Inventory (CMAI) is an empirical scale that measures only observable behaviours, and does not consider patients’ mood or thought content. It was developed in nursing homes for nursing home use47,48 and is subject to floor effects in community dwelling patients with mild degrees of behavioural disturbance. The CMAI uses a 7-point scale to assess the frequency of 29 behaviours commonly seen in nursing home residents. Behaviours are characterized in four clusters: verbally aggressive (e.g., directed at a person or object), verbally nonaggressive (not directed at a specific object or person), physically aggressive (directed), and physically nonaggressive (undirected), but the total score is most commonly used to quantify behavioural disturbance. The scale is filled out by professional caregivers who are usually nurses or nursing assistants. The staff needs to be trained prior to using the instrument. The scale is observational and can be administered using either a short or long form. The CMAI takes an average of 20 minutes to complete. The instrument has found to be reliable and valid.49,50 Additional versions of the instrument have been developed, including a Community form (CMAI-C), a 38-item questionnaire for interviews with caregivers or relatives, and a short form (14 items).

**Conclusion**

Investigators who require a comprehensive instrument to measure neuropsychiatric symptoms in studies of patients with dementia should consider using the NPI, the CERAD-BRSD, or, possibly, the BEHAVE-AD. The NPI is currently the most frequently used scale. For investigators who are particularly focusing on mood, the Cornell Scale for Depression in Dementia and the DMAS are appropriate. The CMAI is particularly useful for personnel with limited training, as it measures only observable behaviours and not phenomenology that would require a professionally trained examiner to elicit.

**References**

8. Tarred PN, Mack JL, Patterson MB, Edland SD, Weiner MF, Fillenbaum G, et al. The Behaviour Rating Scale for Dementia of the Consortium to Establish a Registry for Alzheimer’s Disease: The Behavioral Pathology Committee of the Consortium to