Published in final edited form as:

Int Psychogeriatr. 2014 November ; 26(11): 1805–1848. doi:10.1017/S1041610214001537.

Assessing neuropsychiatric symptoms in people with dementia: a systematic review of measures

Laura N. Gitlin¹, Katherine A. Marx¹, Ian H. Stanley¹, Bryan R. Hansen¹, and Kimberly S. Van Haitsma²

¹Center for Innovative Care in Aging, Johns Hopkins University, Baltimore, Maryland, USA

²Polisher Research Institute, North Wales, Pennsylvania, USA

Abstract

Background—Neuropsychiatric symptoms (NPS) occur in people with dementia throughout disease course and across etiologies. NPS are associated with significant morbidities and hastened disease processes. Nevertheless, people with dementia are not systematically assessed for NPS in clinical settings. We review existing NPS measures for clinical and/or research purposes, and identify measurement gaps.

Methods—We conducted a computerized search of peer-reviewed published studies of measures (January 1, 1980–December 1, 2013) using multiple search terms. Measures selected for review were in English, had adequate psychometric properties, and were developed for or used with people with dementia. Papers describing measures were evaluated by three coders along seven characteristics: behavioral domains, number of items, method of administration, response categories, targeted population, setting, and psychometric properties.

Results—Overall, 2,233 papers were identified through search terms, and 36 papers from manual searches of references. From 2,269 papers, 85 measures were identified of which 45 (52.9%) had adequate psychometric properties and were developed or used with dementia populations. Of these, 16 (35.6%) were general measures that included a wide range of behaviors; 29 (64.4%) targeted specific behaviors (e.g. agitation). Measures differed widely as to behaviors assessed and measurement properties.

© International Psychogeriatric Association 2014

Correspondence should be addressed to: Laura N. Gitlin, PhD, Director/Professor, Center for Innovative Care in Aging, Johns Hopkins University, 525 North Wolfe Street, Suite 316, Baltimore, Maryland 21205, USA. Phone: +410-955-7539; Fax: 410-614-6873. lgitlin1@jhu.edu.

Conflict of interest: None.

Description of authors' roles: L.N. Gitlin conceptualized the paper and designed the approach, drafted the paper. K.A. Marx, helped to conduct the literature review, was one of the coders of measures, helped to construct the tables, drafted sections of the paper and critically read versions. I.H. Stanley helped to conduct the literature review, was one of the coders of measures, helped to construct the tables, drafted sections of the paper, helped to draft and critically read versions. B.R. Hansen helped to conduct the literature review, was one of the coders of measures, and critically read the final version. K.S. Van Haitsma critically read previous draft versions and provided conceptual guidance.

Partial results from this study were presented as a poster at the 2013 Alzheimer's Association International Conference in Boston, Massachusetts and in a symposium at the 66th annual Scientific Meeting of the Gerontological Society of America in New Orleans, Louisiana. A previous version of this review was developed in part for the Centers for Medicare and Medicaid Services Tool Kit to facilitate assessment of neuropsychiatric behaviors and use of non-pharmacologic practices in nursing homes.

Conclusions—A robust set of diverse measures exists for assessing NPS in different settings. No measures identify risk factors for behaviors or enable an evaluation of the context in which behaviors occur. To improve clinical efforts, research is needed to evaluate concordance of behavioral ratings between formal and informal caregivers, and to develop and test measures that can identify known risks for behaviors and the circumstances under which behaviors occur.

Keywords

neuropsychiatric symptoms; measures; assessments; behaviors; dementia; Alzheimer's disease

Introduction

Most persons with dementia, regardless of disease etiology or stage, will experience neuropsychiatric symptoms (NPS; Lyketsos *et al.*, 2011). Also referred to as behavioral symptoms or behavioral and psychological symptoms of dementia (BPSD), NPS include a wide range of behaviors that appear to cluster into five domains: depression, agitation, aggression, apathy, and psychosis (McShane, 2000). Other behavioral symptoms are also common although their phenotype is not well-delineated or understood. Behaviors such as sleep disturbances, anxiety, rejection of care, wandering, ingestion of dangerous objects, and repetitive questioning overlap with but also are distinguished from NPS clusters, and often present as the most challenging and distressful to families (Gitlin *et al.*, 2007; Volicer *et al.*, 2007; Gitlin *et al.*, 2010; Volicer *et al.*, 2013).

NPS can have dire consequences. For people with dementia, untreated symptoms are associated with poor quality of life (Karttunen *et al.*, 2011), premature nursing home placement (Chan *et al.*, 2003), higher healthcare utilization and costs (Murman *et al.*, 2002), and more rapid disease progression (Rabins *et al.*, 2013). Similarly, for informal caregivers, behavioral symptoms are associated with negative health and psychosocial consequences (Brodaty and Arasaratnam, 2012) including increased caregiver depression and burden (Okura and Langa, 2010). For formal providers, caring for persons with behavioral symptoms is associated with higher job dissatisfaction and burnout (Brodaty *et al.*, 2003; Miyamoto *et al.*, 2010).

Despite the immense clinical and public health sequelae of NPS, behaviors typically remain unidentified and therefore untreated in clinical settings. This may be due in part to the lack of understanding by health professionals of the importance of monitoring behavioral symptoms in dementia care and that adequate assessment tools for detection of behaviors exist (Gitlin *et al.*, 2012; Kales *et al.*, 2014). The first step in managing NPS is detecting their occurrences (Gitlin *et al.*, 2012; Kales *et al.*, 2014). The Physician Consortium for Performance Improvement (PCPI) recommends that people with dementia be assessed for occurrences of behavioral symptoms in physician practices on a yearly basis at a minimum and that reliable and validated instruments be used (Odenheimer *et al.*, 2013). It is unclear however the extent to which this recommendation is practiced. Also, there are no guidelines as to what measures to use, or how behaviors should be assessed and by whom, particularly for people with dementia who live in the community. Typically, behavioral symptoms are brought to the attention of physicians or health professionals by concerned family members

or other healthcare providers either following a behavioral event or when troublesome behaviors persistently disrupt care or compromise patient or family safety. This is in contrast to systematic and on-going assessment for detection of behaviors as recommended by PCPI.

To help close the gap between recommended and current practices as it pertains to the assessment of behavioral symptoms, we conducted a systematic review of existing, psychometrically sound measures of NPS. Previous reviews of measures are limited in that they targeted specific health professionals (e.g. nurses; Neville and Byrne, 2001), were not systematic in their approach (Conn and Thorpe, 2007), reviewed a subset of scales (Jeon *et al.*, 2011), or need to be updated (Cummings, 1996). In this review, we sought to be comprehensive and include all identified scales available in English that had been developed and/or tested with dementia populations and which had demonstrated psychometric adequacy. Our primary purpose is to evaluate the state of measurement in this area and compare characteristics of measures. Also, we identify measurement gaps that require future research, which, if pursued, could improve clinical practice.

Methods

Search procedure

A comprehensive computerized search of peer-reviewed published studies (January 1, 1980–December 1, 2013) was conducted in PubMed, CINAHL, CINAHL Plus, PsychInfo, Medline, and Mental Measurement Yearbook using the following search terms: *neuropsychological tests, neuropsychological measurements, dementia, Alzheimer's disease, behavior, delusions, hallucinations, agitation, aggression, depression, anxiety, eating, euphoria, apathy, disinhibition, irritability, motor disturbance, sleep, and vocalizations*. Additionally, a search of review papers and meta-analyses of neuropsychological measures was completed and the reference section crosschecked with the original search. Papers identified through this process were further searched for additional references to measures. All titles and abstracts of papers were reviewed independently by three trained coders (KAM, IHS, and BRH) who then met among themselves and the primary author to reconcile differences. Measures were selected with the following criteria: (a) published and available in English; (b) developed for or tested in people with dementia; (c) a sample size was reported for the testing of the measure; and (d) one or more psychometric properties were reported.

Measures were reviewed for seven characteristics: behavioral domains included in the measure, number of items, method of administration (e.g. self-report, observation), response categories used, targeted population, setting (e.g. nursing homes, hospitals), and psychometric properties (e.g. reported reliability, validity, sensitivity to change).

Results

The initial search yielded 2,233 papers. Through manual searches of these papers, an additional 36 papers were identified (see Figure 1). Among these 2,269 papers, 85 measures were identified, of which 45 (45/85; 52.9%) were included in this review as they had adequate psychometric properties and had been developed for and/or used with people with

dementia (Tables 1 and 2). Measures developed for other populations (e.g. hospital patients) but for which there was evidence of use with persons with dementia were included.

Of 40 measures not included in the review, 16 (40.0%) were excluded because there was insufficient evidence of their use in people with dementia. Eleven (27.5%) measures did not address neuropsychiatric behaviors; eight (20.0%) measures did not report psychometric properties; and one (2.5%) did not report a sample size. Additionally, despite placing limits on the search, four (10.0%) measures were initially identified but upon further inspection were determined to be outside of our search parameters (e.g. published before 1980).

Behaviors

We categorized measures as either general that included a wide range of behavioral domains, or as specific, in which a particular behavioral domain was targeted. Of 45 measures, 16 (35.6%) were general in which the number of behavioral domains for any one measure varied from 1 to 14 (Table 1). There was no consistency across measures as to the specific domains or behavioral items included. Some measures also addressed behaviors that did not reflect NPS such as items related to self care (Helmès *et al.* 1987; Wahle *et al.*, 1996).

Of 45 measures, most ($n = 29$, 64.4%) targeted a specific NPS (Table 2). Measures were identified for eight specific behavioral domains: agitation representing the most number of measures of the specific type ($n = 7$, 24.1%; Cohen-Mansfield *et al.*, 1989; Mungas *et al.*, 1989; Finkel *et al.*, 1993; Rosen *et al.*, 1994; Yudofsky *et al.*, 1997; Hurley *et al.*, 1999; Logsdon *et al.*, 1999), followed by apathy ($n = 5$, 17.2%; Burns *et al.*, 1990; Marin *et al.*, 1991; Robert *et al.*, 2002; Strauss and Sperry, 2002; Sockeel *et al.*, 2006), aggression ($n = 4$, 13.8%; Yudofsky *et al.*, 1986; Ryden, 1988; Patel and Hope, 1992; Perlman and Hirdes, 2008), anxiety ($n = 4$, 13.8%; Beck *et al.*, 1988; LaBarge, 1993; Shankar *et al.*, 1999; Pachana *et al.*, 2007), depression ($n = 4$, 13.8%; Yesavage *et al.*, 1983; Alexopoulos *et al.*, 1988; Sunderland *et al.*, 1988; Kroenke *et al.*, 2001), sleep ($n = 3$, 10.3%; Buysse *et al.*, 1989; Johns, 1991; Tractenberg *et al.*, 2003), depression and anxiety ($n = 1$, 3.4%; Zigmond and Snaith, 1983), and wandering ($n = 1$, 3.4%; Algase *et al.*, 2001; Table 2). No measures were identified that specifically targeted euphoria, delusions, hallucinations, irritability apart from aggression or anxiety, or motor and verbal disturbances.

Number of items

Across all 45 measures, the number of items was highly variable, ranging from 3 to a potential of 142 (Tables 1 and 2). For general measures, the number of items ranged from 10 (Weyer *et al.*, 1997) to 142 (Cummings *et al.*, 1994). The latter – the Neuropsychiatric Inventory (NPI-C) – was comprised of 14 behavioral domains (delusions, hallucinations, agitation, aggression, dysphoria, anxiety, euphoria, apathy, disinhibition, irritability, aberrant motor behavior, nighttime behavior disturbances, changes in appetite and eating behaviors, and aberrant vocalizations). Each domain in the NPI-C had a yes/no screening question with potential follow-up questions (ranging from 7 to 16 items) that specify different possible behaviors of that domain.

For targeted measures, the number of items was also highly variable ranging from 3 (Robert *et al.*, 2002) to 38 (Algase *et al.*, 2004).

Method of administration

We examined who administered measures and format of ascertainment. Of 45 measures, 14 (31.1%) were administered by an interviewer trained in that measure (Hughes *et al.*, 1982; Yesavage *et al.*, 1983; Sunderland *et al.*, 1988; Buysse *et al.*, 1989; Johns, 1991; Devanand *et al.*, 1992; Cummings *et al.*, 1994; Loreck *et al.*, 1994; Sultzer *et al.*, 1995; Tariot *et al.*, 1995; Reisberg *et al.*, 1997; Weyer *et al.*, 1997; Strauss and Sperry, 2002; Tractenberg *et al.*, 2003). Some measures were administered by nurses ($n = 6$, 13.3%; Yudofsky *et al.*, 1986; Helmes *et al.*, 1987; Patel and Hope, 1992; Ray *et al.*, 1992; Finkel *et al.*, 1993; Rosen *et al.*, 1994), clinicians ($n = 5$, 11.1%; Alexopoulos *et al.*, 1988; Burns *et al.*, 1990; Shankar *et al.*, 1999; Robert *et al.*, 2002; Sockeel *et al.*, 2006), or by trained observers ($n = 4$, 8.9%; Burgio *et al.*, 1994; Yudofsky *et al.*, 1997; Hurley *et al.*, 1999; Perlman and Hirdes, 2008). Several of the measures used self-administration by family caregivers ($n = 3$, 6.7%; Teri *et al.*, 1992; Grace *et al.*, 1999; Kolitz *et al.*, 2003) or the person with dementia ($n = 3$, 6.7%; Zigmond and Snaith, 1983; Beck *et al.*, 1988; LaBarge, 1993). Ten (22.2%) measures were designed to be administered by more than one person (Ryden, 1988; Cohen-Mansfield *et al.*, 1989; Mungas *et al.*, 1989; Baumgarten *et al.*, 1990; Marin *et al.*, 1991; Wahle *et al.*, 1996; Logsdon *et al.*, 1999; Algase *et al.*, 2001; Kroenke *et al.*, 2001; Pachana *et al.*, 2007).

Most measures ($n = 30$, 66.7%) were in interview format (Hughes *et al.*, 1982; Yesavage *et al.*, 1983; Zigmond and Snaith, 1983; Alexopoulos *et al.*, 1988; Beck *et al.*, 1988; Ryden, 1988; Buysse *et al.*, 1989; Cohen-Mansfield *et al.*, 1989; Baumgarten *et al.*, 1990; Burns *et al.*, 1990; Johns, 1991; Marin *et al.*, 1991; Devanand *et al.*, 1992; Teri *et al.*, 1992; LaBarge, 1993; Cummings *et al.*, 1994; Loreck *et al.*, 1994; Sultzer *et al.*, 1995; Tariot *et al.*, 1995; Reisberg *et al.*, 1997; Weyer *et al.*, 1997; Grace *et al.*, 1999; Logsdon *et al.*, 1999; Kroenke *et al.*, 2001; Robert *et al.*, 2002; Strauss and Sperry, 2002; Kolitz *et al.*, 2003; Tractenberg *et al.*, 2003; Sockeel *et al.*, 2006; Pachana *et al.*, 2007), whereas 11 (24.4%) used direct observation (Yudofsky *et al.*, 1986; Helmes *et al.*, 1987; Patel and Hope, 1992; Ray *et al.*, 1992; Finkel *et al.*, 1993; Burgio *et al.*, 1994; Rosen *et al.*, 1994; Wahle *et al.*, 1996; Yudofsky *et al.*, 1997; Hurley *et al.*, 1999; Algase *et al.*, 2001), and one (2.2%) used chart review (Perlman and Hirdes, 2008). Three (6.7%) measures utilized two different formats of administration such as interview and chart extraction (Sunderland *et al.*, 1988; Mungas *et al.*, 1989; Shankar *et al.*, 1999).

For each scale, we also examined who provided the behavioral information. Over half of the scales were administered directly only to the person with dementia ($n = 24$, 53.3%; Yesavage *et al.*, 1983; Zigmond and Snaith, 1983; Yudofsky *et al.*, 1986; Helmes *et al.*, 1987; Beck *et al.*, 1988; Sunderland *et al.*, 1988; Buysse *et al.*, 1989; Mungas *et al.*, 1989; Marin *et al.*, 1991; Patel and Hope, 1992; Ray *et al.*, 1992; Finkel *et al.*, 1993; LaBarge, 1993; Burgio *et al.*, 1994; Rosen *et al.*, 1994; Sultzer *et al.*, 1995; Wahle *et al.*, 1996; Yudofsky *et al.*, 1997; Hurley *et al.*, 1999; Algase *et al.*, 2001; Kroenke *et al.*, 2001; Sockeel *et al.*, 2006; Pachana *et al.*, 2007; Perlman and Hirdes, 2008), whereas 12 (26.7%) relied only upon proxy report (Cohen-Mansfield *et al.*, 1989; Baumgarten *et al.*, 1990; Burns

et al., 1990; Devanand *et al.*, 1992; Teri *et al.*, 1992; Cummings *et al.*, 1994; Tariot *et al.*, 1995; Reisberg *et al.*, 1997; Logsdon *et al.*, 1999; Strauss and Sperry, 2002; Kolitz *et al.*, 2003; Tractenberg *et al.*, 2003), and nine (20.0%) were administered to both the person with dementia and a proxy (Hughes *et al.*, 1982; Alexopoulos *et al.*, 1988; Ryden, 1988; Johns, 1991; Loreck *et al.*, 1994; Weyer *et al.*, 1997; Grace *et al.*, 1999; Shankar *et al.*, 1999; Robert *et al.*, 2002).

Among the general measures ($n = 16$), most were administered by trained interviewers ($n = 8$, 50.0%), of which only the clinical version of the NPI (NPI-C) was designed to be completed by a clinically trained individual (Hughes *et al.*, 1982; Devanand *et al.*, 1992; Cummings *et al.*, 1994; Loreck *et al.*, 1994; Sultzer *et al.*, 1995; Tariot *et al.*, 1995; Reisberg *et al.*, 1997; Weyer *et al.*, 1997). Three (18.8%) of the general measures were for self-administration by families (Teri *et al.*, 1992; Grace *et al.*, 1999; Kolitz *et al.*, 2003), two (12.5%) were for nurses to complete (Helmès *et al.*, 1987; Ray *et al.*, 1992) and one (6.3%) was designed for administration by a trained observer (Burgio *et al.*, 1994). Two (12.5%) could be administered by multiple persons including a trained interviewer or nurse or a family caregiver (Baumgarten *et al.*, 1990; Wahle *et al.*, 1996). Twelve (75.0%) of the general scales were in an interview format (Hughes *et al.*, 1982; Baumgarten *et al.*, 1990; Devanand *et al.*, 1992; Teri *et al.*, 1992; Cummings *et al.*, 1994; Loreck *et al.*, 1994; Sultzer *et al.*, 1995; Tariot *et al.*, 1995; Reisberg *et al.*, 1997; Weyer *et al.*, 1997; Grace *et al.*, 1999; Kolitz *et al.*, 2003) and the remaining four (25.0%) relied on observation (Helmès *et al.*, 1987; Ray *et al.*, 1992; Burgio *et al.*, 1994; Wahle *et al.*, 1996).

Four (25.0%) of the general measures asked both a proxy and the person with dementia to provide information (Hughes *et al.*, 1982; Loreck *et al.*, 1994; Weyer *et al.*, 1997; Grace *et al.*, 1999), seven (43.8%) relied on proxy response only (Baumgarten *et al.*, 1990; Devanand *et al.*, 1992; Teri *et al.*, 1992; Cummings *et al.*, 1994; Tariot *et al.*, 1995; Reisberg *et al.*, 1997; Kolitz *et al.*, 2003) and five (31.3%) were data provided by the person with dementia (Helmès *et al.*, 1987; Ray *et al.*, 1992; Burgio *et al.*, 1994; Sultzer *et al.*, 1995; Wahle *et al.*, 1996). Of these five, four were observation and only one, the Neurobehavioral Rating Scale (NRS; Levin *et al.*, 1987; Sultzer *et al.*, 1995), was an interview format directed only towards the person with dementia.

Among the behavior-specific measures ($n = 29$), six (20.7%) were administered by a trained interviewer (Yesavage *et al.*, 1983; Sunderland *et al.*, 1988; Buysse *et al.*, 1989; Johns, 1991; Strauss and Sperry, 2002; Tractenberg *et al.*, 2003), four (13.8%) by a nurse (Yudofsky *et al.*, 1986; Patel and Hope, 1992; Finkel *et al.*, 1993; Rosen *et al.*, 1994), five (17.2%) by a clinician (Alexopoulos *et al.*, 1988; Burns *et al.*, 1990; Shankar *et al.*, 1999; Robert *et al.*, 2002; Sockeel *et al.*, 2006), three (10.3%) addressed the person with dementia directly (Zigmond and Snaith, 1983; Beck *et al.*, 1988; LaBarge, 1993), and trained observers were used by three (10.3%) of the behavior-specific measures (Yudofsky *et al.*, 1997; Hurley *et al.*, 1999; Perlman and Hirdes, 2008). Eight (27.6%) of these measures could be completed by multiple people (Ryden, 1988; Cohen-Mansfield *et al.*, 1989; Mungas *et al.*, 1989; Marin *et al.*, 1991; Logsdon *et al.*, 1999; Algase *et al.*, 2001; Kroenke *et al.*, 2001; Pachana *et al.*, 2007). Most targeted measures were administered, at least in part, through interview format ($n = 18$, 62.1%; Yesavage *et al.*, 1983, 1986; Zigmond and

Snaith, 1983; Alexopoulos *et al.*, 1988; Beck *et al.*, 1988; Ryden, 1988; Buysse *et al.*, 1989; Cohen-Mansfield *et al.*, 1989; Burns *et al.*, 1990; Johns, 1991; Marin *et al.*, 1991; LaBarge, 1993; Logsdon *et al.*, 1999; Algase *et al.*, 2001; Kroenke *et al.*, 2001; Robert *et al.*, 2002; Strauss and Sperry, 2002; Tractenberg *et al.*, 2003; Sockeel *et al.*, 2006; Pachana *et al.*, 2007), seven (24.1%) used observation (Yudofsky *et al.*, 1986; Patel and Hope, 1992; Finkel *et al.*, 1993; Rosen *et al.*, 1994; Yudofsky *et al.*, 1997; Hurley *et al.*, 1999), and three (10.3%) utilized multiple methods (Sunderland *et al.*, 1988; Mungas *et al.*, 1989; Shankar *et al.*, 1999). Only one (3.4%) measure relied solely on chart review (Perlman and Hirdes, 2008), utilizing the Minimum Data Set (MDS) data. Most ($n = 19$, 65.5%) of the targeted measures relied on data exclusively from the person with dementia (Yesavage *et al.*, 1983; Zigmond and Snaith, 1983; Yudofsky *et al.*, 1986; Beck *et al.*, 1988; Sunderland *et al.*, 1988; Buysse *et al.*, 1989; Mungas *et al.*, 1989; Marin *et al.*, 1991; Patel and Hope, 1992; Finkel *et al.*, 1993; LaBarge, 1993; Rosen *et al.*, 1994; Yudofsky *et al.*, 1997; Hurley *et al.*, 1999; Algase *et al.*, 2001; Kroenke *et al.*, 2001; Sockeel *et al.*, 2006; Pachana *et al.*, 2007; Perlman and Hirdes, 2008), five (17.2%) targeted measures were administered to both the person with dementia and a proxy (Alexopoulos *et al.*, 1988; Ryden, 1988; Johns, 1991; Shankar *et al.*, 1999; Robert *et al.*, 2002) and five (17.2%) were administered only to a proxy (Cohen-Mansfield *et al.*, 1989; Burns *et al.*, 1990; Logsdon *et al.*, 1999; Strauss and Sperry, 2002; Tractenberg *et al.*, 2003).

Response categories

Time frames used in measures ranged from a 15-minute observational period (e.g. Yudofsky *et al.*, 1997), to the past three days (e.g. Ray *et al.*, 1992), past week (e.g. Weyer *et al.*, 1997), past two weeks (e.g. Wahle *et al.*, 1996), past month (e.g. Loreck *et al.*, 1994), past year (e.g. Ryden, 1988), or a comparison of premorbid state to present condition (e.g. Burns *et al.*, 1990). Some scales did not specify a response time frame (e.g. LaBarge, 1993).

Measures also varied widely as to the response categories used. For example, the Cohen-Mansfield Agitation Inventory (CMAI; Cohen-Mansfield *et al.*, 1989), a 29-item measure to assess agitation through an informant interview, has a response format that ranges from 1 = never to 7 = several times in an hour in the past two weeks, with possible scores ranging from 0 to 203. Three (6.7%) measures assessed presence and absence of behaviors (Cummings *et al.*, 1994; Robert *et al.*, 2002; Pachana *et al.*, 2007); however, if present, then two scales involved follow-up questions to obtain a more nuanced assessment of the specific characteristics of the behavioral occurrence, their frequency, and severity (Cummings *et al.*, 1994; Robert *et al.*, 2002).

Of the 45 measures, 25 (55.6%) assessed frequency of occurrence although the ascertainment of frequency varied widely. For example, the Dementia Signs and Symptoms Scale (DSS) included 43 items and eight subscales, and answer choices reflected behaviors over the past month ranging from 0 = absent to 3 = daily (Loreck *et al.*, 1994).

Only four (8.9%) measures assessed caregiver reaction to behaviors (Teri *et al.*, 1992; Cummings *et al.*, 1994; Logsdon *et al.*, 1999; Tractenberg *et al.*, 2003).

Target population and setting

Most measures were developed for use with any of the dementias; this was the case for both general and targeted measures. Only one measure was designed specifically for persons with frontotemporal dementia: the Frontal System Behavior Scale (FrSBe; Grace *et al.*, 1999). This is a 46-item scale that assesses three domains of functioning: apathy (e.g. “Speaks only when spoken to”), disinhibition (e.g. “Does or says embarrassing things”), and executive dysfunction (e.g. Says one thing, then does another). Response items range from 1 = Almost Never to 5 = Almost Always, with higher scores representing greater behavioral abnormality (Table 1).

Among measures included in this review, most ($n = 37$, 82.2%) were designed for use in a specific setting (Table 3): 15 (33.3%) for use in home or community-based settings (Hughes *et al.*, 1982; Beck *et al.*, 1988; Baumgarten *et al.*, 1990; Burns *et al.*, 1990; Devannand *et al.*, 1992; Teri *et al.*, 1992; LaBarge, 1993; Loreck *et al.*, 1994; Grace *et al.*, 1999; Logsdon *et al.*, 1999; Robert *et al.*, 2002; Strauss and Sperry, 2002; Kolitz *et al.*, 2003; Tractenberg *et al.*, 2003; Sockeel *et al.*, 2006); six (13.3%) for use in nursing homes (Mungas *et al.*, 1989; Patel and Hope, 1992; Ray *et al.*, 1992; Finkel *et al.*, 1993; Burgio *et al.*, 1994; Hurley *et al.*, 1999); and three (6.7%) for use in the hospital setting (Yudofsky *et al.*, 1986; Sultzer *et al.*, 1995; Yudofsky *et al.*, 1997). Thirteen (28.9%) were designed and tested for psychometrics in multiple settings (Zigmond and Snaith, 1983; Helmes *et al.*, 1987; Ryden, 1988; Sunderland *et al.*, 1988; Cohen-Mansfield *et al.*, 1989; Rosen *et al.*, 1994; Wahle *et al.*, 1996; Reisberg *et al.*, 1997; Weyer *et al.*, 1997; Shankar *et al.*, 1999; Algase *et al.*, 2001; Pachana *et al.*, 2007; Perlman and Hirdes, 2008). Further, eight (17.8%) of the 45 measures did not specify a specific setting (Yesavage *et al.*, 1983; Alexopoulos *et al.*, 1988; Buysse *et al.*, 1989; Johns, 1991; Marin *et al.*, 1991; Cummings *et al.*, 1994; Tariot *et al.*, 1995; Kroenke *et al.*, 2001).

Psychometric properties

As stated, inclusion criteria for this review narrowed the scope of assessment measures chosen for review to those for which psychometric data (e.g. either validity or reliability) were explicitly reported. Also, we examined if sensitivity to change was reported. Among 45 scales included in this review, 42 (93.3%) reported on validity and 44 (97.8%) reported on reliability. Using the specific terminology found in the papers cited in Tables 1 and 2, overall validity reports included convergent validity ($n = 12$; Yesavage *et al.*, 1983; Zigmond and Snaith, 1983; Helmes *et al.*, 1987; Beck *et al.*, 1988; Burns *et al.*, 1990; Marin *et al.*, 1991; Ray *et al.*, 1992; Teri *et al.*, 1992; Weyer *et al.*, 1997; Yudofsky *et al.*, 1997; Strauss and Sperry, 2002; Kolitz *et al.*, 2003), concurrent validity ($n = 15$; Alexopoulos *et al.*, 1988; Mungas *et al.*, 1989; Patel and Hope, 1992; Teri *et al.*, 1992; Finkel *et al.*, 1993; LaBarge, 1993; Cummings *et al.*, 1994; Loreck *et al.*, 1994; Wahle *et al.*, 1996; Hurley *et al.*, 1999; Shankar *et al.*, 1999; Robert *et al.*, 2002; Sockeel *et al.*, 2006; Pachana *et al.*, 2007; Perlman and Hirdes, 2008), construct validity ($n = 10$; Ryden, 1988; Sunderland *et al.*, 1988; Baumgarten *et al.*, 1990; Burns *et al.*, 1990; LaBarge, 1993; Reisberg *et al.*, 1997; Grace *et al.*, 1999; Shankar *et al.*, 1999; Kroenke *et al.*, 2001; Kolitz *et al.*, 2003), content validity ($n = 7$; Ryden, 1988; Cummings *et al.*, 1994; Yudofsky *et al.*, 1997; Hurley *et al.*, 1999; Shankar *et al.*, 1999; Kolitz *et al.*, 2003; Tractenberg *et al.*, 2003), discriminant

validity ($n = 8$; Beck *et al.*, 1988; Burns *et al.*, 1990; Marin *et al.*, 1991; Teri *et al.*, 1992; Yudofsky *et al.*, 1997; Robert *et al.*, 2002; Strauss and Sperry, 2002; Kolitz *et al.*, 2003), criterion validity ($n = 4$; Hughes *et al.*, 1982; Devanand *et al.*, 1992; Kroenke *et al.*, 2001; Strauss and Sperry, 2002), divergent validity ($n = 1$; Devanand *et al.*, 1992), and neuropathological validity ($n = 1$; Hughes *et al.*, 1982). For four measures, the type of validity was unspecified ($n = 4$; Cohen-Mansfield *et al.*, 1989; Rosen *et al.*, 1994; Logsdon *et al.*, 1999; Algase *et al.*, 2001).

Overall reliability reports included internal consistency ($n = 32$; Yesavage *et al.*, 1983; Zigmond and Snaith, 1983; Helmes *et al.*, 1987; Alexopoulos *et al.*, 1988; Beck *et al.*, 1988; Ryden, 1988; Buysse *et al.*, 1989; Cohen-Mansfield *et al.*, 1989; Baumgarten *et al.*, 1990; Burns *et al.*, 1990; Johns, 1991; Marin *et al.*, 1991; Devanand *et al.*, 1992; Patel and Hope, 1992; Teri *et al.*, 1992; Finkel *et al.*, 1993; LaBarge, 1993; Loreck *et al.*, 1994; Weyer *et al.*, 1997; Yudofsky *et al.*, 1997; Grace *et al.*, 1999; Hurley *et al.*, 1999; Logsdon *et al.*, 1999; Shankar *et al.*, 1999; Algase *et al.*, 2001; Kroenke *et al.*, 2001; Robert *et al.*, 2002; Strauss and Sperry, 2002; Kolitz *et al.*, 2003; Sockeel *et al.*, 2006; Pachana *et al.*, 2007; Perlman and Hirdes, 2008), inter-rater (or inter-observer) reliability ($n = 28$; Hughes *et al.*, 1982; Yudofsky *et al.*, 1986; Helmes *et al.*, 1987; Alexopoulos *et al.*, 1988; Ryden, 1988; Sunderland *et al.*, 1988; Cohen-Mansfield *et al.*, 1989; Mungas *et al.*, 1989; Burns *et al.*, 1990; Devanand *et al.*, 1992; Patel and Hope, 1992; Ray *et al.*, 1992; Finkel *et al.*, 1993; Burgio *et al.*, 1994; Cumming *et al.*, 1994; Loreck *et al.*, 1994; Rosen *et al.*, 1994; Tariot *et al.*, 1995; Wahle *et al.*, 1996; Reisberg *et al.*, 1997; Yudofsky *et al.*, 1997; Hurley *et al.*, 1999; Shankar *et al.*, 1999; Robert *et al.*, 2002; Strauss and Sperry, 2002; Sultzer *et al.*, 1995; Sockeel *et al.*, 2006; Pachana *et al.*, 2007), and test-retest reliability ($n = 17$; Yesavage *et al.*, 1983; Beck *et al.*, 1988; Ryden, 1988; Buysse *et al.*, 1989; Baumgarten *et al.*, 1990; Burns *et al.*, 1990; Marin *et al.*, 1991; Patel and Hope, 1992; Cummings *et al.*, 1994; Wahle *et al.*, 1996; Weyer *et al.*, 1997; Logsdon *et al.*, 1999; Shankar *et al.*, 1999; Robert *et al.*, 2002; Strauss and Sperry, 2002; Sockeel *et al.*, 2006; Pachana *et al.*, 2007).

All 16 general assessment scales reported on reliability data, with eight (50.0%) reporting on internal consistency (e.g. Chronbach's α ; Helmes *et al.*, 1987; Baumgarten *et al.*, 1990; Devanand *et al.*, 1992; Teri *et al.*, 1992; Loreck *et al.*, 1994; Weyer *et al.*, 1997; Grace *et al.*, 1999; Kolitz *et al.*, 2003), 11 (68.8%) reporting on inter-rater (or inter-observer) reliability (Hughes *et al.*, 1982; Helmes *et al.*, 1987; Devanand *et al.*, 1992; Ray *et al.*, 1992; Burgio *et al.*, 1994; Cumming *et al.*, 1994; Loreck *et al.*, 1994; Tariot *et al.*, 1995; Wahle *et al.*, 1996; Reisberg *et al.*, 1997; Sultzer *et al.*, 1995), and four (25.0%) reporting on test-retest reliability (Baumgarten *et al.*, 1990; Cummings *et al.*, 1994; Wahle *et al.*, 1996; Weyer *et al.*, 1997). No general assessment scale reported on all three of these reliability metrics.

With regard to validity data, 13 (81.2%) of the general assessment measures reported data on one or more of the following: convergent validity ($n = 5$; Helmes *et al.*, 1987; Ray *et al.*, 1992; Teri *et al.*, 1992; Weyer *et al.*, 1997; Kolitz *et al.*, 2003), concurrent validity ($n = 4$; Teri *et al.*, 1992; Cummings *et al.*, 1994; Loreck *et al.*, 1994; Wahle *et al.*, 1996), construct validity ($n = 4$; Baumgarten *et al.*, 1990; Reisberg *et al.*, 1997; Grace *et al.*, 1999; Kolitz *et al.*, 2003), content validity ($n = 2$; Cummings *et al.*, 1994; Kolitz *et al.*, 2003), discriminant

validity ($n = 2$; Teri *et al.*, 1992; Kolitz *et al.*, 2003), criterion validity ($n = 2$; Hughes *et al.*, 1982; Devanand *et al.*, 1992), divergent validity ($n = 1$; Devanand *et al.*, 1992), and neuropatho-logical validity ($n = 1$; Hughes *et al.*, 1982).

A similar pattern emerged for the 29 specific behavior scales. The majority presented reliability data ($n = 28$; 96.6%), of which 24 reported internal consistency (e.g. Chronbach's α ; Yesavage *et al.*, 1983; Zigmond and Snaith, 1983; Alexopoulos *et al.*, 1988; Beck *et al.*, 1988; Ryden, 1988; Buysse *et al.*, 1989; Cohen-Mansfield *et al.*, 1989; Burns *et al.*, 1990; Johns, 1991; Marin *et al.*, 1991; Patel and Hope, 1992; Finkel *et al.*, 1993; LaBarge, 1993; Yudofsky *et al.*, 1997; Hurley *et al.*, 1999; Logsdon *et al.*, 1999; Shankar *et al.*, 1999; Algase *et al.*, 2001; Kroenke *et al.*, 2001; Robert *et al.*, 2002; Strauss and Sperry, 2002; Sockeel *et al.*, 2006; Pachana *et al.*, 2007; Perlman and Hirdes, 2008), 17 reported inter-rater reliability (Yudofsky *et al.*, 1986; Alexopoulos *et al.*, 1988; Ryden, 1988; Sunderland *et al.*, 1988; Cohen-Mansfield *et al.*, 1989; Mungas *et al.*, 1989; Burns *et al.*, 1990; Patel and Hope, 1992; Finkel *et al.*, 1993; Rosen *et al.*, 1994; Yudofsky *et al.*, 1997; Hurley *et al.*, 1999; Shankar *et al.*, 1999; Robert *et al.*, 2002; Strauss and Sperry, 2002; Sockeel *et al.*, 2006; Pachana *et al.*, 2007), and 13 reported test-retest reliability (Yesavage *et al.*, 1983; Beck *et al.*, 1988; Ryden, 1988; Buysse *et al.*, 1989; Burns *et al.*, 1990; Marin *et al.*, 1991; Patel and Hope, 1992; Logsdon *et al.*, 1999; Shankar *et al.*, 1999; Robert *et al.*, 2002; Strauss and Sperry, 2002; Sockeel *et al.*, 2006; Pachana *et al.*, 2007).

With regard to validity data, the following were reported for the specific behavior scales: convergent validity ($n = 7$; Yesavage *et al.*, 1983; Zigmond and Snaith, 1983; Beck *et al.*, 1988; Burns *et al.*, 1990; Marin *et al.*, 1991; Yudofsky *et al.*, 1997; Strauss and Sperry, 2002), concurrent validity ($n = 11$; Alexopoulos *et al.*, 1988; Mungas *et al.*, 1989; Patel and Hope, 1992; Finkel *et al.*, 1993; LaBarge, 1993; Hurley *et al.*, 1999; Shankar *et al.*, 1999; Robert *et al.*, 2002; Sockeel *et al.*, 2006; Pachana *et al.*, 2007; Perlman and Hirdes, 2008), construct validity ($n = 6$; Ryden, 1988; Sunderland *et al.*, 1988; Burns *et al.*, 1990; LaBarge, 1993; Shankar *et al.*, 1999; Kroenke *et al.*, 2001), content validity ($n = 5$; Ryden, 1988; Yudofsky *et al.*, 1997; Hurley *et al.*, 1999; Shankar *et al.*, 1999; Tractenberg *et al.*, 2003), discriminant validity ($n = 6$; Beck *et al.*, 1988; Burns *et al.*, 1990; Marin *et al.*, 1991; Yudofsky *et al.*, 1997; Robert *et al.*, 2002; Strauss and Sperry, 2002), criterion validity ($n = 2$; Kroenke *et al.*, 2001; Strauss and Sperry, 2002), and type not specified ($n = 4$; Cohen-Mansfield *et al.*, 1989; Rosen *et al.*, 1994; Logsdon *et al.*, 1999; Algase *et al.*, 2001).

Of the 45 measures reviewed, sensitivity to change was discussed for 31 (68.9%) of the measures, although the reporting quality was highly variable (Hughes *et al.*, 1982; Yesavage *et al.*, 1983; Zigmond and Snaith, 1983; Yudofsky *et al.*, 1986; Helmes *et al.*, 1987; Levin *et al.*, 1987; Alexopoulos *et al.*, 1988; Ryden, 1988; Sunderland *et al.*, 1988; Buysse *et al.*, 1989; Mungas *et al.*, 1989; Marin *et al.*, 1991; Devanand *et al.*, 1992; Patel and Hope, 1992; Ray *et al.*, 1992; Teri *et al.*, 1992; Cummings *et al.*, 1994; Loreck *et al.*, 1994; Rosen *et al.*, 1994; Tariot *et al.*, 1995; Wahle *et al.*, 1996; Weyer *et al.*, 1997; Grace *et al.*, 1999; Logsdon *et al.*, 1999; Shankar *et al.*, 1999; Kroenke *et al.*, 2001; Robert *et al.*, 2002; Kolitz *et al.*, 2003; Sockeel *et al.*, 2006; Boddice *et al.*, 2008; Perlman and Hirdes, 2008). Reports ranged from actual tests and data for sensitivity to change analyses (e.g. Weyer *et al.*, 1997) to statements acknowledging the need for future studies to test for sensitivity to change (e.g.

Yesavage *et al.*, 1983). For the remaining 14 (31.1%) measures, there was no indication or discussion of sensitivity to change (Beck *et al.*, 1988; Cohen-Mansfield *et al.*, 1989; Baumgarten *et al.*, 1990; Burns *et al.*, 1990; Johns, 1991; Finkel *et al.*, 1993; LaBarge, 1993; Burgio *et al.*, 1994; Reisberg *et al.*, 1997; Yudofsky *et al.*, 1997; Hurley *et al.*, 1999; Algase *et al.*, 2001; Strauss and Sperry, 2002; Tractenberg *et al.*, 2003).

Discussion

Undetected and therefore untreated NPS have potent effects on disease course, disease management, and health-related morbidity and mortality for people with dementia, their family members, and formal providers. Thus, assessing behavioral symptoms in clinical settings using reliable and valid measures is critical. Also, with the need to develop and test treatment strategies for effectively managing behavioral symptoms, it is important to identify adequate and standardized measures that can be used in clinical trials testing either pharmacologic or non-pharmacologic treatment strategies.

This systematic review uncovered 45 measures published in English with reported psychometric properties that can identify behavioral occurrences among people with dementia and which can be used in a multitude of clinical and research contexts. Measures varied widely in their characteristics reflecting in part the lack of consensus as to what constitutes NPS, and how behaviors should be identified, characterized, classified, and assessed. However, it also reflects the variegated needs of clinical settings and research purposes. Thus, this array of measures provides opportunities for assessing behaviors for almost any clinical or research purpose or context. Given that all 45 measures were selected because they had at least minimal adequate psychometric properties (validity and/or reliability), we show that clinicians and researchers have a robust set of tools from which to choose.

It is not possible to indicate if one measure is superior to another as use depends upon a number of important considerations. Choice of a measure should be guided by several considerations including the purpose of measurement (need for a general screen or a more targeted behavioral assessment), the setting in which the assessment of behaviors will occur, who will assess for behaviors and how assessment will occur (observation, proxy interview), and the amount of time and resources available for assessing for behavioral symptoms. Recent algorithmic decision-trees for behavioral management (Gitlin *et al.*, 2012) suggest that clinicians should start with a general measure of behavioral symptoms that can serve as an all-purpose screen for a wide range of behavioral domains. Then, based on the screen, a specific, behavior-targeted measure could be introduced as a follow-up to provide greater specificity as to the type of behaviors manifested per domain and to obtain a more nuanced understanding of the presenting behavior.

In selecting a measure to use from these 45, it would also be important to determine the setting and population for which it was validated. For example, a measure developed for hospital use may not be appropriate for the nursing-home setting. Some measures require clinical input to complete although most can be administered by persons of unspecified backgrounds who become familiar with the instrument. Most measures rely on proxy report

and are dependent upon the reporter's memory of behavioral events. Direct or video-captured observations for documenting behavioral occurrences may afford real time ratings that are more objective than patient or proxy report. However, a direct observational approach would appear most feasible in long-term care settings; to date there is a lack of adequate tools and analytic strategies for capturing video-recorded behavioral occurrences in home environments for either clinical or research purposes, although this is an important direction for future research. Additionally, the choice of measure will depend upon the behaviors of most concern in either the clinical or research context. For example, a study on agitation would obviously want to include measures targeting this one domain of behaviors.

The NPI-C is worth highlighting as it stands out as one of the most efficient measures; it involves multiple behavioral domains at a general level as well as targets specific behaviors within domains; it can be used in multiple clinical settings as well as for research purposes; and various scores can be generated based on one's purpose. For example, from the NPI-C, the following scores can be generated: a total score reflecting frequency, a total score reflecting severity of behaviors, a total frequency times severity behavior score, number of behavioral domains affected, number of behavioral items endorsed for any one domain, caregiver level of upset for any one behavioral item, domain and overall. Additionally, the NPI-C is the only measure to be validated in multiple countries, and to yield subscales that have been validated for use as stand-alone measures (e.g. depression, agitation subscales). There are also briefer derivatives (NPI, NPI-Q) that facilitate ease of use in a variety of clinical settings. Finally, the NPI-C integrates both a general and targeted approach: it offers general screening for the presence or absence of a behavioral domain, followed by more detailed questions about specific behaviors within a domain that is endorsed as occurring. Follow-up questions elicit the frequency and severity of specific behaviors representing the behavioral cluster (e.g. agitation).

Noteworthy is that most measures were developed up to 30 years ago. Yet, assessment of behavioral symptoms does not appear to be part of routine care in setting such as primary care in which most persons with dementia receive on-going medical attention. This may reflect in part limited awareness as to the importance of monitoring behavioral symptoms, lack of knowledge that standardized assessment tools exist, belief that clinical judgment is better than validated tools, lack of time during the medical encounter, and concern that if behaviors are detected, there are no evidence-based practices for their management (Murphy *et al.*, 2014). Research to identify current practices and the barriers and facilitators to detecting behaviors using standardized measures and then managing behaviors in different practice settings would be an important endeavor for advancing dementia care in this area.

There are important limitations to this group of measures and which point to the need for significant measurement development in order to enhance clinical practice as it concerns behavioral detection. First, papers reporting on measures do not typically discuss the theoretical underpinnings of item selection and scale construction. Scales have been developed mostly by and within specific disciplines (e.g. neurology, psychiatry, social science) and thus implicitly represent different conceptualizations of behaviors. Item selection has tended to be based primarily upon empirical observations of persons with dementia and there is a lack of a shared framework for defining behaviors, and their

underlying causes and measurement across instrumentation. This is an observation initially reported by Cummings in 1996 and which persists up to today. Second, there are no measures to our knowledge that systematically identify and assess the ecosystem or contextual characteristics in which behaviors occur. Characterizing behaviors independent of context has been the primary focus of existing scales. Existing measures decontextualize behaviors, focusing narrowly on the reporting of the frequency and/or severity of occurrences. This may be due to the implicit assumptions of most measures that behaviors are a direct consequence of neuropathology and as such the brain–behavior relationship dominates the approach to measurement.

However, current thinking is that behavioral symptoms may be a consequence of the confluence of multiple factors including neuropathology as well as medical and physical and social environmental conditions. Neurodegenerative changes may enhance vulnerabilities to stressors in the physical and social environment (Gitlin *et al.*, 2012). Most behavioral symptoms such as agitation or aggression cannot be accounted for alone by neurological impairment. Rather, they appear to reflect unmet needs or a sensitivity to a combination of issues including comorbidities, pain, sleep disturbances, anxiety, fear, underlying infection, and/or factors related to the physical and social environment such as excessive clutter, and complex caregiver communications (Hodgson *et al.*, 2010; 2013). Each of these factors has been implicated in behavioral disturbances.

Currently, measures assume a reactive stance, capturing behavioral events during or after their occurrence. However, attention to potential modifiable risk factors would lead to a preventive care approach in which strategies could be introduced to minimize their contribution to behavioral manifestations. As the goal should be to prevent behavioral symptoms, identifying specific, known risk factors and assessing for those factors on a routine basis would be important. Research has shown for example that aggressive behaviors are associated with pain and caregiver burden (Kunik *et al.*, 2010). As these factors are modifiable, early detection of the presence of these conditions may prevent or minimize the occurrences of aggressive behaviors. Therefore, developing measures that capture known risk factors and the contextual characteristics in which behaviors occur would be important to pursue.

One promising measurement approach involves using smart home technologies or computer assisted video and audiography to capture behaviors in real time. This would enable documentation of antecedent conditions and consequences from which to better characterize the contexts in which behaviors occur in real time. These approaches may lead to measures that systematically capture contextual factors of behavioral occurrences. For example, deriving simple checklists of the presence or absence of key factors identified as contributing to behaviors could be developed and integrated into a measurement approach to more fully understand behavioral occurrences. A related point is that while most measures identified the frequency of behavioral occurrences, only a few evaluated the severity of behaviors to patients or considered caregiver level of upset. None assessed the level of caregiver confidence or self-efficacy in managing behaviors. Assessing proximally related factors such as caregiver reaction and confidence could provide a more holistic understanding of context and direct clinical intervention.

A third measurement consideration is that most scales depend upon a proxy to report observed behaviors. However, it is unclear as to the relationship of proxy report to direct observation and clinician ratings. There is some evidence that family report of behavioral occurrences is influenced by their own mood and distress (de Vugt *et al.*, 2004). There is also some evidence to suggest that both clinician experiences and the institutional culture may affect the choice of intervention used, suggesting that a similar dynamic may be at play when rating the behavior itself (Cohen-Mansfield *et al.*, 2014). Level of knowledge and understanding of behaviors, job satisfaction, and frustrations caring for persons with dementia, and personal values may all influence scoring, particularly for the severity rating of a particular behavior. Thus, research is needed to examine the extent to which behavioral reports by families are congruent with clinician observations and what factors influence ratings by both clinicians and families.

Yet another area of need is for the development of clinical practice guidelines for assessing NPS. A protocol might include regular and on-going brief screening for risk factors and actual occurrences of NPS, followed by in-depth examination of frequency, severity, and caregiver reactions to behavioral occurrences in addition to obtaining a deep description of the biopsychosocial and environmental context in which the behaviors occurred (Gitlin *et al.*, 2012; Kales *et al.*, 2014). Of importance would be for research to evaluate the treatment benefits of integrating routine behavioral detection and assessment in clinical settings.

This review is subject to several limitations. First, only published studies of scales available in English were included. We were unable to determine if identified scales have been translated into other languages; nor were we able to review scales that have been developed in other languages but which are not available in English. There may be scales developed in other languages from which to choose. Second, we were unable to discern the feasibility or practicalities of using any one measure nor the extent to which they are used. Papers did not report the barriers or facilitators of using a particular measure in a specific context. Third, most publications did not indicate whether a measure was sensitive to change nor the impact of its use over multiple testing occasions. Thus, we were unable to draw firm conclusions in this regard. However, attention to the sensitivity of a scale is important as assessments may occur on repeated occasions, and also be used to evaluate behavioral changes due to a treatment. Finally, the science of behavioral symptoms of dementia is evolving. For example, rejection to care behaviors or disorders of personality has recently been identified as dementia-related behaviors distinct from agitation or other behavioral forms. Thus, a limitation of this review may be its scope or focus. That is, as behavioral types are further identified, a review of their measurement should be considered in future reviews.

Beyond the limitations posed by the measures themselves as well as this review, we demonstrate that there are many measures to use for the purposes of assessing NPS in persons with dementia. These measures have strong psychometric properties, and are available to clinicians and researchers alike for use in a variety of settings. From an assessment perspective, attention to behavioral detection is possible and the lack of systematic monitoring of behaviors is not warranted. One hesitation to assess for behaviors by clinicians may be the perception that there is an inadequate evidence-base for treatments. However, this is not the case and there is an emerging evidence base that supports

intervening if behavioral symptoms are present. As to current treatments, pharmacologic agents are typically favored although they have been shown to have limited utility in behavioral management. The most commonly used atypical anti-psychotics may pose more harm than benefit, carry a FDA black box warning for older dementia patients, and tend to be ineffective for behaviors that families and formal providers identify as most problematic such as rejection of care or agitated-type behaviors (Sink *et al.*, 2005; Kales *et al.*, 2007; Gitlin, *et al.*, 2010; Kales *et al.*, 2011). However, a recent clinical trial found citalopram to be effective for agitation in patients with dementia although with cardiac and cognitive adverse effects (Porsteinsson *et al.*, 2014). Noteworthy is that non-pharmacologic treatments (e.g. educating caregivers, identifying and addressing unmet needs, simplifying the physical environment and caregiver communications, establishing structured daily routines or using meaningful activities) have a growing evidence base and are widely endorsed by multiple medical organizations as first-line treatments (Lyketsos *et al.*, 2006; Rabins *et al.*, 2007; Salzman *et al.*, 2008; Brodaty *et al.*, 2012; Gitlin *et al.*, 2012; Kales *et al.*, 2014; Samus *et al.*, 2014).

Thus, given the number and range of available measures, the known harms of behavioral symptoms, and emerging algorithms and evidence-base for treating behavioral symptoms, the ascertainment of behavioral symptoms should be a key feature of comprehensive dementia care in all settings in which persons with dementia reside or are treated.

Acknowledgments

This research was supported by a grant from the National Institute on Aging (R01AG041781; L.N. Gitlin), the Alzheimer's Association (NPSASA-10-174625; L.N. Gitlin), and The Commonwealth Fund (K.S. Van Haitsma). We would like to thank Ann M. Kolanowski, PhD, Hartford Center of Geriatric Nursing Excellence School of Nursing, Penn State, University Park, Pennsylvania, USA for her support and guidance in this research.

References

- Alexopoulos GS, Abrams RC, Young RC, Shamoian CA. Cornell scale for depression in dementia. *Society for Biological Psychiatry*. 1988; 23:271–284.
- Algase DL, Beattie ERA, Bogue EL, Yao L. The Algase wandering scale: initial psychometrics of a new caregiver reporting tool. *American Journal of Alzheimer's Disease and Other Dementias*. 2001; 16:141–152.
- Algase DL, Beattie ERA, Song JA, Milke D, Duffield C, Cowan B. Special section – behavioral symptoms of dementia: their measurement and intervention. Validation of the Algase wandering scale (Version 2) in across cultural sample. *Aging & Mental Health*. 2004; 8:133–142. [PubMed: 14982718]
- Auer SR, Monteiro IM, Reisberg B. The empirical behavioral pathology in Alzheimer's disease (E-BEHAVE-AD) rating scale. *International Psychogeriatrics*. 1996; 8:247–266. [PubMed: 8994895]
- Baumgarten M, Becker R, Gauthier S. Validity and reliability of the dementia behavior disturbance scale. *Journal of the American Geriatrics Society*. 1990; 38:221–26. [PubMed: 2313003]
- Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. *Journal of Consulting and Clinical Psychology*. 1988; 56:893–897. [PubMed: 3204199]
- Belanger HG, Brown LM, Crowell TA, Vanderploeg RD, Curtiss G. The key behaviors change inventory and executive functioning in an elderly clinic sample. *The Clinical Neuropsychologist*. 2002; 16:251–257. [PubMed: 12607138]
- Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the hospital anxiety and depression scale: an updated literature review. *Journal of Psychosomatic Research*. 2002; 52:69–77. [PubMed: 11832252]

- Boddice G, Pachana NA, Byrne GJ. The clinical utility of the geriatric anxiety inventory in older adults with cognitive impairment. *Nursing Older People*. 2008; 20:36–39. [PubMed: 18982897]
- Brodaty H, Arasaratnam C. Meta-analysis of non-pharmacological interventions for neuropsychiatric symptoms of dementia. *American Journal of Psychiatry*. 2012; 169:946–953. [PubMed: 22952073]
- Brodaty H, Draper B, Low LF. Nursing home staff attitudes towards residents with dementia: strain and satisfaction with work. *Journal of Advanced Nursing*. 2003; 44:583–590. [PubMed: 14651681]
- Burgio LD, et al. Studying disruptive vocalization and contextual factors in the nursing home using computer-assisted real-time observation. *Journals of Gerontology: Psychological Sciences*. 1994; 49:P230–P239.
- Burns A, Folstein S, Brandt J, Folstein M. Clinical assessment of irritability, aggression, and apathy in Huntington and Alzheimer disease. *The Journal of Nervous and Mental Disease*. 1990; 178:20–26. [PubMed: 2136908]
- Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Research*. 1989; 28:193–213. [PubMed: 2748771]
- Byrne GJ, Pachana NA. Development and validation of a short form of the geriatric anxiety inventory—the GAI-SF. *International Psychogeriatrics*. 2011; 23:125–131. [PubMed: 20561386]
- Chan DC, Kasper JD, Black BS, Rabins PV. Presence of behavioral and psychological symptoms predicts nursing home placement in community-dwelling elders with cognitive impairment in univariate but not multivariate analysis. *Journals of Gerontology: Medical Sciences*. 2003; 58A: 548–554.
- Clarke DE, van Reekum R, Simard M, Streiner DL, Freedman M, Conn D. Apathy in dementia and examination of the psychometric properties of the apathy evaluation scale. *Journal of Neuropsychiatry and Clinical Neuroscience*. 2007; 19:57–64.
- Cohen-Mansfield J. Instruction Manual for the Cohen-Mansfield Agitation Inventory (CMAI). 1991. Available at: http://www.dementia-assessment.com.au/symptoms/cmai_manual.pdf
- Cohen-Mansfield J, Golander H, Arnheim G, Cohen R. Reactions and interventions for delusions in nursing home residents with dementia. *American Journal of Alzheimer's Disease and Other Dementias*. 2014; 10.1177/1533317514522850.
- Cohen-Mansfield J, Marx M, Rosenthal A. A description of agitation in a nursing home. *Journals of Gerontology*. 1989; 44:M77–M84. [PubMed: 2715584]
- Conn D, Thorpe L. Assessment of behavioural and psychological symptoms associated with dementia. *The Canadian Journal of Neurological Sciences*. 2007; 34:S67–S71. [PubMed: 17469686]
- Cummings JL. Theories behind existing scales for rating behavior in dementia. *International Psychogeriatrics*. 1996; 8(Suppl. 3):293–300. [PubMed: 9154578]
- Cummings JL, et al. The neuropsychiatric inventory: comprehensive assessment of psychopathology in dementia. *Neurology*. 1994; 44:2308–2314. [PubMed: 7991117]
- de Medeiros K, et al. The Neuropsychiatric Inventory – Clinician rating scale (NPI-C): reliability and validity of a revised assessment of neuropsychiatric symptoms in dementia. *International Psychogeriatrics*. 2010; 22:984–994. [PubMed: 20594384]
- Devanand DP, et al. Behavioral syndromes in Alzheimer's disease. *International Psychogeriatrics*. 1992; 4:161–184. [PubMed: 1288661]
- de Vugt ME, et al. Do caregiver management strategies influence patient behaviour in dementia? *International Journal of Geriatric Psychiatry*. 2004; 19:85–92. [PubMed: 14716704]
- Dujardin K, Sockeel P, Delliaux M, Destee A, Defebvre L. The Lille Apathy Rating scale: validation of a caregiver-based version. *Movement Disorders*. 2008; 23:845–849. [PubMed: 18307266]
- Fillenbaum GG, Peterson B, Morris JC. Estimating the validity of the Clinical Dementia Rating scale: the CERAD experience. *Aging Clinical and Experimental Research*. 1996; 8:379–385.
- Finkel S, Lyons JS, Anderson RL. Reliability and validity of the Cohen-Mansfield Agitation Inventory in institutionalized elderly. *International Journal of Geriatric Psychiatry*. 1992; 7:487–490.
- Finkel SI, Lyons JS, Anderson RL. A Brief Agitation Rating Scale (BARS) for nursing home elderly. *Journal of the American Geriatrics Society*. 1993; 41:50–52. [PubMed: 8418123]

- Flint AJ, Rifat S. Factor structure of the Hospital Anxiety and Depression Scale in older adults with major depression. *International Journal of Geriatric Psychiatry*. 2002; 17:117–123. [PubMed: 11813272]
- Fydrich T, Dowdall D, Chambless DL. Reliability and validity of the Beck Anxiety Inventory. *Journal of Anxiety Disorders*. 1992; 6:55–61.
- Gitlin LN, Kales HC, Lyketsos CG. Nonpharmacologic management of behavioral symptoms in dementia. *Journal of the American Medical Association*. 2012; 308:2020–2029. [PubMed: 23168825]
- Gitlin LN, Winter L, Dennis MP, Hauck WW. A non-pharmacological intervention to manage behavioral and psychological symptoms of dementia and reduce caregiver distress: design and methods of project ACT. *Clinical Interventions in Aging*. 2007; 2:695–703. [PubMed: 18225471]
- Gitlin LN, Winter L, Dennis MP, Hodgson N, Hauck WW. Targeting and managing behavioral symptoms in individuals with dementia: a randomized trial of a nonpharmacological intervention. *Journal of the American Geriatrics Society*. 2010; 58:1465–1474. [PubMed: 20662955]
- Grace J, Stout JC, Malloy PF. Assessing frontal lobe behavioral syndromes with the Frontal Lobe Personality Scale. *Assessment*. 1999; 6:269–284. [PubMed: 10445964]
- Greene JG, Smith R, Gardiner M, Timbury GC. Measuring behavioral disturbance of elderly demented patients in the community and its effects on relatives: a factor analytic study. *Age and Aging*. 1982; 11:121–126.
- Helmes E, Csapo KG, Short JA. Standardization and validation of the Multidimensional Observation Scale for Elderly Subjects (MOSES). *Journals of Gerontology*. 1987; 42:395–405. [PubMed: 3598087]
- Hodgson N, Gitlin LN, Winter L, Czekanski K. Undiagnosed illness and neuropsychiatric behaviors in community-residing older adults with dementia. *Alzheimer's Disease and Associated Disorders*. 2010; 24:603–609.
- Hodgson N, Gitlin LN, Winter L, Hauck W. Caregiver's perceptions of the relationship of pain to behavioral and psychiatric symptoms in older community residing adults with dementia. *Clinical Journal of Pain*. 2013; 30:421–427. [PubMed: 24281271]
- Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *The British Journal of Psychiatry*. 1982; 140:566–572. [PubMed: 7104545]
- Hurley AC, et al. Measurement of observed agitation in patients with dementia of the Alzheimer type. *Journal of Mental Health and Aging*. 1999; 5:117–132.
- Jeon YH, et al. Recommended measures for the assessment of behavioral disturbances associated with dementia. *American Journal of Geriatric Psychiatry*. 2011; 19:403–415.
- Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep*. 1991; 14:540–545. [PubMed: 1798888]
- Johns MW. Reliability and factor analysis of the Epworth sleepiness scale. *Sleep*. 1992; 15:376–381. [PubMed: 1519015]
- Kales HC, Gitlin LN, Lyketsos CG. Management of neuropsychiatric symptoms of dementia in clinical settings: recommendations from a multidisciplinary expert panel. *Journal of the American Geriatrics Society*. 2014; 62:762–769. [PubMed: 24635665]
- Kales HC, et al. Mortality risk in patients with dementia treated with antipsychotics versus other psychiatric medications. *American Journal of Psychiatry*. 2007; 164:1568–1576. [PubMed: 17898349]
- Kales HC, et al. Trends in antipsychotic use in dementia 1999–2007. *Archives of General Psychiatry*. 2011; 68:190–197. [PubMed: 21300946]
- Karttunen K, et al. Neuropsychiatric symptoms and quality of life in patients with very mild and mild Alzheimer's disease. *International Journal of Geriatric Psychiatry*. 2011; 26:473–482. [PubMed: 21445998]
- Kaufer DI, et al. Validation of the NPI-Q, a brief clinical form of the neuropsychiatric inventory. *The Journal of Neuropsychiatry and Clinical Neurosciences*. 2000; 12:233–239. [PubMed: 11001602]
- Kolitz BP, Vanderploeg RD, Curtiss G. Development of the key behaviors change inventory: a traumatic brain injury behavioral outcome assessment instrument. *Archives of Physical Medicine and Rehabilitation*. 2003; 84:277–284. [PubMed: 12601661]

- Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*. 2001; 16:606–613. [PubMed: 11556941]
- Kunik ME, et al. Causes of aggressive behavior in patients with dementia. *Journal of Clinical Psychiatry*. 2010; 71:1145–1152. [PubMed: 20361896]
- LaBarge E. A preliminary scale to measure the degree of worry among mildly demented Alzheimer's disease patients. *Physical and Occupational Therapy in Geriatrics*. 1993; 11:43–57.
- Levin HS, et al. The neurobehavioral rating Scale: Assessment of the behavioral sequelae of head injury by the clinician. *Journal of Neurology, Neurosurgery, and Psychiatry*. 1987; 50:183–193.
- Logsdon RG, et al. Assessment of agitation in Alzheimer's disease: the agitated behavior in dementia scale. Alzheimer's disease cooperative study. *Journal of the American Geriatrics Society*. 1999; 47:1354–1358. [PubMed: 10573447]
- Loreck D, Bylsma F, Folstein M. The Dementia Signs and Symptoms Scale: a new scale for comprehensive assessment of psychopathology in Alzheimer's disease. *American Journal of Geriatric Psychiatry*. 1994; 2:52–59. [PubMed: 21629007]
- Lyketsos CG, et al. Position statement of the American Association for Geriatric Psychiatry regarding principles of care for patients with dementia resulting from Alzheimer disease. *American Journal of Geriatric Psychiatry*. 2006; 14:561–573. [PubMed: 16816009]
- Lyketsos CG, et al. Neuropsychiatric symptoms in Alzheimer's disease. *Alzheimer's & Dementia*. 2011; 7:532–539.
- Marin RS. Apathy: a neuropsychiatric syndrome. *The Journal of Neuropsychiatry and Clinical Neurosciences*. 1991; 3:243–254. [PubMed: 1821241]
- Marin RS, Biedrzycki RC, Firinciogullari S. Reliability and validity of the Apathy Evaluation Scale. *Psychiatry Research*. 1991; 38:143–162. [PubMed: 1754629]
- McShane R. What are the syndromes of behavioral and psychological symptoms of dementia? *International Psychogeriatrics*. 2000; 12:147–153.
- Miyamoto Y, Tachimori H, Ito H. Formal caregiver burden in dementia: impact of behavioral and psychological symptoms of dementia and activities of daily living. *Geriatric Nursing*. 2010; 31:246–253. [PubMed: 20682402]
- Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *International Psychogeriatrics*. 1997; 9:173–176. [PubMed: 9447441]
- Mungas D, Weiler P, Franzi C, Henry R. Assessment of disruptive behavior associated with dementia: the disruptive behavior rating scales. *Journal of Geriatric Psychiatry and Neurology*. 1989; 2:196–202. [PubMed: 2635016]
- Murman DL, Chen Q, Powell MC, Kuo SB, Bradley CJ, Colenda CC. The incremental direct costs associated with behavioral symptoms in AD. *Neurology*. 2002; 59:1721–1729. [PubMed: 12473759]
- Murphy K, et al. Understanding diagnosis and management of dementia and guideline implementation in general practice: a qualitative study using the theoretical domains framework. *Implementation Science*. 2014; 9:31. [PubMed: 24581339]
- Neville CC, Byrne GJ. Behaviour rating scales for older people with dementia: which is the best for use by nurses? *Australasian Journal on Ageing*. 2001; 20:166–172.
- Odenheimer G, et al. Quality improvement in neurology: dementia management quality measures (executive summary). *Neurology*. 2013; 81:1545–1549. [PubMed: 24068786]
- Okura T, Langa KM. Caregiver burden and neuropsychiatric symptoms in older adults with cognitive impairment: the Aging, Demographics, and Memory Study (ADAMS). *Alzheimer Disease and Associated Disorders*. 2010; 25:116–121. [PubMed: 21192239]
- Osman A, Kopper BA, Barrios FX, Osman JR, Wade T. The Beck anxiety inventory: reexamination of factor structure and psychometric properties. *Journal of Clinical Psychology*. 1998; 53:7–14. [PubMed: 9120035]
- Pachana NA, Byrne GJ, Siddle H, Koloski N, Harley E, Arnold E. Development and validation of the geriatric anxiety inventory. *International Psychogeriatrics*. 2007; 19:103–114. [PubMed: 16805925]
- Patel V, Hope RA. A rating scale for aggressive behavior in the elderly: the RAGE. *Psychological Medicine*. 1992; 22:211–221. [PubMed: 1574558]

- Perlman CM, Hirdes JP. The aggressive behavior scale: a new scale to measure aggression based on the minimum data set. *Journal of the American Geriatrics Society*. 2008; 56:2298–2303. [PubMed: 19093929]
- Porsteinsson AP, et al. Effect of citalopram on agitation in Alzheimer disease: the CitAD randomized clinical trial. *JAMA*. 2014; 311:682–691. [PubMed: 24549548]
- Rabins PV, et al. APA work group on Alzheimer's disease and other dementias. American Psychiatric Association practice guideline for the treatment of patients with Alzheimer's disease and other dementias. *American Journal of Psychiatry*. 2007; 164:5–56. [PubMed: 18340692]
- Rabins PV, et al. Predictors of progression to severe Alzheimer's disease in an incidence sample. *Alzheimer's & Dementia*. 2013; 9:204–207.
- Ray WA, Taylor JA, Lichtenstein MJ, Meador KG. The nursing home behavior problem scale. *Journals of Gerontology: Medical Sciences*. 1992; 47:M9–M16.
- Reisberg B, Auer SR, Monteiro IM. Behavioral pathology in Alzheimer's disease (BEHAVE-AD) rating scale. *International Psychogeriatrics*. 1997; 8:301–308. [PubMed: 9154579]
- Robert PH, et al. The apathy inventory: assessment of apathy and awareness in Alzheimer's disease, Parkinson's disease and mild cognitive impairment. *International Journal of Geriatric Psychiatry*. 2002; 17:1099–1105. [PubMed: 12461757]
- Rosen J, et al. The Pittsburgh Agitation Scale: a user-friendly instrument for rating agitation in dementia patients. *The American Journal of Geriatric Psychiatry*. 1994; 2:52–59. [PubMed: 21629007]
- Ryden MB. Aggressive behavior in persons with dementia who live in the community. *Alzheimer's Disease and Associated Disorders*. 1988; 2:342–355.
- Ryden MB, Bossenmaier M, McLachlan C. Aggressive behavior in cognitively impaired nursing home residents. *Research in Nursing & Health*. 1991; 14:87–95. [PubMed: 2047539]
- Salzman C, et al. Elderly patients with dementia-related symptoms of severe agitation and aggression: consensus statement on treatment options, clinical trials methodology, and policy. *The Journal of Clinical Psychiatry*. 2008; 69:889–898. [PubMed: 18494535]
- Samus QM, et al. A multidimensional home-based care coordination intervention for elders with memory disorders: the maximizing independence at home (MIND) pilot randomized trial. *American Association for Geriatric Psychiatry*. 2014; 22:398–414.
- Shah A, Evans H, Parkash N. Evaluation of three aggression/agitation behavior rating scales for use on an acute admission and assessment psychogeriatric ward. *International Journal of Geriatric Psychiatry*. 1998; 13:415–420. [PubMed: 9658278]
- Shankar KK, Walker M, Frost D, Orrell MW. The development of a valid and reliable scale for rating anxiety in dementia (RAID). *Aging & Mental Health*. 1999; 3:39–49.
- Sink KM, Holden KF, Yaffe K. Pharmacological treatment of neuropsychiatric symptoms of dementia: a review of the evidence. *Journal of the American Medical Association*. 2005; 293:596–608. [PubMed: 15687315]
- Sockeel P, Dujardin K, Devos D, Deneve C, Destee A, Defebvre L. The Lille Apathy Rating Scale (LARS), a new instrument for detecting and quantifying apathy: validation in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*. 2006; 77:579–584.
- Steer, RA.; Beck, AT. Beck Anxiety Inventory. In: Zalaquett, CP.; Wood, RJ., editors. *Evaluating Stress: A Book of Resources*. Lanham, MD: Scarecrow Education; 1997.
- Strauss ME, Sperry SD. An informant-based assessment of apathy in Alzheimer disease. *Neuropsychiatry, Neuropsychology, & Behavioral Neurology*. 2002; 15:176–183.
- Sultzer DL, Berisford A, Gunay I. The neurobehavioral rating scale: reliability in patients with dementia. *Journal of Psychiatry*. 1995; 29:185–191.
- Sunderland T, et al. A new scale for the assessment of depressed mood in dementia patients. *The American Journal of Psychiatry*. 1988; 145:955–959. [PubMed: 3394879]
- Tariot PN, et al. The behavior rating scale for dementia of the consortium to establish a registry for Alzheimer's disease. *The American Journal of Psychiatry*. 1995; 152:1349–1357. [PubMed: 7653692]

- Teri L, Truax P, Logsdon R, Uomoto J, Zarit S, Vitaliano PP. Assessment of behavioral problems in dementia: the revised memory and behavior problems checklist. *Psychology and Aging*. 1992; 7:622–631. [PubMed: 1466831]
- Tractenberg RE, Singer CM, Cummings JL, Thai LJ. The sleep disorders inventory: an instrument for studies of sleep disturbance in persons with Alzheimer's disease. *Journal of Sleep Research*. 2003; 12:331–337. [PubMed: 14633245]
- Volicer L, Bass EA, Luther SL. Agitation and resistiveness to care are two separate behavioral syndromes of dementia. *Journal of the American Medical Directors Association*. 2007; 8:527–532. [PubMed: 17931577]
- Volicer L, van der Steen JT, Frijters DH. Involvement in activities and wandering in nursing home residents with cognitive impairment. *Alzheimer's Disease and Associated Disorders*. 2013; 27:272–277.
- Wahle M, Häller S, Spiegel R. Validation of the NOSGER (Nurses' Observation Scale for Geriatric Patients). *International Psychogeriatrics*. 1996; 8:525–547. [PubMed: 9147168]
- Wetherell JL, Gatz M. The Beck anxiety inventory in older adults with generalized anxiety disorder. *Journal of Psychopathology and Behavioral Assessment*. 2005; 27:17–24.
- Weyer G, Erzigkeit H, Kanowski S, Ihl R, Hadler D. Alzheimer's disease assessment scale: reliability and validity in a multicenter trial. *International Psychogeriatrics*. 1997; 9:123–138. [PubMed: 9309486]
- Yesavage JA, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of Psychiatric Research*. 1983; 17:37–49. [PubMed: 7183759]
- Yudofsky SC, Kopecky HJ, Silver JM, Endicott J. The Overt Agitation Severity Scale for the objective rating of agitation. *Journal of Neuropsychiatry*. 1997; 9:541–548.
- Yudofsky SC, Silver JM, Jackson W, Endicott J, Williams D. The Overt Aggression Scale for the objective rating of verbal and physical aggression. *The American Journal of Psychiatry*. 1986; 143:35–39. [PubMed: 3942284]
- Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*. 1983; 67:361–370. [PubMed: 6880820]

Figure 1.
Search flowchart. PwD = Person with dementia.

Table 1
Summary of general measures of neuropsychiatric symptoms for persons with dementia ($n = 16$)

Assessment Tool	Behavioral Domains ^a	Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
			Who ^a	How ^b	To whom ^c				
1 Alzheimer's disease assessment scale non-cog* (Weyer <i>et al.</i> , 1997)	$N = 10$ Tremors Pacing Motor restlessness Tearfulness Depression Delusions Hallucinations Appetite Concentration Uncooperativeness	10 items	TI	I	D, P	Based on past week Rated 0 = not present to 5 = severe Range 0–50 Higher scores indicate greater behavioral issues	Not specified	AD patients in community and nursing homes	Reliability: Internal consistency: $\alpha = 0.83$ Test-retest: $r = 0.977$ Validity: Convergent validity: significant correlations with NOSGER Patients Mood subscale ($r = 0.69$), social behavior ($r = 0.69$), and disturbing behavior ($r = 0.51$; $p < 0.05$) Sensitivity to change: Indicated
2 Behavioral pathology in Alzheimer's disease* (BEHAVE-AD; Reisberg <i>et al.</i> , 1997) Also available the E-BEHAVE-AD (Auer <i>et al.</i> , 1996)	$N = 7$ Delusions Hallucinations Activity disturbances Aggressiveness Diurnal rhythm disturbances Affective disturbance Anxiety/phobia	26 items	TI TO	I	P	Based on past two weeks Rated 0 = not present to 3 (each category 3 is different) Range 0–75 (only first 25 items totaled) Higher scores indicate greater behavioral issues	20 minutes	AD patients; outpatient and nursing home patients	Reliability: Inter-rater: intraclass correlation coefficient $r = 0.96$ ($p < 0.01$) Validity: Construct validity: supported by the differences between the nature and course of behavioral symptoms of AD and those of the cognitive and functional symptoms; Sensitivity to change: Not indicated
3 Behavioral syndromes scale for dementia* (BSSD; Devanand <i>et al.</i> , 1992)	$N = 5$ Disinhibition (including agitation, aggression, and wandering) Catastrophic reactions Apathy/indifference Sundowning Denial	24 items	TI	I	P	Based on past week Rated 0 = no information to 6 = extreme Denial rated (0–4) Global rating for each domain Higher scores indicate greater behavioral issues	20–30 minutes	Probable AD in outpatient setting	Reliability: Internal consistency: Catastrophic reactions: $\alpha = 0.69$ – 0.78 Disinhibition: $\alpha = 0.73$ – 0.82 Apathy/indifference: $\alpha = 0.82$ – 0.83 Sundowning: $\alpha = 0.70$ – 0.76 Inter-rater: intraclass correlation coefficients for the five domains: Catastrophic reactions: 0.64 – 0.85

Assessment Tool	Behavioral Domains ^a	Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties	
			Who ^a	How ^b	To whom ^c					
4	CERAD behavior rating scale for dementia * (BRSD; Tariot <i>et al.</i> , 1995)	N = 8 Depressive features Psychotic features Defective self-regulation Irritability/agitation Vegetative features Apathy Aggression Affective lability	Original 51 items (48 quantitative and three open-ended) Revised 46 items (three quantitative items dropped and three open-ended items consolidated into one question) 17-item shortened version (unable to locate published reliability and validity for the short form)	TI	I	P	Based on past month five items (diurnal patterns of confusion and changes in interest, appetite, weight, and sexual drive) scored as present or absent. Other items rated 0 = has not occurred since illness began to 4 = present 16 days or more in the past month, more than half the days in the month Higher scores indicate greater behavioral issues Scoring available for “has occurred since illness began but not in past month”	20–30 minutes	Dementia patients in various settings	Disinhibition: 0.83–0.90 Apathy indifference: 0.65–0.85 Sundowning: 0.53–0.95 Denial: 0.40–0.84 Validity: Divergent validity: demonstrated by weak to moderate correlations between domains Criterion validity: demonstrated in several ways including an association with mMMSE score Sensitivity to change: Indicated Reliability: Inter-rater: ranged from 91.3% to 100%. Item <i>K</i> s ranged from 0.77–1.00. Validity: Not reported Sensitivity to change: Indicated
5	Clinical Dementia Rating scale * (CDR; Hughes <i>et al.</i> , 1982; Fillenbaum <i>et al.</i> , 1996; Morris, 1997)	N = 6 Memory Orientation Judgment and problem solving community affairs Home and hobbies Personal Care	48 items for informant 27 items for m person with AD	TI	I	D, P	Rating based on trained interviewers’ judgment based on semi-structured interview of caregiver and person with AD Each domain rated 0 = none to 3 = severe	40 minutes	AD patients in the community	Reliability: Inter-rater: 83% Validity: Criterion validity: correlations with neuropsychological measures for both global and individual scores Neuropathological validity: detecting the

Assessment Tool	Behavioral Domains ^a	Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
			Who ^a	How ^b	To whom ^c				
6	Computer-assisted Behavioral Observation Systems* (CABOS; Burgio <i>et al.</i> , 1994)	12 h of observation per patient (four 3-h blocks)	TO	O	D	Location Activity in environment Sound in environment Social environment Physical restraint	12 h per patient	Nursing home patients with probable dementia	presence or absence dementia Sensitivity to change: Indicated Reliability: Inter-observer: κ 's ranged from 1.0 (location – hairdresser) to 0.67 (activity – transfer) Validity: Not reported Sensitivity to change: Not indicated
7	Dementia behavior disturbance scale* (DBD; Baumgarten <i>et al.</i> , 1990)	28 items	TI, F	I	P	Based on prior week Rated 0 = never to 4 = all the time Range 0–112 Higher scores indicate more disturbance	15 minutes	Dementia patients living in the community	Reliability: Internal consistency: $\alpha = 0.83–0.84$ Test-retest: $r = 0.71$ Validity: Construct validity: correlation with Greene's Behavioural and Mood Disturbance Scale, $r = 0.73$ (Green <i>et al.</i> , 1982) Sensitivity to change: Not indicated
8	Dementia Signs and Symptoms scale* (DSS; Loreck <i>et al.</i> , 1994)	43 items	TI	I	D, P	Rating based on occurrence and severity in past month Rated 0 = absent to 3 = daily Higher scores indicate greater behavioral issues	30 minutes	AD patients in clinical settings	Reliability: Internal consistency: α ranged from 0.37 for hallucinations to 0.82 for behaviors. Average internal consistency was 0.60. Inter-rater: 0.92–0.99 Validity: Concurrent validity: significant correlations between subscales of DSS and established measures of constructs, ranging from the depression subscale of DSS and the CUSPAD ($r = 0.49$, $p = 0.001$) and the Mania subscale of DSS and the Young Mania scale ($r = 0.94$, $p < 0.001$). Sensitivity to change: Indicated

Assessment Tool	Behavioral Domains ^a	Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
			Who ^a	How ^b	To whom ^c				
9 Frontal System Behavior scale * (FISBe; Grace <i>et al.</i> , 1999) (formerly the Frontal Lobe Personality Scale)	N = 3 Apathy Disinhibition Executive dysfunction	46 items	F	I	D, P	Rating based on pre-illness and current, or just current. Frequency rated 1 = almost never to 5 = almost always, reversed for positive items Sub-scores and total score (range 46–230) Higher score equals more behavioral abnormality.	10 minutes to administer; 10–15 minutes to score	Outpatients with damage to the frontal lobe, TBI, AD, and PD	Reliability: Internal consistency: $\alpha = 0.93-0.95$ Validity: Construct validity: family ratings of patient pre-morbid behavior and post-illness/injury frontal were not highly correlated ($r = 0.30$; $p = 0.16$) and pre and post scores were significantly different ($t(22) = -6.21$, $p < 0.001$). Sensitivity to change: Indicated
10 Key behavior change inventory * (KBCI; Belanger <i>et al.</i> , 2002; Kolitz <i>et al.</i> , 2003)	N = 8 Inattention Impulsivity Unawareness of problems Apathy Interpersonal difficulties Communication problems Somatic difficulties Emotional adjustment	64 items	F	I	P	Rating period not stated. 4-point scale (false not true to very true) Half of items are worded positively, half negatively. Range of scores not available. Greater score equals greater impairment.	Not specified	Traumatic brain injury and mild dementia in clinics	Reliability: Internal consistency: $\alpha = 0.82-0.91$ Validity: Content validity: external item review by panel of experts Construct validity: significant group differences between controls and those with TBI ($F(16,178) = 9.15$, $p < 0.001$). Convergent validity: “Convergent validity was demonstrated by significant correlations between the KBCI scales hypothesized to relate to executive functioning and at least one cognitive executive measure, with the exception of the KBCI Impulsivity scale.” Discriminant validity: non-significant correlations with measures of language (modified Boston Naming Test (BNT)), visuospatial abilities (Judgment of Line Orientation), memory (overall recognition hit rate for the CERAD list

Assessment Tool	Behavioral Domains ^a	Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
			Who ^a	How ^b	To whom ^c				
11	Multi-dimensional observation scale for elderly patients* (MOSES; Helmes <i>et al.</i> 1987)	N = 5 Self-care Disoriented behavior Depressed/anxious mood Irritable behavior Withdrawn behavior	40 items	N	O	D	Based on past week Rated either on 1–4 or 1–5 scale with different response sets for each item Range 40–180 Higher scores indicate greater behavioral issues.	Not specified Older adults in psychiatric facilities, nursing homes, homes for the aged, and continuing care hospitals – patients with dementia were not excluded but specific psychometric on this population are not given.	learning task), and global cognitive functioning learning task), and global cognitive functioning learning task), and global cognitive functioning Sensitivity to change: Indicated Reliability: Internal consistency: Self-care: $\alpha = 0.82$ Disorientation: $\alpha = 0.87$ Depression: $\alpha = 0.80$ Irritability: $\alpha = 0.79$ Withdrawn: $\alpha = 0.78$ Inter-rater: Self-care: $r = 0.97$ Disorientation: $r = 0.84$ Depression: $r = 0.58$ Irritability: $r = 0.72$ Withdrawn: $r = 0.75$ Validity: Convergent validity: correlation with Physical and Mental Impairment-of-functioning Evaluation (PAMIE) subscales significant at $p < 0.001$. Self-care: $r = 0.91$ Disorientation: $r = 0.81$ Depression: $r = 0.65$ Irritability: $r = 0.77$ Withdrawn: $r = 0.78$ Convergent validity: depression correlated with Zung Depression Status Inventory: $r = 0.49$, $p < 0.005$; self-care correlated with Robertson Short Mental Status Questionnaire: $r = 0.53$, $p < 0.001$; Disorientation with Robertson Short Mental Status Questionnaire: $r = 0.77$, $p < 0.001$ Sensitivity to change: Indicated Reliability: Inter-rater: $r = 0.93$, $p < 0.001$ Validity: Not reported Sensitivity to change:
12	The Neurobehavioral Rating scale** (NRS; Levin <i>et</i>	N = 6 Cognition Agitation/disinhibition Behavioral retardation Anxiety/Depression	27 items	TI	I	D	Rating period not stated Scored 0 = not present to 6 = extremely severe Range 0–162	Patients with head trauma, HIV infection, and dementia	learning task), and global cognitive functioning learning task), and global cognitive functioning learning task), and global cognitive functioning Sensitivity to change: Indicated Reliability: Inter-rater: $r = 0.93$, $p < 0.001$ Validity: Not reported Sensitivity to change:

Assessment Tool	Behavioral Domains ^a	Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
			Who ^a	How ^b	To whom ^c				
<i>al.</i> , 1987; Sultzer <i>et al.</i> , 1995)	Verbal output disturbance Psychosis					Higher scores indicate greater behavioral issues			Indicated
13 Neuropsychiatric inventory* (NPI; Cummings <i>et al.</i> , 1994) Also available are the NPI-C and the NPI-Q.	N = 12 Delusions Hallucinations Dysphoria Anxiety Agitation Euphoria Apathy Irritability Disinhibition Aberrant motor behavior Nighttime behavior disturbances Changes in appetite and eating behaviors (the last two were additions to the original) Caregiver distress	12–91 Varies based on domain screening questions Each domain has a screening (Y/N response) r question. If yes, subsequent questions in that domain (7–9 items per domain) are asked and assessed for frequency, severity, and caregiver distress.	TI	I	P	Based on past month Yes/No to behavior present Frequency rated 1 = occasionally, less than once per week to 4 = very frequently, once or more per day or continuously Severity rated 1 = mild, produces little stress in patient to 3 = marked, a major source of behavioral abnormality Caregiver distress rated 0 = not distressing to 5 = extreme distress Total score for each domain calculated by multiplying frequency by severity. Add domain totals for total NPI score. Higher scores indicate greater behavioral issues.	10 minutes but depends on number of r behaviors present	Dementia patients, no specific setting stated	Reliability: Inter-rater: varied from 93.6% to 100% Test-retest: 0.79 ($p < 0.01$) for frequency and 0.86 ($p < 0.01$) for severity at three weeks Validity: Content validity: a Delphi panel to review the behaviors of apathy, irritability, disinhibition, and euphoria as there was no “gold standard.” Concurrent validity: scores on relevant scales were compared to the BEHAVE-AD and HAM-D. All correlations reached the 0.05 level of significance and all but one reached the 0.01 level of significance. Sensitivity to change: Indicated
NPI-C* (de Medeiros <i>et al.</i> , 2010)	N = 14 Delusions Hallucinations Agitation Aggression Dysphoria Anxiety Euphoria Apathy Disinhibition Irritability Aberrant motor behavior Nighttime behavior disturbances Changes in appetite and eating behaviors Aberrant vocalizations Caregiver distress	14–142 Varies based on domain screening questions Each domain has a screening (Y/N response) question. If yes, subsequent question in that domain (7–16 items per domain) are asked of the caregiver for frequency, severity, and caregiver distress.	C	I	P, D, C	Based on past month Yes/no to behavior present Frequency rated 1 = occasionally, less than once per week to 4 = very frequently, once or more per day or continuously Severity rated 1 = mild, produces little stress in patient to 3 = marked, a major source of behavioral abnormality Caregiver distress rated 0 = not distressing to 5 = extreme distress	Not specified	Dementia patients, no specific setting stated	Reliability: Inter-rater: majority of items had intraclass correlations (ICCs) ranging from 0.70–0.96 (some items had lower ICCs, particularly in the hallucinations domain). Validity: Convergent validity: Pearson correlations of NPI-C constructs to established scales ranging from 0.31 (apathy) to 0.61 (depression) Sensitivity to change: Indicated.

Assessment Tool	Behavioral Domains ^a	Administration				Estimated Time to Administer	Setting and Target Population	Psychometric Properties
		Number of Items	Who ^a	How ^b	To whom ^c			
NPI-Q* (Kaufert et al., 2000)	N = 12 Delusions Hallucinations Dysphoria Anxiety Agitation Euphoria Apathy Irritability Disinhibition Aberrant motor behavior Nighttime behavior disturbances Changes in appetite and eating behaviors (the last two were additions to the original) Caregiver distress	12 screening questions from NPI, severity, and caregiver distress	F	I	P	Not specified	Dementia patients, no specific setting stated	Reliability: Test-retest: correlations were 0.80 (total symptoms, $p < 0.0001$) and 0.94 (distress scores, $p < 0.0001$). Validity: Convergent validity: NPI-Q and NPI Sensitivity to change: Indicated
14 Nurses' observation scale for geriatric patients* (NOSGER; Wahle et al., 1996)	N = 6 Memory IADLs ADLs Mood Social behavior Disturbing behavior	30 items	N, F	O	D	Not specified	Older adults at home or in an institution (healthy, mild dementia, and advanced dementia)	Reliability: Inter-rater: Memory: $r = 0.85$ IADL: $r = 0.89$ ADL: $r = 0.88$ Mood: $r = 0.76$ Social behavior: $r = 0.68$ Disturbing behavior: $r = 0.70$ ($p < 0.001$ for all subscales) Test-retest: Memory: $r = 0.91$

Assessment Tool	Behavioral Domains ^a	Administration			Number of Items	Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
		Who ^a	How ^b	To whom ^c					
15	The nursing home behavior problem scale* (NHBPS; Ray <i>et al.</i> , 1992)	N	O	D	29 items	Based on past three days Rated 0 = never to 4 = always Range 0–116 Higher scores indicate greater behavioral issues.	3–5 minutes per resident	Nursing home residents	LADL: $r = 0.92$ ADL: $r = 0.88$ Mood: $r = 0.85$ Social behavior: $r = 0.87$ Disturbing behavior: $r = 0.84$ ($p < 0.001$ for all subscales) Validity: Concurrent validity: Memory compared measures of cognition (digit span forward and backward, trail-making; $r = 0.43–0.70$, $p < 0.001$) LADL compared with ADL and PLUT ($r = 0.60–0.68$, $p < 0.001$) ADL: compared with LADL and PLUT3 ($r = 0.73–0.80$, $p < 0.001$) Social behavior compared with PLUT ($r = 0.74$, $p < 0.001$) Not done for mood or disturbing behavior Sensitivity to change: Indicated Reliability: Inter-rater: $r = 0.75–0.83$ Validity: Convergent Validity: Correlation with NOSIE: $r = -0.747$ Correlation with CMAI: $r = 0.911$ Sensitivity to change: Indicated
16	Revised memory and behavior problem checklist* (RMBPC; Teri <i>et al.</i> , 1992)	F	I	P	24 items	Based on past week Behaviors rated on frequency: 0 = never occurs to 4 = occurs daily or more often Range 0–96 Higher scores indicate greater frequency of behavioral issues. Caregiver reaction rated by degree behavior is upsetting/bothersome	15–20 minutes	Dementia patients in outpatient clinic	Reliability: Internal consistency: patient behavior frequency (overall): $\alpha = 0.84$ Caregiver reaction (overall): $\alpha = 0.90$ Validity: Concurrent validity: significant correlations between Depression subscale of RMBC and the Hamilton Depression Rating Scale (HDRS; $r =$

Assessment Tool	Behavioral Domains ^a	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
		Who ^a	How ^b	To whom ^c				
					0 = not at all to 4 = extremely Range 0–96 Higher scores indicate greater distress.			0.44, $p < 0.01$) and a diagnosis of major depression ($r = 0.36$, $p < 0.01$). Disruption subscale of RMBC also correlated with the HDRS ($r = 0.19$, $p < 0.05$). Convergent validity: significant correlations between the Memory-Related Problems subscale of RMBC and MMSE score ($r = -0.48$, $p < 0.01$) and diagnosis of dementia ($r = 0.45$, $p < 0.01$). Discriminant validity: no significant correlations between Depression and Disruption subscales of RMBC and MMSE score ($r = -0.04$ and $r = -0.09$, respectively; $p = ns$) or diagnosis of dementia ($r = 0.05$ and $r = 0.10$, respectively; $p = ns$). No significant correlations between Memory-Related Problems subscale and HDRS score ($r = 0.00$, $p = ns$) or diagnosis of major depression ($r = -0.01$, $p = ns$). Sensitivity to change: Indicated

^aThe Behavioral Domain column lists area using the labeling of behaviors as reported within the cited paper.

^bN = nurse; F = family caregiver; TI = trained interviewer; TO = trained observer; C = clinician.

^cO = observation; I = interview.

^dD = person with dementia; P = proxy respondent; C = clinician.

* Scale developed for use in persons with dementia.

** Scale developed for use in a population other than dementia, tested to be reliable and valid in persons with dementia.

*** Scale developed for use in general population, used in dementia population but psychometrics untested or inconsistent results.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

- α = Cronbach's alpha.
- BEHAVE-AD = Behavioral Pathology in Alzheimer's Disease.
- BRSD = CERAD Behavior Rating Scale for Dementia.
- BSSD = Behavioral Syndromes Scale for Dementia.
- CABOS = Computer Assisted Behavioral Observation Systems.
- CDR = Clinical Dementia Rating Scale.
- DBD = Dementia Behavior Disturbance Scale.
- DSS = Dementia Signs and Symptoms Scale.
- FrSBc = Frontal System Behavioral Scale.
- KBCI = Key Behavior Change Inventory.
- MOSES = Multidimensional Observation Scale for Elderly Subjects.
- NHBPS = The Nursing Home Behavior Problem Scale.
- NOSGER = Nurses' Observation Scale for Geriatric Patients.
- NPI = Neuropsychiatric Inventory.
- NPI-C = Neuropsychiatric Inventory – Clinician.
- NPI-Q = Neuropsychiatric Inventory – Questionnaire.
- NRS = The Neurobehavioral Rating Scale.
- RMBPC = Revised Memory and Behavior Problem Checklist.

Table 2
Summary of specific measures of behaviors to assess neuropsychiatric symptoms in persons with dementia (N = 29)

Agitation Scales (N = 7)						
Assessment Tool	Number of Items	Administration			Response Categories	Estimated Time to Administer
		Who ^d	How ^b	To whom ^c		
1 Agitated Behavior in Dementia scale* (ABID; Logsdon et al., 1999)	16 items	TI, F	I	P	Frequency rated on past two weeks – each week rated separately. Frequency rated 0 = did not occur in the week to 3 = occurred daily or more often. Two weekly scores are added together for a final score on each item of 0 to 6. Range 0–48 Higher scores indicate greater agitation. Caregiver reaction only rated once in two weeks. Caregiver reaction rated 0 = not upsetting to 4 =extremely upsetting Reaction range 0 to 64. Higher scores indicate greater reaction.	<20 minutes
2 Brief Agitation Rating Scale* (BARS; Finkel et al., 1993) (based on Cohen-Mansfield Agitation Inventory)	10 items	N	O	D	Based on prior two weeks Rated 1 = none to 7 = several times a day Range 10–70 Higher scores indicate greater agitation.	Not specified
3 Cohen-Mansfield Agitation	29 items Short form is 14 items.	TI, F	I	P	Based on prior two weeks	<30 minutes
						<p>Reliability: Internal consistency: 0.70 Test-retest: 0.60–0.73 Validity: Type not specified: validity confirmed with correlation to RMBPC ($r = 0.74, p < 0.0001$), BRSD ($r = 0.65, p < 0.0001$), and the CMAI ($r = 0.62, p < 0.0001$). Sensitivity to change: Indicated</p> <p>Reliability: Internal consistency: $\alpha = 0.74–0.82$ Inter-rater: intraclass correlation $r = 0.73$ Validity: Concurrent validity: significant correlation with BEHAVE-AD and BSSD Sensitivity to change: Not indicated</p> <p>Reliability: Reliability: Originally designed for nursing home patients but also used</p>

Agitation Scales (N = 7)

Assessment Tool	Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
		Who ^a	How ^b	To whom ^c				
Inventory* (CMAI; Cohen-Mansfield <i>et al.</i> , 1989; Cohen-Mansfield, 1991; Finkel <i>et al.</i> , 1992)	Community form is 37 items.				Rated 1 = never to 7 = several times in an hour Range 0–203 Higher scores indicate greater agitation.		in community settings.	Internal consistency: $\alpha = 0.86$ – 0.91 based on shift worked Inter-rater: 0.41 for the total score Reliability for the short form: exact agreement = 0.82 point discrepancy = 0.93 <u>Validity:</u> Type not specified; Pearson product-moment correlations between CMAI and Behave-AD and BSSD range from 0.0304–0.5177 depending on shift. <u>Sensitivity to change:</u> Not indicated
4 Disruptive Behavior Rating Scales* (DBRS; Mungas <i>et al.</i> , 1989)	21 items in four dimensions of disruptive behavior Physical aggression Verbal aggression Agitation Wandering	TO and N	O, CR	D	Daily for a week Rated 0 = insufficient data to 5 = behavior occurs and has a severe effect or results in extreme intervention (life-threatening injury) Range 0–105 Higher scores indicate greater agitation.	5–10 minutes	Dementia patients in nursing facilities	<u>Reliability:</u> Inter-rater: Physical aggression: $r = 0.91$ Verbal aggression: $r = 0.83$ Agitation: $r = 0.84$ Wandering: $r = 0.71$ Total: $r = 0.93$ <u>Validity:</u> Concurrent validity: total score correlation with nurse's assessment rating for severity ($r = 0.73$, $p < 0.001$) and with distress ($r = 0.85$, $p < 0.001$) <u>Sensitivity to change:</u> Indicated
5 Overt Agitation Severity Scale* (OASS; Yudofsky <i>et al.</i> , 1997)	Twelve behavior units divided into three subgroups: vocalization and oral/facial movement Upper torso and upper extremity movement Lower extremity movements	TO	O	D	Rated during 15-minute observation period Intensity in three domains scored as 1 to 4 with each domain having different descriptions of intensity.	15 minutes	Adult psychiatric patients, including those with dementia	<u>Reliability:</u> Internal consistency: split-half ranging from 0.88–0.91 (depending on rater) Inter-rater: Pearson correlation coefficient ($r = 0.90$, $p < 0.01$) <u>Validity:</u> Convergent validity: strong association with

Agitation Scales (N = 7)

Assessment Tool	Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
		Who ^a	How ^b	To whom ^c				
6	Pittsburgh Agitation Scale* (PAS; Rosen <i>et al.</i> , 1994)	N	O	D	Item frequency rated as 0 = not present to 4 = always present. Item frequency rated as 0 = not present to 4 = always present. Intensity and frequency are multiplied for each item to give a severity score. Severity scores are totaled for the OASS total score. Higher scores indicate greater agitation.	<5 minutes	In-patient unit for dementia patients with behavioral problems and nursing home patients with dementia	PAS ($r = 0.81, p < 0.01$ for rater 1 and $r = 0.82, p < 0.01$ for rater 2) Discriminant validity: established by low correlation between OASS and OAS ($r = 0.28, p < 0.01$) Content validity: expert consensus Sensitivity to change: Not indicated
7	Scale for the Observation of Agitation in Persons with Dementia of the Alzheimer type* (SOAPD; Hurley <i>et al.</i> , 1999)	TO	O	D	Period of observation ranged from 1 to 8 h. Scale is 0 to 4; each group has different scoring criteria based on the behavior of interest. Scores are not totaled.	5 minutes	Dementia patients in long-term care facilities	Reliability: Inter-rater: intraclass correlation $r = +0.82 - +0.93$ for total score; individual item $r = +0.54 - +0.88$ Validity: Type not specified: validity is confirmed by the difference in scores when interventions to reduce agitation were initiated compared to no interventions. Sensitivity to change: Indicated Reliability: Internal consistency: $\alpha = 0.70$ Inter-rater: κ 's ranged from 0.55 to 0.90 Validity: Concurrent validity: significant correlations between SOAPD domains and SCMAI subscales Content validity: panel of nine experts Sensitivity to change: Not indicated

Apathy Scales (N = 7)

Agitation Scales (N = 7)

Assessment Tool	Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
		Who ^a	How ^b	To whom ^c				
1 Apathy Evaluation Scale* (AES; Marin, 1991; Marin <i>et al.</i> , 1991; Clarke <i>et al.</i> , 2007) Three versions: self: AES-S, informant: AES-I, clinician: AES-C	Eighteen core items	D, F, C	I	D	Based on current functioning or for patients hospitalized within 3–4 days. Rate based on past four weeks. Rated 1 = not at all true/characteristic to 4 = very true/characteristic (Three items are negatively worded and would need to be reversed scored for a total score) Range 18–72 Lower scores indicate more apathy	10–20 minutes	Adults, 18+ years in various settings	Reliability: Internal consistency: 0.86–0.094 In dementia patients: AES-C: $\alpha = 0.90$ AES-I: $\alpha = 0.90$ Test-retest $\alpha = 0.76$ – 0.94 Validity: Convergent validity: assessed by examining the correlation between the three versions of the AES (i.e. self, clinician, and informant): $r = 0.43$, $p < 0.01$ to 0.72 , $p < 0.01$. Discriminant validity: assessed by examining the correlation between apathy and depression (for self-rated ($r = 0.43$) and informant-rated ($r = 0.27$, $p < 0.01$)) and anxiety (for the clinician ($r = 0.35$, $p < 0.01$) and self-ratings ($r = 0.42$)). In dementia patients the AES-I provided the greatest sensitivity at 92.9%. AES-C 85.7% AES-S 61.5% Sensitivity to change: Indicated
2 Apathy Inventory* (IA; Robert <i>et al.</i> , 2002)	Three items	C	I	D (IA-patient), P (IA-caregiver)	Based on change since onset of the illness also can be used over a specified time period Items are present or absent. If present, frequency rated 1 = occasionally, to 4 = very frequently Severity rated 1 = mild to 3 = marked The AI-caregiver score range 0–36	Not specified	MCI, Parkinson's, and dementia outpatients	Reliability: Internal consistency: $\alpha = 0.84$ Inter-rater: $\kappa = 0.99$ Test-retest: $\kappa = 0.99$, 0.97 , and 0.99 for emotional blunting, lack of initiative, and lack of interest respectively, and 0.96 overall Validity: Concurrent validity: correlation between the lack-of-initiative ($r = 0.23$, $p < 0.01$) and lack-

Agitation Scales (N = 7)

Assessment Tool	Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
		Who ^a	How ^b	To whom ^c				
3 Dementia Apathy Interview and Rating scale* (DAIR; Strauss and Sperry, 2002)	Sixteen items Follow-up question determines behavioral changes from prior to AD diagnosis	TI	I	P	<p>Higher scores indicate greater apathy. Higher scores indicate greater apathy. Higher scores indicate greater apathy.</p> <p>In the AI-patient interview, patients report presence or absence of three AI items. If present, patient rates intensity 1 = mild to 12 = severe. Range 0–36 Higher scores indicate more severe apathy.</p> <p>Based on past month Rated 0 = no or almost never to 3 = yes, almost always Only items representing a change in behavior are included in the final apathy score. Higher scores represent greater apathy.</p>	Not specified	Patients in clinic with probable AD	<p>of-interest ($r = 0.63$, $p < 0.001$) items and the NPI apathy subscale score. Discriminant validity: AI caregivers distinguish AD patients and controls, with AD patients having significantly higher score on lack of initiative and global score than control. <u>Sensitivity to change:</u> Indicated</p> <p>Reliability: Internal consistency: overall: $\alpha = 0.89$; in-person: $\alpha = 0.91$; telephone: $\alpha = 0.94$ Inter-rater Reliability: determined by a second rater's rating of ten audio-taped interviews: $r = 1.00$, $p < 0.01$ Test-retest: assessed using 20 randomly selected caregivers with assessments on average 56 days apart: $r = 0.85$, $p < 0.001$ <u>Validity:</u> Convergent validity: correlation between apathy score and an independent clinician's blind assessment of apathy: $r = 0.31$, $p < 0.05$ to 0.46, $p < 0.01$ Criterion validity: optimal cut-points and associated sensitivity and specificity not determined Discriminant validity: very poor correlation between apathy score and depression: $r = 0.08$. <u>Sensitivity to change:</u> Not indicated</p>

Agitation Scales (N = 7)

Assessment Tool	Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
		Who ^a	How ^b	To whom ^c				
4 Irritability-Apathy Scale* (IAS; Burns <i>et al.</i> , 1990)	Ten items in two domains: Irritability Apathy	C	I	P	Rated compared to before onset of illness Irritability Question 1 rated 1 = not at all irritable to 5 = extremely irritable Questions 2-5 rated 1 = never to 3 = always Total possible = 17 Higher scores indicate greater irritability. Apathy Rated 1 to 5 with each question having different responses Total possible = 25 Higher scores indicate greater apathy	Not specified	Patients with AD or Huntington's disease in community	Reliability: Internal consistency: Irritability: $\alpha = 0.82$; apathy: $\alpha = 0.78$ Inter-rater: irritability $r = 1.00$; apathy: $r = 0.85$ Test-retest: irritability: $r = 0.81$; apathy: $r = 0.76$ Validity: Discriminant validity: no significant correlation between apathy and premorbid traits (i.e. being "good tempered," "bad tempered," "happy," or a "worrier"). Construct validity: IAS apathy subscale differentiated between controls and AD, and controls and HD ($p < 0.05$) Convergent validity: irritability score highly associated with Psychogeriatric Dependency Rating Scale ($r = 0.87, p < 0.001$). Sensitivity to change: Not indicated
5 Lille Apathy Rating Scale* (LARS; Sockeel <i>et al.</i> , 2006; Dujardin <i>et al.</i> , 2008)	Thirty-three items in nine domains: reduction in everyday productivity Lack of interest Lack of initiative Extinction of novelty-seeking Extinction of motivation Blunting of emotional responses Lack of concern Poor social life Extinction of self awareness	C	I	D	Based on past four weeks Items 1-3 rated (2 to -2) based on time to reply and number of activities named. Remaining 30 items are rated -1 to 1. Range -36 to +36 Higher and more positives core indicates greater severity of apathy.	Not specified	Parkinson's disease patients (including those with dementia) in the community	Reliability: Internal consistency: $\alpha = 0.80$ Inter-rater Reliability: intraclass correlation coefficient = 0.98 Test-retest reliability at four-months: $r = 0.95$ Validity: Concurrent validity: significant correlations between LARS and global scores for AES ($r = 0.87$) The validity of the LARS for assessing the presence and severity of apathy has been demonstrated in

Int Psychogeriatr. Author manuscript; available in PMC 2015 May 18.

Agitation Scales (N = 7)

Assessment Tool	Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
		Who ^a	How ^b	To whom ^c				
4 Ryden Aggression Scale* (RAS; Ryden, 1988; Ryden <i>et al.</i> , 1991) Also available the RAS-2	Twenty-five items in three subscales: Physically aggressive behavior Verbal aggression Sexual aggression	F, N	I	P, D	Based on past year Frequency Rated 0 = never to 5 = one or more times daily Range 0–125 Higher scores indicate greater aggressive behavior.	20 minutes	Community and hospital patients with dementia	<u>Reliability:</u> Internal consistency: $\alpha = 0.88$ Inter-rater: $r = 0.88$ Test-retest: $r = 0.86$ at 8–12 weeks <u>Validity:</u> Construct validity: RAS1 to RAS2 $r = 0.65$, $p < 0.001$ Content validity: literature and expert review Sensitivity to change: Indicated

Anxiety Scales (N = 4)

1 Beck Anxiety Inventory*** (BAI; Beck <i>et al.</i> , 1988; Fydrich <i>et al.</i> , 1992; Steer and Beck, 1997; Osman <i>et al.</i> , 1998)	Twenty-one items	D	I	D	Based on past week Rated 0 = not at all to 3 = severely, it bothered me a lot Range 0–63 Higher scores indicate greater anxiety. 0–21 = low anxiety 22–35 = moderate anxiety 36+ = potential for concern	Time to complete: 10 minutes; time to score: 5 minutes	General population in community settings, has been used in dementia population but psychometrics are mixed	<u>Reliability:</u> Internal consistency: $\alpha = 0.92$ Test-retest: 0.75 ($df = 81$); one week interval <u>Validity:</u> Convergent validity: correlations between BAI and HAM-A and HAM-D were 0.51 ($df = 150$) and 0.25 ($df = 153$), respectively. Discriminant validity: correlation between BAI and CCL-A (0.51, $df = 151$), CCL-D (0.22, $df = 150$), and HS (0.15, $df = 158$) One study (Wetherell and Gatz, 2005) found limitations with the use in older adults and another questioned its use in patients with Parkinson's. <u>Sensitivity to change:</u> Not indicated
2 Geriatric Anxiety Inventory** (GAI; Pachana <i>et</i>	Twenty items Five items	N, D N, D	II	D D	Based on past week Rated 0 (disagree) to 1 (agree)	Not specified Not specified	Older adults community dwelling and nursing homes,	<u>Reliability:</u> Internal consistency: $\alpha = 0.91$ among normal

Agitation Scales (N = 7)

Assessment Tool	Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
		Who ^a	How ^b	To whom ^c				
<i>al.</i> , 2007; Bodice <i>et al.</i> , 2008) GAI Short Form ** (GAI-SF; Byrne and Pachana, 2011)					Range 0-20 Scores of 9 indicate clinical anxiety symptomatology. Based on past week Rated 0 (disagree) to 1 (agree) Range 0-5 Scores of >3 indicate clinical anxiety symptomatology		have been used in patients with dementia. Older adults community dwelling	elderly; $\alpha = 0.93$ in psychogeriatric sample Inter-rater: $r = 0.99$, $p < 0.0000$ in psychogeriatric sample Test-retest: $r = 0.91$, $p < 0.0000$ in psychogeriatric sample Short form: Internal consistency: $\alpha = 0.81$ in GAI-SF Correlation between GAI and GAI-SF was 0.88. Area under ROC curve: 0.80 (95% CI 0.64–0.97) <u>Validity:</u> Concurrent validity: significant correlations in expected directions with the GADS ($r = 0.57$, $p < 0.001$), STAI ($r = -0.44$, $p < 0.001$), BAI ($r = 0.63$, $p < 0.001$), Penn State Worry Questionnaire ($r = 0.70$, $p < 0.001$), and the Positive and Negative Affect Schedule – negative ($r = 0.58$, $p < 0.001$) and positive ($r = 0.34$, $p < 0.001$) subscales. Cut point 10/11: Specificity: 0.84 Sensitivity: 0.75 Short form: Cut point 2/3: Specificity: 0.87 Sensitivity: 0.75 Short form: Concurrent validity: significant correlation with the STAI ($r = 0.48$, $p < 0.001$) <u>Sensitivity to change:</u> Indicated
3 Rating Anxiety in Dementia* (RAID; Shankar <i>et al.</i> , 1999)	Twenty items (six sub-groups)	C	I, CR	P, D	Based on past two weeks Rated 0 = absent to 3 = severe	10–15 minutes	Persons with dementia in hospitals, nursing homes, and community	<u>Reliability:</u> Internal consistency: $\alpha = 0.83$

Agitation Scales (N = 7)

Assessment Tool	Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
		Who ^a	How ^b	To whom ^c				
RAID with structured interview also available RAID with structured interview also available RAID with structured interview also available RAID with structured interview also available					Range 0-60 11 suggests significant clinical anxiety			Inter-rater: κ 's ranged from 0.51 to 1 and overall agreement ranged from 82-100%. Test-retest: κ 's ranged from 0.53 to 1 and overall agreement ranged from 84-100%. Validity: Content validity: panel of experts and professionals working with older dementia patients Concurrent validity: Correlation with Carer's rating (0.73) Only 38 of the 83 participants were able to complete the other measures of anxiety: Clinical Anxiety Scale (0.54), Anxiety Status Inventory (0.62). All correlations were significant at $p < 0.001$. A modified version of the RAID with the depression items removed was compared to the CSDD (0.2). Construct validity: principal component analysis found a 5-factor structure of 18 items and accounted for 63.8% of variance. KMO = 0.768. Sensitivity to change: Indicated.
4 The Worry Scale* (LaBarge, 1993)	Eight items	D	I	D	Rating period not stated Rated 5 = Always to 1 = Never Two items are reverse coded. Range 8-40 Higher scores indicate greater worry.	Not specified	Persons with dementia living in the community	Reliability: Internal consistency: $\alpha = 0.85$ Validity: Construct validity: factor analysis found one dimension with factor weights of 0.448-0.776. Concurrent validity: correlations with State Trait Anxiety Inventory ($r = 0.55, p < 0.0001$)

Agitation Scales (N = 7)

Assessment Tool		Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
			Who ^a	How ^b	To whom ^c				
Depression Scales (N = 4)									
1	Cornell Scale for Depression in Dementia* (CSDD; Alexopoulos <i>et al.</i> , 1988)	Nineteen items	C	I	D, P	Based on week prior except for weight loss, loss of interest, and lack of energy which are evaluated in the past month. Rated 0 = absent to 2 = severe Range 0–38 Higher scores indicate greater depressive symptomatology.	30 minutes (20 minutes with caregiver and 10 minutes with patient)	Dementia patients in various settings	Reliability: Internal consistency: $\alpha = 0.84$ Inter-rater Reliability: $\kappa = 0.67$ Validity: Concurrent validity: significant correlation between score on CSDD and Research Diagnostic Criteria for depression diagnosis ($r = 0.83, p < 0.001$) Sensitivity to change: Indicated
2	The Dementia Mood Assessment Scale* (DMAS; Sunderland <i>et al.</i> , 1988)	Twenty-four items in two subgroups: depression severity of dementia	TI	I, CR	D	Based on past week Items 1–17 rate severity of depression 0 = within normal limits to 6 = most severe Items 18–24 rate severity of dementia 0 = within normal limits to 6 = most severe Only items 1–17 are considered in total score. Range 0–102 Higher scores indicate greater depression symptomatology.	20–30 minutes	Dementia patients – inpatient or outpatient	Reliability: Inter-rater Reliability: Depression items: ($r = 0.69–0.74, p < 0.0001$) Other items: $r = 0.28$ (mania)–0.77 (functional impairment), $p < 0.01$ for all Reliable in mild to moderate AD but not in severe AD Validity: Construct validity: correlation with global depression scores ($r = 0.73$) Sensitivity to change: Indicated
3	The Geriatric Depression Scale*** (GDS; Yesavage <i>et al.</i> , 1983)	Thirty items	TI	I	D	Based on past week Yes/No response Several items are reverse scored. 0 = not indicative of depression 1 = indicative of	5–10 minutes	General population has been used in dementia population with mixed results on reliability and validity.	Reliability: Internal consistency: $\alpha = 0.94$ Test–retest: correlation was 0.85 ($p < 0.001$) Validity: Convergent validity: correlations of 0.84 ($p <$

Sensitivity to change:
Not indicated

Depression Scales (N = 4)

Reliability:
Internal consistency: $\alpha = 0.84$
Inter-rater Reliability: $\kappa = 0.67$
Validity:
Concurrent validity: significant correlation between score on CSDD and Research Diagnostic Criteria for depression diagnosis ($r = 0.83, p < 0.001$)
Sensitivity to change: Indicated

Reliability:
Inter-rater Reliability: Depression items: ($r = 0.69–0.74, p < 0.0001$)
Other items: $r = 0.28$
(mania)–0.77 (functional impairment), $p < 0.01$ for all
Reliable in mild to moderate AD but not in severe AD
Validity: Construct validity: correlation with global depression scores ($r = 0.73$)
Sensitivity to change: Indicated

Reliability:
Internal consistency: $\alpha = 0.94$
Test–retest: correlation was 0.85 ($p < 0.001$)
Validity:
Convergent validity: correlations of 0.84 ($p <$

Agitation Scales (N = 7)

Assessment Tool		Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
			Who ^a	How ^b	To whom ^c				
4	Patient Health Questionnaire – 9*** (PHQ-9; Kroenke <i>et al.</i> , 2001)	Nine items	TI, D	I	D	depression Range 0–30 depression Range 0–30 Higher scores indicate greater depression symptomatology. Based on past two weeks Rated 0 = not at all to 3 = nearly every day Total scores range from 0–27 Higher scores indicated more depressive symptomatology. Five or items scoring >2 indicates major depression. Maps to DSM-IV-TR	5 minutes	General population in a variety of settings has been used in patients with dementia.	0.001) with the Zung Self-Rated Depression Scale (SDS) and 0.83 (<i>p</i> < 0.001) with the Hamilton Rating Scale for Depression (HRS-D) for Depression (HRS-D) Sensitivity to change: Indicated Reliability: Internal consistency: <i>a</i> = 0.86–0.89 Validity: Criterion validity: TROC analysis found the area under the curve was 0.95. Construct validity: strong correlation with mental health portion of SF-20 (0.73). Sensitivity to change: Indicated
Sleep Scales (N = 2)									
1	Epworth Sleepiness Scale*** (ESS; Johns, 1991; Johns, 1992)	Eight items	TI	I	P, D	Based on recent times (no exact period given) Rated 0 = no chance of dozing to 3 = high chance of dozing Range 0–24 Higher scores indicate greater daytime sleepiness. Scores of 10 indicate above normal daytime sleepiness	Not specified	General population in any healthcare setting has been used in studies of people with dementia.	Reliability: Internal consistency: <i>a</i> = 0.74–0.88 Validity: Not reported Sensitivity to change: Not indicated
2	Pittsburgh Sleep Quality Index*** (PSQI; Buysse <i>et al.</i> , 1989)	Nineteen items self-rated Five items rated by bedpartner or roommate are part of the original scale but are not	TI	I	D	Based on past month First four items ask for time or amount of sleep Items 5–18 rated 0 = not in the past month	5–10 minutes for patient to complete; 5 minutes to score	General population in all healthcare settings has been used with patients with dementia	Reliability: Internal consistency: <i>a</i> = 0.83 Test-retest: no difference found (<i>r</i> = 2.32, <i>p</i> = 0.03) Validity:

Agitation Scales (N = 7)

Assessment Tool	Number of Items	Administration			Response Categories	Estimated Time to Administer	Setting and Target Population	Psychometric Properties
		Who ^a	How ^b	To whom ^c				
3 The Sleep Disorders Inventory* (SDI; Tractenberg <i>et al.</i> , 2003)	Seven items	TI	I	P	to 3 = three or more times a week One global item rated 0 = very good to 3 = very bad Scoring is done in seven components and then all components are totaled. Scores can range from 0 to 21. A score of 5+ indicates poor sleeper.	Not specified	AD patients in community and living with caregiver	Sensitivity 89.6%, specificity 86.5% ($\kappa = 0.75, p < 0.001$) in distinguishing good and poor sleepers Sensitivity to change: Indicated
					Based on past two weeks Frequency rated 0 = not present to 4 = once or more per day (every night) Severity rated 0 = not present to 3 = marked Caregiver distress rated 0 = not at all to 5 = very severely/extremely average. Frequency and average severity scores are multiplied to calculate total score. Range 0–12 Higher scores indicated more sleep disturbance.			
								Reliability: Not reported Validity: Content validity: correlation with NPI sleep subscale, $r = 0.341, p < 0.05$ Sensitivity to change: Not indicated

Depression and Anxiety Scales (n = 1)

1 Hospital Anxiety and Depression Scale*** (HADS; Zigmond and Snaith, 1983; Bjelland <i>et al.</i> , 2002; Flint and Rifat, 2002)	Fourteen items in two domains: Anxiety Depression	D	I	D	Based on past week Scored from 0–3 Specific response wording varies with each item. Total for each subscale ranges from 0–21.	<10 minutes	General population in community and hospital settings has been used in patients with mild dementia but psychometrics have not been tested in this population.	Reliability: Internal consistency: General population: Depression: $\alpha = 0.67–0.90$ Anxiety: $\alpha = 0.68–0.93$ Older Adults: Depression: $\alpha = 0.77$ Anxiety: $\alpha = 0.76$ Validity:
---	---	---	---	---	--	-------------	---	--

Agitation Scales (N = 7)						
Assessment Tool	Number of Items	Administration			Response Categories	Estimated Time to Administer
		Who ^a	How ^b	To whom ^c		
					Higher scores indicate greater symptoms Higher scores indicate greater symptoms Higher scores indicate greater symptoms	
						Psychometric Properties Convergent validity: ranged from 0.49 to 0.83 Sensitivity and specificity were found to be approximately 0.80. Sensitivity to change: Indicated
Wandering Scales (N = 1)						
1	Algase Wandering Scale* (AWS; Algase <i>et al.</i> , 2001) Also available is the AWS (V2; Algase <i>et al.</i> , 2004)	TO, N	O	D	Rating based on general knowledge of person and not over a specified time period. Timed and coded ambulation in public areas of a nursing home Wandering classified by pattern and rhythm Pattern consisted of random lapping, and pacing. Rhythm based on cycle of locomotive and non-locomotive. The 28 items are times or episodes when wandering occurs, such as walks between lunch and dinner, aimless walks, or bumping into obstacles when walking. Range and interpretation of scores not provided	Dementia patients in assisted-living or nursing home Reliability: Internal consistency: $\alpha = 0.87$ Validity: Type not specified: all but the routinized subscale significantly correlated with staff reports of patient wandering ($p < 0.01$) Sensitivity to change: Not indicated

^a N = nurse; F = family caregiver; TI = trained interviewer; TO = trained observer; C = clinician; D = person with dementia.

^b O = observation; I = interview, CR = chart review.

^c D = person with dementia; P = proxy respondent; C = clinician.

* Scale developed for use in persons with dementia.

** Scale developed for use in a population other than dementia, tested to be reliable and valid in persons with dementia.

*** Scale developed for use in general population, used in dementia population but psychometrics untested or inconsistent results.

α = Cronbach's alpha.

ABID = Agitated Behavior in Dementia Scale.

ABS = Aggressive Behavior Scale.

ADL = Activities of Daily Living.

AES-C = Apathy Evaluation Scale–Clinician.

AES-I = Apathy Evaluation Scale–Informant.

AES-S = Apathy Evaluation Scale–Self.

AES = Apathy Evaluation Scale.

AWS = Algea Wandering Scale.

V2 = version 2.

BAI = Beck Anxiety Inventory.

BARS = Brief Agitation Rating Scale.

BDI = Beck Depression Inventory.

CCL-A = Cognition Checklist for Anxiety.

CCL-D = Cognition Checklist for Depression.

CMAI = Cohen Mansfield Agitation Inventory.

CSD = Cornell Scale for Depression in Dementia.

DAIR = Dementia Apathy Interview and Rating Scale.

DBRS = Disruptive Behavior Rating Scales.

DMAS = The Dementia Mood Assessment Scale.

DSM-IV-TR = Diagnostic and Statistical Manual of Mental Disorders IV.

ESS = Epworth Sleepiness Scale.

GAI = Geriatric Anxiety Inventory.

GDS = Geriatric Depression Scale.

HADS = Hospital Anxiety and Depression Scale.

HAM-A = Hamilton Rating Scale for Anxiety.

HAM-D = Hamilton Rating Scale for Depression.

HD = Huntington's Disease.

IADL = Instrumental Activities of Daily Living.

IA = Apathy Inventory.

IAS = Irritability Apathy Scale.

LARS = Lille Apathy Rating Scale.

MCI = Mild Cognitive Impairment.

MDS = Minimum Data Set.

mMMSE = Modified Mini-Mental Status Exam.

MOAS = Modified Overt Aggression Scale.

NOISE = Nurse Oriented Scale for Inpatient Evaluation.

OAS = Overt Aggression Scale.

OASS = Overt Agitation Severity Scale.

PANSS = Positive and Negative Symptom Scale.

PAS = Pittsburgh Agitation Scale.

PD = Parkinson's Disease.

PHQ-9 = Patient Health Questionnaire–9.

PLUT = Plutchik Scale.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

PSQI = Pittsburgh Sleep Quality Index.
RAGE = Rating Scale for Aggressive Behavior in the Elderly.
RAID = Rating Anxiety in Dementia.
RAS = Ryden Aggression Scale.
SANS = Scale for the Assessment of Negative Symptoms.
SDI = The Sleep Disorders Inventory.
SF-20 = 20-item Short Form Survey.
SOAPD = Scale to Assess Observed Agitation in Persons with Dementia of the Alzheimer Type.
TBI = Traumatic Brain Injury.

Table 3
Comparison of measures along settings, and numbers and types of items (N = 45)

Scale	Settings						Behavioral Categories							
	NH	ALF	Home	Hosp	None Spec	# Items	General	Agitation	Apathy	Aggressi on	Anxiety	Depression	Sleep	Wandering
ADAS	X		X			10	X							
BEHAVE-AD	X		X			26	X							
BSSD			X			24	X							
CERAD-BRSD					X	51	X							
CDR			X			75	X							
CABOS	X					N/A	X							
DBD			X			28	X							
DSS			X			43	X							
FrSBc			X			46	X							
KBCI			X			64	X							
MOSES	X			X		40	X							
NRS				X		27	X							
NPI					X	12-91	X							
NOSGER	X	X	X			30	X							
NHBPS	X					29	X							
RMBCPC			X			24	X							
ABID			X			16		X						
BARS	X					10		X						
CMAI	X		X			29		X						
DBRS	X					21		X						
OASS				X		12		X						
PAS	X			X		N/A		X						
SOAPD	X					7		X						
AES					X	18			X					
IA			X			3			X					
DAIR			X			16			X					
IAS			X			10			X					

Scale	Settings					# Items	Behavioral Categories							
	NH	ALF	Home	Hosp	None Spec		General	Agitation	Apathy	Aggressi on	Anxiety	Depression	Sleep	Wandering
LARS			X			33			X					
ABS	X			X		4				X				
OAS				X		4				X				
RAGE	X					21				X				
RAS			X	X		25				X				
BAI			X			21					X			
GAI	X		X			20					X			
RAID	X		X	X		20					X			
Worry Scale			X			8					X			
CSDD					X	19						X		
DMAS			X	X		24						X		
GDS					X	30						X		
PHQ-9					X	9							X	
ESS					X	8							X	
PSQI					X	19							X	
SDI			X			7								
HADS			X	X		14					X	X		
AWS	X	X				29								X

NH = nursing home.

ALF = assisted living facility.

HOME = refers to any community based setting including the person's home, outpatient clinics, or community centers.

HOSP = Hospital.

None spec = setting not specified by papers.

General = broad range of behavioral categories included.

Depress/anxiety = includes scales focusing on depression only, anxiety only or depression and anxiety combined.

N/A = not applicable or difficult to determine as number.