The BEHAVE-AD Assessment System: A Perspective, A Commentary on New Findings, and A Historical Review

Barry Reisberg a–c Isabel Monteiro a, b Carol Torossian a, b Stefanie Auer d Melanie B. Shulman a, b Santosh Ghimire a, b Istvan Boksay a, b Francoise Guillo BenArous a, b Ricardo Osorio a, b Aninditha Vengassery a, b Sheema Imran a–c Hussam Shaker a–c Sadaf Noor a, b Shazia Naqvi a, b Sunnie Kenowsky c Jinfeng Xu b

a Aging and Dementia Research Center, b Alzheimer’s Disease Center and c Fisher Alzheimer’s Program, New York University School of Medicine, New York, N.Y., USA; d M.A.S. Alzheimerhilfe and Department for Clinical Neurosciences and Preventative Medicine, Danube University, Krems, Austria

Key Words
Dementia · Alzheimer’s disease · Behavioral and psychological symptoms of dementia · Psychiatric status rating scales · Assessments · Pharmacotherapy · Management · Antipsychotics · Anxiolytics · Patient outcome assessment

Abstract
Background: Behavioral and psychological symptoms of dementia (BPSD) and associated disturbances in Alzheimer’s disease (AD) are a source of distress and burden for spouses, professional caregivers, and others with responsibilities for the care of individuals with AD. BPSD with behavioral disturbances are also associated with more rapid institutionalization and increased morbidity and mortality for persons with AD. Objectives: In this review and commentary, we discuss the history of the development of BPSD and behavioral disturbance assessments, which are distinct from those evaluating cognitive and functional symptoms of AD. In particular, we review the informant-based Behavioral Pathology in Alzheimer’s Disease Rating Scale (BEHAVE-AD), the related, potentially more sensitive, BEHAVE-AD Frequency-Weighted Severity Scale (BEHAVE-AD-FW), and the direct subject evaluation-based Empirical BEHAVE-AD Rating Scale (E-BEHAVE-AD). The kinds of medications that alleviate behavioral symptoms on these measures as well as the problems and possibilities for further advances with these medications are discussed. Finally, the importance of distinguishing BPSD and behavioral disturbance remediation in AD from the treatment of cognitive decline and other aspects of AD is emphasized in the context of appropriate assessment methodology. The objective of this
The paper is to provide a framework for further advances in the treatment of BPSD and associated behavioral disturbances in AD and, consequently, a framework for continuing improvements in the lives of individuals with AD and those who share the burden of the disease with the AD person.

**Introduction and Overview**

A little more than a quarter century after the Behavioral Pathology in Alzheimer’s Disease Rating Scale (BEHAVE-AD) was published [1] and more than a decade since the publication of the more detailed and potentially more sensitive BEHAVE-AD Frequency-Weighted Severity Scale (BEHAVE-AD-FW) [2], new insights are emerging with respect to the nature of

### Abbreviations used in this paper

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAGP</td>
<td>American Association for Geriatric Psychiatry</td>
</tr>
<tr>
<td>AD</td>
<td>Alzheimer’s disease</td>
</tr>
<tr>
<td>ADCS</td>
<td>Alzheimer’s Disease Cooperative Studies</td>
</tr>
<tr>
<td>APA</td>
<td>American Psychiatric Association</td>
</tr>
<tr>
<td>BEHAVE-AD</td>
<td>Behavioral Pathology in Alzheimer’s Disease Rating Scale</td>
</tr>
<tr>
<td>BEHAVE-AD-FW</td>
<td>Behavioral Pathology in Alzheimer’s Disease Frequency-Weighted Severity Scale</td>
</tr>
<tr>
<td>BPRS</td>
<td>Brief Psychiatric Rating Scale</td>
</tr>
<tr>
<td>BPSD</td>
<td>behavioral and psychological symptoms of dementia</td>
</tr>
<tr>
<td>CASI</td>
<td>Cognitive Abilities Screening Instrument</td>
</tr>
<tr>
<td>CATIE-AD</td>
<td>Clinical Antipsychotic Trials of Intervention Effectiveness – Alzheimer’s Disease</td>
</tr>
<tr>
<td>CDR</td>
<td>Clinical Dementia Rating scale</td>
</tr>
<tr>
<td>CGIC</td>
<td>Clinical Global Impression of Change assessment</td>
</tr>
<tr>
<td>CGI-S</td>
<td>Clinical Global Impression of Severity assessment</td>
</tr>
<tr>
<td>CIBI</td>
<td>Clinician Interview Based Impression assessment</td>
</tr>
<tr>
<td>CIBIC</td>
<td>Clinician’s Interview-Based Impression of Change assessment</td>
</tr>
<tr>
<td>CI-PCM</td>
<td>Comprehensive, Individualized Person-Centered Management program</td>
</tr>
<tr>
<td>CMAI</td>
<td>Cohen-Mansfield Agitation Inventory</td>
</tr>
<tr>
<td>CT</td>
<td>Computed tomographic scans</td>
</tr>
<tr>
<td>DSM-III-R</td>
<td>Diagnostic and Statistical Manual of Mental Disorders Third Edition-Revised</td>
</tr>
<tr>
<td>E-BEHAVE-AD</td>
<td>Empirical Behavioral Pathology in Alzheimer’s Disease Rating Scale</td>
</tr>
<tr>
<td>FAST</td>
<td>Functional Assessment Staging scale</td>
</tr>
<tr>
<td>FDA</td>
<td>United States Food and Drug Administration</td>
</tr>
<tr>
<td>FDG PET</td>
<td>18F-2-deoxy-2-fluro-D-glucose positron emission tomography</td>
</tr>
<tr>
<td>GDS</td>
<td>Global Deterioration Scale</td>
</tr>
<tr>
<td>ICC</td>
<td>intraclass correlation coefficient</td>
</tr>
<tr>
<td>IPSC-E</td>
<td>inventory of Psychiatric and Somatic Complaints in the Elderly</td>
</tr>
<tr>
<td>MCI</td>
<td>Mild Cognitive Impairment</td>
</tr>
<tr>
<td>MMSE</td>
<td>Mini-Mental Status Examination</td>
</tr>
<tr>
<td>NBRPS</td>
<td>Neurobehavioral Rating Scale</td>
</tr>
<tr>
<td>NCIC</td>
<td>No Cognitive Impairment</td>
</tr>
<tr>
<td>NINCDS-ADRDA</td>
<td>National Institute for Neurological and Communicative Diseases and Stroke-Alzheimer’s Disease and Related Disorders Association</td>
</tr>
<tr>
<td>NPI</td>
<td>Neuropsychiatric Inventory</td>
</tr>
<tr>
<td>NYU-CIBIC-Plus</td>
<td>New York University Clinician’s Interview-Based Impression of Change plus Caregiver Input assessment</td>
</tr>
<tr>
<td>SCAG</td>
<td>Sandoz Clinical Assessment Geriatric scale</td>
</tr>
<tr>
<td>SCI</td>
<td>Subjective Cognitive Impairment</td>
</tr>
<tr>
<td>SD</td>
<td>standard deviation</td>
</tr>
<tr>
<td>SPAD</td>
<td>Symptoms of Psychosis in Alzheimer’s Disease rating scale</td>
</tr>
</tbody>
</table>
behavioral and psychological symptoms of dementia (BPSD) [3, 4] in Alzheimer’s disease (AD) [5, 6]. In this review and commentary, we initially discuss the most recent findings regarding delusions in AD and offer a perspective on these findings. We then provide a historical discussion on the development of the Global Deterioration Scale (GDS) [7] and the associated delineation of both a severe dementia stage not addressed or conflated by the contemporaneously published Clinical Dementia Rating (CDR) scale [8] and mild cognitive impairment (MCI) as well as subjective cognitive impairment (SCI) stages also not addressed or conflated by the CDR staging procedure [9–11]. The implications of these early and late GDS stages for some of the new findings of Nakatsuka et al. [5] are elaborated. Subsequently, we discuss the development of the BEHAVE-AD, which rigorously separates the behavioral disturbance symptoms from cognitive and functional symptoms in AD. This discussion includes the construct validity, reliability, and criterion validity studies of this instrument for the assessment of AD-related BPSD. Further, we discuss apparent errors in prescribing practices for BPSD medications and the results and consequences of these errors that, to some extent, still exist.

In order to more sensitively and precisely assess the BEHAVE-AD symptoms, we subsequently developed the BEHAVE-AD-FW, which adds a frequency weighting to the severity weightings of the traditional BEHAVE-AD measure [2]. The conceptual framework, methodology, assessment validity, reliability, and other properties of this potentially more sensitive and more informative expansion of the BEHAVE-AD are discussed.

The BEHAVE-AD and the BEHAVE-AD-FW are both informant-based instruments. Direct observation of BPSD symptomatology can also be of considerable value. For this purpose, we developed the Empirical BEHAVE-AD Rating Scale (E-BEHAVE-AD) [12]. The conceptual framework, reliability, criterion validity, and utility of this direct observation assessment instrument are reviewed.

The BEHAVE-AD and the E-BEHAVE-AD have been very widely used. For example, the BEHAVE-AD has been used in approvals for risperidone (Risperdal) for the treatment of BPSD-related conditions in many nations throughout the world. In addition, the BEHAVE-AD and the E-BEHAVE-AD have been used as elements of the syncretic New York University Clinician’s Interview-Based Impression of Change plus Caregiver Input (NYU-CIBIC-Plus) assessment [13, 14]. In this context, these measures have been used as part of a pivotal outcome measure in the worldwide initial approvals of 2 of the 3 most widely used medications for the treatment of AD in the world today, namely, rivastigmine (Exelon) [15, 16] and memantine (Namenda, Ebixa, Axura) [17]. The BEHAVE-AD-FW, which should have approximately 2.5 times the sensitivity of the traditional BEHAVE-AD, holds great promise as a sensitive outcome measure in future BPSD clinical trial investigations [2].

There is widespread recognition that the nature and content of BPSD in AD and closely related dementias such as cerebrovascular dementia, which generally occurs in association with AD, are characteristic for this condition. For example, Jeste and Finkel [18] have elaborated the numerous differences between psychosis in AD and related dementias and psychosis in schizophrenia. The BEHAVE-AD and the associated assessments (i.e., the BEHAVE-AD-FW and the E-BEHAVE-AD) incorporate characteristic BPSD of AD, which are associated with disturbances and which, with increasing magnitude of the symptomatology, require treatment. In addition, the BEHAVE-AD and the associated measures rigorously exclude primarily cognitive and/or functional domain symptoms from the rating scales [9, 19] and assess symptoms that differ fundamentally from the cognitive and functional symptoms of AD. For example, the symptoms identified with the BEHAVE-AD and the BEHAVE-AD-FW have all been demonstrated to peak at a stage prior to the final GDS and Functional Assessment Staging (FAST) [20] stage of AD [21, 22]. This is in marked distinction to the cognitive and functional symptoms of AD, which decline continuously throughout the course of AD [9, 19, 23]. The symptoms identified by the
BEHAVE-AD and the associated assessments also differ from the cognitive and functional domain symptoms in AD in their relative responsiveness both to nonpharmacological management-type effects, such as those associated with being placed in the placebo arm of a clinical trial, and to treatments with medications associated with a reduction in ‘disturbances’ including anxiolytic as well as antipsychotic medications [24–28]. Furthermore, these positive effects on the BEHAVE-AD symptoms have been demonstrated to be independent of the effects of the management-placebo interventions, or the pharmacologic interventions, on cognition.

Finally, this review concludes with a discussion of the differences between the approach taken by the BEHAVE-AD and the associated assessments and that taken by the Neuropsychiatric Inventory (NPI) [29, 30]. The objectives of the BEHAVE-AD and the BEHAVE-AD-FW are to sensitively assess potentially remediable behavioral disturbances in AD patients. In contrast, the NPI assesses neuropsychiatric symptomatology much more broadly. For example, from the perspective of AD, the NPI assesses one category of symptoms that appears to relate to manif (mania-like) disorders and which has not been found to be related to AD [31]. Also, the NPI assesses symptoms such as weight gain, which, in addition to being unrelated to the behavioral disturbances in AD or to AD symptoms more generally, is, in the context of the treatment of behavioral disturbances in AD, much more likely to reflect side effects of BPSD behavioral disturbance medications than to relate to the treatment of BPSD disturbances. In the context of the treatment of BPSD disturbances in AD, side effects of medications should probably not be assessed with the same instrument that assesses targeted therapeutic symptomatology.

Finally, and perhaps most importantly, in the context of the treatment of BPSD-associated behavioral disturbances in AD, the NPI, even for the most germane symptomatic categories, does not assess individual symptoms. With the NPI, only the symptom categories are assessed on severity and frequency measures. For example, for the very important ‘Delusions’ category, the NPI does not assess the severity or frequency of the individual delusional symptoms. Only the general Delusions category is assessed for severity and frequency. In contrast, the BEHAVE-AD and the BEHAVE-AD-FW assess 7 separate delusional symptoms. In the 40-site study of Katz et al. [25] conducted at centers throughout the United States, each of the 7 delusional symptoms and symptomatic categories showed a significant response to the medication treatment in comparison with the placebo treatment using only the traditional BEHAVE-AD [1]. With the BEHAVE-AD-FW, each of these delusional symptomatic categories is assessed not only on the severity dimension but also on the frequency-weighted severity dimension. Therefore, from the standpoint of the treatment of BPSD behavioral disturbances in AD, the BEHAVE-AD and, especially, the BEHAVE-AD-FW should be much more sensitive and much more specific for AD-related symptomatology than the NPI.

Herein, we review these measures, i.e. the BEHAVE-AD, the BEHAVE-AD-FW, and the E-BEHAVE-AD, in detail. This review includes current findings, the historical development of these measures, scale properties including the sensitivity of the measures to therapies, and, finally, future directions.

**Recent Findings of Nakatsuka et al. [5] with the BEHAVE-AD-FW**

The publication of Nakatsuka et al. [5] is an important work, which will hopefully serve to advance the current understanding of delusions in AD. The insights from this study may also be important in terms of current research and treatment of BPSD in AD and closely related dementias of late life more generally [3, 4]. The important finding of their study [5] is that, unlike delusions in schizophrenia, which tend to occur in a closely associated manner, delusions in AD, when viewed from a frequency-weighted severity perspective, are disparate.
entities, which are not necessarily related to each other. Therefore, the assessment, treatment, and origins of the major delusional entities in AD might preferably be investigated separately.

Specifically, Nakatsuka et al. [5] have found that, in terms of the frequency-weighted severity scores on the BEHAVE-AD-FW [2], the specific delusion that ‘residence is not home’ peaks in frequency-weighted severity scores in more severely impaired community-residing individuals with probable AD [32], i.e., individuals with Mini-Mental State Examination (MMSE [33]) scores in the range of 0–9. In contrast, the frequency-weighted severity scores of other delusions, specifically assessed using the BEHAVE-AD-FW, may show very different patterns as AD evolves from less severe cognitively and functionally based stages [7, 8] to more severe stages. However, perhaps in part because of sample size limitations, the differences in manifest frequency-weighted severity delusion scores between cognition-based severity groups were not significant for delusional entities other than the ‘residence is not home’ delusion in this study [5]. Therefore, the patterns observed by Nakatsuka et al. [5] for delusions apart from the ‘residence is not home’ delusion should be considered tentative at the present time.

For example, the study observed that the delusion of theft (‘people are stealing things’ in the BEHAVE-AD-FW) peaks in frequency-weighted severity scores in AD patients with MMSE scores of 20–30 and, with the subsequent evolution of AD cognition-based severity, appears to be less manifest [5]. Other specific delusions in AD were found in this study to show more variegated patterns of frequency-weighted severity scores as AD progresses. An example of this latter pattern is the delusion of ‘infidelity’ (social and/or sexual unfaithfulness on the part of the spouse, children, and/or caregivers of individuals with dementia). This infidelity delusion peaks in frequency-weighted severity manifestation in the mild AD subjects studied by the investigators and becomes less manifest in moderate AD subjects with an MMSE score of 10–19. However, the delusion of infidelity seems to be more manifest once again as the AD process advances to the most severely impaired subject group studied by Nakatsuka et al. [5]. Changes in suspiciousness and delusions, apart from the 5 manifestations individually assessed with the BEHAVE-AD-FW, appear to show frequency-weighted manifestation patterns, which increase with each increasing cognitive severity level studied by Nakatsuka et al. [5]. However, the increments are not significantly different and much less dramatic than those observed for the ‘residence is not home’ BEHAVE-AD-FW delusion.

Comparison of Findings with the BEHAVE-AD-FW and a Study of the Incidence of BEHAVE-AD Symptoms

Comparison of BEHAVE-AD-FW Scores and BEHAVE-AD Incidence Findings for Delusions in General

The results obtained by Nakatsuka et al. [5] in a Japanese community-based outpatient setting using the BEHAVE-AD-FW assessment can be compared with those of an incidence study of BEHAVE-AD delusions, which used the traditional BEHAVE-AD, unmodified by the frequency of occurrence, in a US-based population nearly a quarter century ago [22]. Like the Japanese study of the BEHAVE-AD-FW, the incidence study required that all impaired subjects with dementia fulfill the American Psychiatric Association (APA) Diagnostic and Statistical Manual Third Edition-Revised (DSM-III-R) criteria for dementia as well as the National Institute for Neurological and Communicative Diseases and Stroke-Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA) criteria for a diagnosis of probable AD [32, 34]. The US study examined the incidence of the 7 delusional types identified with the tradi-
tional BEHAVE-AD assessment [1]. The traditional BEHAVE-AD is a severity scale without a frequency-weighted dimension [1]. However, the incidence study did not take the severity of the symptomatology or the frequency of occurrence into consideration – only the presence or absence of delusional symptoms was studied. The results indicated that the incidence of 5 of the 7 BEHAVE-AD delusional symptoms specifically assessed with this rating instrument peaked in subjects with a moderate dementia stage (GDS stage 5 [7]). These subjects had mean MMSE scores (±SD) of 14.1 ± 3.0. The other 2 of the 7 delusional symptomatic types assessed in this study peaked in the moderately severe AD subjects (GDS stage 6), who had a mean MMSE score of 4.9 ± 4.6. Specifically, the incidence of the delusion of ‘infidelity’ (social and/or sexual unfaithfulness) and the incidence of delusions other than the 5 delusions specifically assessed with the BEHAVE-AD were the highest in the subjects with mean MMSE scores of 4.9 ± 4.6, equivalent to the ‘low MMSE group’ studied by Nakatsuka et al. [5].

Comparison of BEHAVE-AD-FW Scores and BEHAVE-AD Incidence Findings for the ‘Residence Is Not Home’ Delusion

In the Japanese study using the BEHAVE-AD-FW [5], it was found that, with respect to individual delusions, the low MMSE AD group (i.e., subjects with MMSE scores of 0–9) had the highest frequency-weighted severity scores for the ‘residence is not home’ delusion and that these frequency-weighted severity delusion scores were significantly greater than those obtained for the intermediate MMSE group (i.e., subjects with MMSE scores of 10–19). Furthermore, the frequency-weighted severity scores of the ‘residence is not home’ delusion for the total sample of Nakatsuka et al.’s [5] study showed a low but significant correlation with total MMSE [and Cognitive Abilities Screening Instrument (CASI) [35]] scores.

These results can be compared to those obtained in the study of the stage-specific incidence of the ‘residence is not home’ delusion published in 1989. The incidence of this specific delusion was found to be about 50% less common in mild AD subjects with a mean MMSE score of 19.1 ± 3.2, than in the moderate and moderately severe AD subjects with mean MMSE scores of 14.1 ± 3.0 and 4.9 ± 4.6, respectively. Specifically, the incidence of this specific delusion increased from 12.5% in the mild AD subjects to 28.6% in the moderate AD subjects and 27.7% in the moderately severe AD subjects. Therefore, we might surmise from the findings of Nakatsuka et al. [5] that, in comparison with AD persons with moderate global severity, the severity and frequency rather than the incidence of this ‘residence is not home’ delusion increase in AD persons with moderately severe global severity. Clearly, all these potentially disturbing phenomena in terms of delusions in persons with AD – their incidence, severity, and frequency – require further study at the present time.

Hence, the examination of the earlier study on the incidence of BPSD symptoms [22] only reinforces the fundamental conclusion of the study of Nakatsuka et al. [5] that the delusions of AD should be studied separately in future treatment and phenomenological studies and that the BEHAVE-AD-FW is a useful approach for such studies. The publication of the study of Nakatsuka et al. [5] is also an apposite occasion for a brief review of where we are at present in terms of the assessment and treatment of BPSD in persons with AD and how we had arrived at the present juncture.

The Development of Modern Global Dementia Severity Measures and Their Relevance for the Recent Findings with the BEHAVE-AD-FW

At the beginning of the 1980s, there was an increasing interest in the treatment of the condition then termed ‘senile dementia’. As a result of the autopsy studies of Tomlinson, Blessed, and Roth [36–38], AD, which had previously been considered to be a rare entity occurring exclu-
sively before the age of 65 years, was beginning to be seen as the major cause of late-life
dementia. Consequently, senile dementia was classified as ‘senile dementia, Alzheimer’s type’,
‘multi-infarct dementia’, or senile dementia secondary to both of these conditions [39]. At that
time, various medications and nutraceuticals were administered and/or approved in different
countries for what was then termed ‘mental decline in the elderly’ or, perhaps more commonly,
‘cerebrovascular insufficiency’. The most popular of these substances were termed ‘vasodi-
lators’, and in the United States and many other nations, the most commonly used vasodilator
was a medication called hydergine, which is a combination of 3 hydrogenated alkaloids of ergot.
One indication of the popularity of hydergine treatment at that time was a study of Medicaid
patients in California, which found that hydergine ranked second in terms of total drug expen-
ditures for all medications in the institutionalized population [40]. Numerous placebo-controlled
studies were purported to have demonstrated the efficacy of hydergine treatment. For example,
a review from 1981 identified 12 such studies [39]. Most of these studies used a rating scale
known as the ‘SCAG’, the Sandoz Clinical Assessment Geriatric Scale [41]. This scale assessed
cognition-related symptoms, functioning, mood-related symptoms, and dizziness, all on a qual-
itative 7-point scale from 1 (absent) to 7 (severe). Despite the popularity of hydergine and the
apparent efficacy of this medication on the SCAG scale, its effects never translated into objective
evidence of efficacy on cognitive tests, and the true utility of hydergine treatment for any recog-
nizable disorder never became clear [42]. Nevertheless, during this period, hydergine was
found to be the 11th most widely prescribed pharmaceutical in the world [43].

Hence, there was a need in the early 1980s for assessment instruments that could
describe: (1) the evolution of the new, increasingly recognized entity, for which the term
‘Alzheimer’s disease’ was suggested; (2) the cognitive symptoms occurring in normal aging
and the evolution of these symptoms as AD progresses, and (3) the functional changes
occurring in normal aging and the evolution of these changes as AD progresses. It was also
clear at that time that behavioral symptoms accompanied the aging of the brain and the AD
process. These symptoms also needed a better definition, and their relationship to the AD
process needed to be described.

One response to these needs was to carefully observe ‘normal aged’ and AD persons. The
objective was to establish symptomatic commonalities as well as differences. The observa-
tions were informed using the assessment instruments available at that time, such as the
Inventory of Psychic and Somatic Complaints in the Elderly (IPSC-E), a measure developed at
the US National Institute of Mental Health by Allen Raskin [44] and Thomas Crook and first
used in a hyperbaric oxygen study with elderly subjects conducted in association with inves-
tigators at the New York University School of Medicine [45].

At that time, in the late 1970s and early 1980s, a 7-point rating scale was used at the
former Geriatric Study and Treatment Program of the Department of Psychiatry of the New
York University Medical Center to score the severity of cognitive decline or impairment.
This scale ranged from no cognitive decline (a score of 1) to very severe cognitive decline
(a score of 7). Subjects were scored on that continuum based upon their performance on
assessments from the Guild Memory Tests [46], WAIS vocabulary scores [47], age, and
other factors. After some years of careful systematic observation, it was possible to describe
comprehensive clinically relevant concomitants of the 7 psychometric levels. The compre-
hensive clinically relevant descriptions included the most outstanding cognitive, functional,
affective, and behavioral disturbance characteristics of each level. These manifestations
were published together with the psychometric descriptions and emerging brain change
data from computerized tomographic (CT) and $^{18}$F-2-deoxy-2-fluro-D-glucose positron
emission tomographic (FDG PET) scan data as the global deterioration scale for primary
degenerative dementia [7]. As stated in the first line of the 1982 GDS publication, primary
degenerative dementia was the terminology used at that time as the clinical ‘equivalent to
the neuropathologically defined Alzheimer’s disease [7]. This GDS publication, describing the course of normal brain aging, the entities subsequently termed subjective cognitive impairment, mild cognitive impairment, and mild, moderate, moderately severe, and severe AD, was released 2 years before the publication of the McKhann et al. criteria [32] in 1984. The criteria proposed in McKhann et al. [32] made it scientifically acceptable to refer to the clinical entity of AD, which, despite Alois Alzheimer’s classic clinical case description [48], was at that time considered to be exclusively a diagnosis made after a postmortem study of the brain.

As noted above, the GDS is comprised of clinical descriptions of the 7 levels of AD development. The descriptions are in the form of words and phrases, which describe progressive changes succinctly in terms of the cognitive, functional, affective, and behavioral disturbance changes which occur with the progression of what is still frequently seen as ‘normal brain aging’ as well as of MCI and AD. Apart from traditional reliability studies in ambulatory patients, ambulatory research center, long-term care facility, and nursing home settings [for a review, see 9], the specific content validity of the phraseology in the GDS has been the subject of independent investigation. The specific content validity of the GDS was assessed in a unique study [50] conducted under the direction of John Overall, who is also the developer of the Brief Psychiatric Rating Scale (BPRS), which has been (and perhaps is still, at the time of this publication) the most widely used rating scale in schizophrenia research for many years [49]. The phraseology of the GDS was divided into 30 component phrases. Spouses of AD patients and clinic attendees without dementia were asked to assess the presence or absence of symptomatology in terms of the GDS phrases. Using a statistical technique known as the principal components analysis, a new scale was constructed. The investigators found that the phrases tended to cluster into naturally occurring stages or phases. The specific hierarchy of the clustering of the phrases very closely mirrored that of the GDS assessment [50].

The GDS has been very widely used throughout the world. For example, of the 4 medications currently used and approved by the US Food and Drug Administration (FDA) for the treatment of AD, the GDS was used in the pivotal trials associated with worldwide approvals for 2 of these medications, namely, rivastigmine [15, 16] and memantine [17]. Also, since 2004, the Alzheimer’s Association, the leading advocacy organization for AD in the world, has prominently posted the GDS on its website to explicate the course of brain aging and AD for the world [51].

**Relevance of GDS Stage 7 (Severe Dementia) for the Interpretation of the BEHAVE-AD-FW Findings of Nakatsuka et al. [5]**

The GDS includes several stages of AD which were not studied by Nakatsuka et al. [5] and which could add to their reported findings. One of these stages is GDS stage 7, i.e., severe dementia. This is a stage in which ‘verbal abilities are lost’ and in which ‘patients are incontinent’; eventually, they also ‘lose the ability to walk’. This final stage of AD is important in that a substantial proportion of all AD patients are in this stage [52] and much of the burden of AD occurs in this stage. A series of derivative measures were developed from the GDS, specifically, the Brief Cognitive Rating Scale (BCRS), assessment measures and the FAST scale [20, 53]. These measures, particularly the FAST procedure, permit the detailed characterization of this final stage of AD, in which 6 successive functional substages, each with a mean duration of 1 year or more, are identifiable [10, 54, 55]. The FAST procedure also identifies 5 functional substages corresponding to the moderately severe dementia stage (stage 6 on the GDS). Hence, it is possible to identify a total of 11 successive functional substages corresponding to what is commonly referred to simply as ‘severe dementia’. These detailed assessments of the moderately severe and severe dementia stages of AD serve many purposes. One
of these purposes is that understanding the continuing process of AD can potentially decrease the neglect of the affected individuals and can potentially be of importance for providing quality care, remediating deficits, and reducing human suffering, neglect, and distress [56, 57].

The recognition of these later stages and substages of AD is also of great relevance for a proper scientific review of BPSD and, more particularly, delusional phenomena with the evolution of AD. The study of Nakatsuka et al. [5] did not exclude subjects with severe AD (GDS and FAST stage 7). Nevertheless, on the basis of their findings, it is likely that their sample included very few, if any, subjects with this dementia severity. The study by Reisberg et al. [22] of the stage-specific incidence of BPSD included 17 subjects in GDS stage 7. All of these 17 GDS stage 7 subjects had MMSE scores of 0 (bottom). An essential conclusion of that study was that "all [BPSD] symptoms were noted less frequently in GDS stage 7 than in GDS stage 6 subjects" [22]. Furthermore, no delusional symptoms were noted in any of the GDS stage 7 (severe dementia) subjects. This is, of course, not surprising since even early GDS stage 7 subjects are largely non-verbal, as described in detail with the FAST procedure. Nevertheless, this perspective and review is of obvious importance in the interpretation of the findings of the recent Japanese BEHAVE-AD-FW study [5]. The latter study included subjects with DSM-III-R dementia [34] and a CDR score of ≥1 [8]. There were no explicit exclusions of more severe or GDS stage 7 subjects. To the extent that subjects in the severity range of GDS stage 7 may have been included in Nakatsuka et al.'s outpatient, Taji Clinic sample, their data would have been biased towards an absence of delusions in the more severe subject sample(s). Hence, the absence of any lower-range dementia severity exclusion makes the findings more difficult to interpret but does not diminish their importance. This is particularly the case as the final CDR stage, CDR stage 3, includes both the severity range of late GDS stage 6 subjects as well as the severity range of all GDS stage 7 subjects [8, 11, 14, 58]. Therefore, even a retrospective analysis using the CDR staging would not be useful in identifying those subjects who had entered the severity range equivalent to GDS stage 7 and who consequently do not any longer express delusional symptomatology, either because they no longer have the verbal skills to express these symptoms or because they no longer have cognitive ideation associated with delusions. In the study of Nakatsuka et al. [5], the MMSE might, however, be useful in the retrospective identification of subjects who might have had the equivalent of a GDS stage 7 level of cognition. Studies have indicated that most (but not all) GDS stage 7 subjects have MMSE scores of 0 (bottom) and most (but not all) late GDS stage 6 subjects achieve an MMSE score of >0 [23]. Therefore, the authors of the Tajiri Clinic study might wish to retrospectively examine their data further in a subsequent study excluding subjects with an MMSE score of 0 (bottom). In this future retrospective examination of their data, the investigators might find an even stronger effect with respect to the 'residence is not home' delusion increasing with the advance of AD severity within the range of MMSE scores from 30 to 1.

Relevance of GDS Stage 3 (MCI) for the Understanding of BPSD (Including Delusions)

The Tajiri Clinic study of delusional symptoms [5] also did not include subjects with MCI. MCI is a condition initially characterized and used in association with the GDS stage 3 [7, 59–61]. This GDS stage 3 was initially described in part as a condition in which 'the earliest clear-cut clinical deficits appear... however, objective evidence of memory deficit is obtained only through an intensive interview' [7]. It was noted that in this stage, 'decreased performance becomes manifest in demanding employment and social situations' and 'coworkers become aware of the [person’s] relatively poor performance’ [7]. This GDS stage 3 was distinguished from the preceding GDS stage 2 in which 'the [person] complains of memory deficit. Most frequently [persons]... complain of forgetting where familiar objects have been placed and of forgetting names that they formerly knew well. There is no objective evidence of
memory deficit in the clinical interview and no objective deficits in employment or social situations. The individual ... displays appropriate concerns about the symptoms' [7].

The GDS stage 3, for which the terminology 'mild cognitive impairment' was coined, was also distinguished from a subsequent stage, later characterized as the stage of mild dementia or mild AD [9]. In this GDS stage 4 of mild dementia, 'a clear-cut deficit is apparent in a careful clinical interview. The deficits are apparent in many areas... [Persons] display decreased knowledge of recent events in their own lives and current events in the world around them' [7]. In addition, in this GDS stage 4, individuals 'can no longer perform complex tasks accurately and efficiently' [7]. Careful comparison with the CDR staging phraseology indicates that the CDR 0.5 stage appears to encompass both the entities currently referred to as MCI and also mild dementia or mild AD [9].

In the initial GDS publication from 1982, the GDS stage 3 was referred to as a stage of 'mild cognitive decline'. In 1988, data were published indicating that the cohort in the GDS stage 3 performed significantly more poorly than the cohort in the preceding GDS stage 2 on the MMSE [33], on each of the tests developed by Blessed et al. [62] (including the Blessed Dementia Scale, the Blessed Information Test, the Blessed Memory Test, and the Blessed Concentration Test) and on 10 of 11 psychometric tests studied [59]. Hence, it was clear that GDS stage 3 patients not only experienced a decline but were impaired. Therefore, beginning in 1988, the terminology 'mild cognitive impairment' was used in reference to GDS stage 3 subjects [59–61].

Nakatsuka et al. [5] did not study subjects with MCI. However, a broader view of their findings should include them. The 1989 incidence study of symptoms identified with the BEHAVE-AD examined otherwise healthy individuals with MCI [22]. Of 14 MCI subjects in the sample, only 1 subject manifested a delusion, specifically the delusion of theft (i.e., that 'people are stealing things'). Hence, this study of otherwise healthy individuals with MCI does not indicate that delusions are common in individuals with MCI. This is in marked contrast to several other behavioral symptoms that can be very common in individuals with MCI. For example, anxiety with regard to upcoming events, e.g., the person repeatedly saying 'When are we going?', was found to occur in 43% of MCI subjects, and 'other anxieties' occurred in 46% of MCI subjects. Also, more than 20% of subjects with MCI manifested purposeless activity, non-verbal anger or negativity, tearfulness, and depressed mood (frequently manifested by negative statements, e.g., the person saying 'I wish I were dead'). Although beyond the scope of the study from Nakatsuka et al. [5], these symptoms could informatively be studied using the frequency-weighted severity approach employed in the Japanese Tajiri Clinic publication. Also, it would be useful to study these symptoms in MCI on an individual basis rather than simply as collective entities, again, in accord with the approach taken by Nakatsuka et al. [5] for delusional symptoms.

Relevance of GDS Stage 2 (SCI) for the Understanding of BPSD (Including Delusions)

It is now increasingly recognized that there is a behavioral stage in the development of AD before the MCI stage. This stage is described in the GDS as GDS stage 2 [7]. It is a stage, in which individuals manifest subjective deficits in the absence of objective deficits on mental status examination or psychometric testing. In otherwise healthy individuals, longitudinal studies have demonstrated that this stage, now termed the stage of SCI, eventuates in AD. More specifically, in 1986 it was estimated that the pre-MCI, GDS stage 2, i.e., the SCI stage, lasts a mean of 15 years prior to the advent of MCI [54]. At that time, it was also estimated that the MCI stage lasts a mean of 7 years prior to the advent of the mild dementia stage of AD [54]. These temporal estimates from 1986 were based upon clinical observations as well as epidemiologic data available at that time with respect to the prevalence of dementia, AD, and cognitive decline.
In the case of MCI, the temporal duration estimate published in 1986 [54] was supported initially by an approximately 4-year longitudinal study published in 1999 [63]. Subsequently, studies in clinic attendees around the world found a mean conversion rate to dementia in MCI subjects of 15.01% per year [64]. This conversion rate is within 1% of the conversion rate which would have been anticipated if the clinic attendees had been evenly distributed throughout the MCI stage at baseline and if MCI is a stage lasting exactly a mean of 7 years [i.e., for a 7-year stage and an evenly distributed population, one would expect a conversion rate to dementia of 14.29% per annum; the difference between the observed annual conversion rate of 15.01% per annum in clinic attendees in published studies (from the 2004 review) [64] and the hypothesized conversion rate of 14.29% (originally posited in 1986) [54] for a stage lasting precisely 7 years is 0.72%, i.e., less than a 1% difference!].

The temporal duration of the SCI stage has also been examined in a prospective longitudinal study. Specifically, 20 years after the 15-year estimate of the duration of the SCI stage was first published in 1986 [54], Prichep et al. [65] published a 9-year prospective longitudinal study of outcomes in SCI subjects. In this study, 44 SCI (GDS stage 2) subjects were followed over a mean of 8.9 ± 1.8 years. Assuming that the subjects in this study were uniformly distributed throughout the GDS stage 2 at baseline, an analysis indicated that the hypothesized duration of the SCI stage based upon the 1986 temporal estimate [54] and the observed duration of the SCI stage based upon the 2006 study of Prichep et al. [65] differed by only 2% [66]. Specifically, Prichep et al. [65] observed that 27 of the 44 SCI subjects declined to MCI or, in some cases, to dementia over the subsequent 8.9-year mean interval. Assuming a perfectly uniformly distributed baseline GDS stage 2 population and a stage lasting exactly 15 years, it was hypothesized that 59.33% of the subjects would convert to a more severe stage over 8.9 years (8.9 years × 6.666%/year conversion rate for a stage lasting precisely 15 years = 59.33%). The observed conversion rate in the study from Prichep et al. [65] was 61.36% (27 converters to MCI or dementia from a sample of 44 subjects = 61.36%). The difference between the observed change rate and the hypothesized change rate was 2.03%. Hence, the results from this study strongly support the 1986 hypothesis that SCI is a stage in the eventual development of MCI and subsequent dementia, and, more particularly, the dementia of AD. Of the decliners in this study, 20 declined to an MCI diagnosis (GDS stage 3) and 7 declined to a dementia diagnosis. Of those subjects who received a dementia diagnosis, 6 fulfilled the NINCDS-ADRDA criteria for AD and 1 subject was diagnosed as having dementia associated with cerebrovascular disease [65]. Therefore, the longitudinal study of Prichep et al. [65] indicates that the hypothesis that SCI is a stage lasting approximately 15 years prior to the MCI stage of eventual AD in otherwise healthy individuals is correct.

In addition to the longitudinal data indicating that SCI is a stage in the development of AD, there is, at the present time, considerable physiologic data indicating that the changes seen in SCI are on a continuum with those observed in MCI and AD (reviewed in Reisberg et al. [67]). For example, apart from significant differences between SCI and MCI subject groups, which have long been observed (e.g., Reisberg et al. [59]), recent studies have noted physiologic differences between SCI and 'No Cognitive Impairment' (NCI; GDS stage 1) subject groups. For example, decrements in the cerebral metabolism of SCI subjects have been observed compared with NCI subjects [68]. A very different physiologic parameter, i.e., an abnormal CSF Aβ42-to-tau ratio, has been found in SCI subjects compared with similarly aged NCI subjects [69]. Similar differences between individuals with SCI and individuals with NCI have also been reported in terms of hippocampal volume [70] and cerebral amyloid deposition measured by carbon 11-labeled Pittsburgh compound B PET imaging [71], to cite some of the major examples from this rapidly accumulating literature.

In addition to the studies establishing the temporal duration of the SCI stage and the physiologic differences between SCI and NCI, as well as SCI and MCI subjects, one other
important area of investigation has been published pertaining to the demonstration of the NCI to SCI, to MCI, to dementia (especially AD) continuum. This is the demonstration of the differences in outcomes between pre-MCI persons, SCI persons, and persons without SCI (i.e., GDS stage 1). Specifically, a 7-year prospective longitudinal study compared the outcome in healthy older persons free of subjective complaints of cognitive impairment, defined as being in GDS stage 1, with the outcome in otherwise healthy older persons with SCI, defined as being in GDS stage 2. The results indicated that otherwise healthy community-residing older persons presenting to an outpatient cognitive research center with SCI are much more likely to manifest subsequent decline than similarly aged [and educated] persons presenting without these complaints [72]. This study found that ‘with the Weibull proportional hazards model that controlled for baseline demographic variables [age, gender, and education] and follow-up time and examined the hazard rates of progressing… the hazard ratio of SCI subjects to NCI subjects was 4.5 (95% confidence limits, 1.9 to 10.3)’ [72]. Additionally, when subjects declined, the SCI subjects declined more quickly than the NCI subjects (p < 0.0001, for both the Wilcoxon test and the log-rank test). Using an accelerated failure time model that controlled for age, gender, education, and follow-up time and examined the time to decline on the basis of the baseline group, it was found that SCI subjects took an average of 60% shorter time to decline than NCI subjects.

Other studies have also found that subjective complaints of memory/cognitive impairment are associated with a higher risk of cognitive decline or dementia [e.g., 73–77]. For example, Wang et al. [74] found that what they termed ‘subjective memory deterioration’ was associated with cognitive decline and incident dementia. They observed a hazard ratio for incident dementia at the age of 70 years of 6.0 (95% CI = 2.1–18), a very similar finding to that of Reisberg et al. [72] described above. Therefore, it is now very clear from multiple perspectives that there is an SCI stage of eventual AD prior to MCI. However, what are the implications of this stage in terms of the assessment of delusions and other behavioral disturbances in AD?

In the 1989 study of the stage-specific incidence of potentially remediable behavioral symptoms [22], 15 subjects in the pre-MCI stage, i.e., the SCI stage, were evaluated. All of these subjects were healthy older persons who fulfilled the medical exclusion criteria from McKhann et al. [32] for probable AD, except, of course, that they did not have dementia. This means that an extensive workup was performed on these subjects including corpuscular blood counts, a comprehensive metabolic panel, testing for thyroid abnormalities, assessment of serum B12 and folate, testing for syphilis, and an MRI brain scan, as well as an extensive medical history and psychiatric, neurologic, physical, and medical examinations. Subjects with significant medical, neurological, or psychiatric conditions that might interfere with cognition, or progress or recur so as to interfere with cognition, were excluded. These exclusions would potentially apply to subjects who presented with delusional disorder or psychosis. Hence, if SCI subjects were assessed as presenting with a delusional disorder or psychosis, they would have been excluded from the 1989 study of the stage-specific incidence of behavioral symptomatology [22]. However, if the subjects were assessed as, apart from subjective complaints of cognitive impairment, otherwise healthy, they were included in the study. Therefore, it is possible that this study may have underestimated the true incidence of delusions in the SCI stage of eventual AD.

Of the 15 subjects studied in the SCI stage in the study from Reisberg et al. [22], 2 subjects presented with symptomatology in the Paranoid or Delusional Ideation category of the BEHAVE-AD. Both subjects had ‘suspiciousness or paranoia, apart from the specific delusional types identified by the BEHAVE-AD’. The most common SCI behavioral symptom identified in this study was tearfulness, which was present in one-third of the subjects studied in the SCI stage. Anxieties and phobias were also common. The specific anxiety identified by the
BEHAVE-AD because of its frequency in AD, i.e., ‘anxiety with respect to upcoming events’, occurred in 1 of the 15 SCI subjects studied, and ‘other anxieties’ occurred in more than a quarter of these subjects. The ‘fear of being alone’, which is very common in GDS stages 5 and 6 AD persons, was not observed in the SCI subjects studied. However, one-fifth of the SCI subjects manifested ‘other phobias’.

Hence, even the very early SCI stage of eventual AD, which begins more than 20 years before dementia becomes evident, appears to be occasionally accompanied by ‘paranoid and delusional’ type symptoms identified with the BEHAVE-AD, as well as by other BPSD. This SCI stage can also be effectively studied using the frequency-weighted severity approach employed in the investigation by Nakatsuka et al. [5] of persons with overt dementia associated with AD.

The Development of Behavioral Assessments for AD: History and Background of the Development of the BEHAVE-AD

The history of the development of the GDS has already been discussed in this review. The GDS notes the major clinical symptomatology of the stages of AD, including cognition, functioning, and behavioral disturbance-related symptoms. As discussed previously in this review, the GDS, published in 1982, includes the description of pre-dementia AD stages, including the MCI stage, which only began to be widely recognized in 2001 (almost 20 years after the GDS description), and the SCI stage, which appears to be increasingly recognized only now in 2013 (more than 30 years after the GDS description) as a clinical stage of eventual AD. Also, the GDS includes a severe stage, which, as evidenced by the study of Nakatsuka et al. [5], is still widely ignored at the present time. Although the GDS notes behavioral symptomatology appropriately in the description of the stages, the criteria for the selection of GDS stages are based upon cognition and functioning.

As also reviewed in the preceding sections of this publication, the GDS staging descriptions were subsequently elaborated in more detailed descriptions of the cognitive and functional evolution of AD [20, 53, 54, 78, 79]. The BCRS and FAST axes were developed and enumerated to be optimally concordant with the corresponding GDS stages as well as with each other [9, 20, 53, 78, 79]. Since the creation of the GDS, it has been apparent that the mood and remediable behavioral disturbance elements of the stages of AD also needed elaboration. The ‘mood axis’ of the BCRS was developed and published, describing the most prominent mood and behavioral changes in each of the GDS stages [80]. Through systematic observations, the most prominent psychosis like symptoms in AD were identified, and a rating scale for these symptoms was developed and published [81]. Shortly afterward, the broader aspects of potentially remediable behavioral symptoms in AD began to be described in detail [21]. Finally, in 1987, a rating scale for the assessment of potentially remediable behavioral and psychological symptoms in AD and the disturbances associated with the symptoms was published: the BEHAVE-AD [1]. This scale assesses 25 characteristic potentially remediable, behavioral and psychological symptoms individually on a 4-point severity scale in AD persons in the following 7 categories: (A) Paranoid and Delusional Ideation, (B) Hallucinations, (C) Activity Disturbances, (D) Aggressiveness, (E) Diurnal Rhythm Disturbances, (F) Affective Disturbances, and (G) Anxieties and Phobias. The BEHAVE-AD publication [1] contained data on the apparent responsiveness of the BEHAVE-AD symptoms to an antipsychotic (neuroleptic) medication in wide usage at that time, namely, thioridazine.
The BEHAVE-AD

The BEHAVE-AD was unique and a major step forward in AD assessment in that it (1) comprehensively identified and assessed in detail potentially neuroleptic responsive AD symptoms and (2) excluded cognition- and function-related symptoms. The importance and rationale for separating BPSD from cognitive and functional symptoms of dementia was subsequently discussed and reviewed extensively in a separate publication [19]. The BEHAVE-AD assessments were demonstrated to be reliable [24, 82–84], and their reliability was found to be excellent and similar to the reliability of the most reliable cognitive dementia assessments’ [85] (see table 1 for a summary of reliability studies and other aspects of the BEHAVE-AD assessment instrument).

For example, the BEHAVE-AD was assessed for reliability in a study which compared clinic interrater reliability, telephone versus clinic same-rater reliability, and telephone versus clinic interrater reliability [84]. The 17 subjects in this study ranged in severity from MCI (GDS stage 3, n = 5) to mild, moderate, or moderately severe AD (GDS stages 4–6, n = 12). The interrater reliability coefficient (intraclass correlation coefficient, ICC) for the clinic interrater study, in which 2 clinicians scored simultaneously, blind to each other’s ratings and any previous ratings, was 0.98 (p ≤ 0.001). For the telephone versus clinic rater evaluations, the clinicians scored on different days, and the evaluations were separated by a mean of 6.2 ± 5.6 days. For the same clinician, who was blind to all previous evaluations, the ICC for the evaluation by phone and, on a different day, in person was 0.89 (p ≤ 0.001). For both clinicians, each scoring on different days, one by phone and, subsequently, the other in person, the ICC was 0.94 (p ≤ 0.001).

Another reliability study of the BEHAVE-AD was conducted in a nursing home in a suburb of Paris, France, using a French translation of the BEHAVE-AD [83]. All subjects (n = 20) were native-born French speakers. Their mean age was 75.5 ± 9.2 years (range 58–90), and their severity ranged from GDS stage 4 to GDS stage 7. The MMSE scores ranged from 0 to 25, with a mean score of 11.4 ± 7.1. The subjects’ diagnoses included multi-infarct dementia (n = 2), dementia ‘due to other causes’ (n = 3), and probable AD (n = 15). The subjects were evaluated by 2 experienced geropsychiatrists. The designated interviewer elicited the relevant information, and both clinicians separately rated the responses. After the designated clinician obtained all information, the second rater had the opportunity to clarify any uncertainties by questioning the caregiver. The designated interviewer was alternated. The ICCs for the BEHAVE-AD total scores were 0.96 (p < 0.01) for the rater consistency coefficient and 0.95 (p < 0.01) for the rater agreement coefficient.

Subsequently, the BEHAVE-AD assessments were widely used. Perhaps most importantly for the history of BPSD treatment in AD, 3 major multicenter phase 3 clinical trials of Risperdal (risperidone) for the treatment of BPSD in AD were conducted [25–27]. A review with a pooled analysis of the 3 trials has been published [28]. The review notes that ‘the three trials had similar designs and used the same well validated and accepted scales… to evaluate neuropharmacological interventions in patients with BPSD’. The outcome measures were the BEHAVE-AD and the Cohen-Mansfield Agitation Inventory (CMAI) [86]. One of these measures was the primary outcome measure in the studies, and the other measure was the secondary outcome measure, depending upon the trial. In the 3 controlled trials, subjects assigned to the active treatment received risperidone for 12 weeks in dosages ranging from 0.25 mg daily to 1.0 mg b.i.d. (2 mg per day). ‘In general, the starting dose of risperidone was 0.25 mg bid.’ The 3 randomized controlled trials found that ‘at the recommended doses, risperidone displayed a favorable risk-benefit profile’ [28]. ‘The mean dose of risperidone across all trials was 1.0 mg/day.’ The mean change at the end point was significantly higher for risperidone than for placebo in the pooled analysis (–6.1 vs. –3.6 for BEHAVE-AD total scores (p < 0.001) and
A Conceptual framework

The BEHAVE-AD [1] differs from all prior scales in the fundamental principles of item selection. These are:
1. The BEHAVE-AD was designed to specifically reflect and measure the characteristic behavioral symptoms which occur in AD.
2. All items are designed to be optimally independent of the comparatively very-difficult-to-treat cognitive symptomatology of AD.
3. All items reflect behaviors that are potentially disturbing to caregivers.
4. All items were selected to reflect behaviors which were likely to be potentially remediable through the use of psychopharmacologic medications which were available at the time of the publication of the BEHAVE-AD in 1987 [1].

B Methodology

The BEHAVE-AD is comprised of 25 symptomatic items describing behavioral disturbances. The symptomatic items fall within 7 categories. Each symptomatic item is assessed on a 4-point severity scale. On this scale, a score of ‘0’ indicates that the item is not present; a score of ‘1’ indicates the presence of the symptom; a score of ‘2’ indicates that the symptom is present, generally with an emotional component, and a score of ‘3’ indicates that the symptom is present, generally with an emotional and physical component. Information regarding the symptoms is obtained by interviewing one or more knowledgeable informant(s). The magnitude of symptomatic severity is assessed over the prior 2-week interval. The 25 symptoms fall into the following categories: (A) Paranoid and Delusional Ideation (7 items); (B) Hallucinations (5 items); (C) Activity Disturbances (3 items); (D) Aggressiveness (3 items); (E) Diurnal Rhythm Disturbances (1 item); (F) Affective Disturbances (2 items), and (G) Anxieties and Phobias (4 items). Total scores for each of the 7 categories and for the entire rating scale are obtained. In addition, a global scale rating is obtained of the magnitude of the BEHAVE-AD-assessed symptomatology. This global scale also utilizes a 4-point assessment of symptomatic severity from ‘0’ = not at all troubling to the informant(s) [or caregiver(s)] or dangerous to the subject, to ‘4’ = severely troubling to the informant(s) [or caregiver(s)] or dangerous to the subject.

C Construct validity

The BEHAVE-AD symptoms uniformly differ from cognitive and functional symptoms which occur in AD.

<table>
<thead>
<tr>
<th>Cognitive and functional symptoms in AD</th>
<th>BEHAVE-AD-identified behavioral disturbance symptoms in AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>All these symptoms demonstrate continued worsening with the advance of AD.</td>
<td>All these symptoms peak in occurrence and magnitude of severity at some stage prior to the final GDS and FAST scale stage 7 of AD [22].</td>
</tr>
<tr>
<td>The origin of these symptoms relates strongly to the evolution of the brain pathology in AD (e.g., AD-related neurofibrillary pathology, progressive neurometabolic decrement, etc.).</td>
<td>The origin of these symptoms relates strongly to environmental factors, interpersonal factors, sensory changes (especially in the case of hallucinations), and neurochemical brain changes with the advance of AD.</td>
</tr>
</tbody>
</table>

D Reliability

<table>
<thead>
<tr>
<th>Authors [Ref.]</th>
<th>Methodology</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patterson et al. [82]</td>
<td>Assessment of percent agreement on BEHAVE-AD items.</td>
<td>Percent agreement for 23 of the 25 BEHAVE-AD symptomatic items ranged from 82 to 100%. For ‘tearfulness’, the percent agreement was 76% and for ‘auditory hallucinations’, the percent agreement was 78%. For the entire 25-item scale, the percent agreement was 90%.</td>
</tr>
</tbody>
</table>
Table 1 (continued)

<table>
<thead>
<tr>
<th>Authors [Ref.]</th>
<th>Methodology</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levy et al. [24]</td>
<td>Sequential interviews of caregivers by a geriatrician and a geropsychiatrist.</td>
<td>Spearman rank correlation coefficients were 1.0 for the BEHAVE-AD total score and the global assessment item.</td>
</tr>
<tr>
<td>Sclan et al. [83], Study 1</td>
<td>Interrater reliability with 2 post-doctoral fellow clinicians rating simultaneously. The designated interviewer elicited information. After the first rater obtained all information, the second rater had the opportunity to clarify any uncertainties by questioning the informant. The designated rater was alternated for each subsequent subject. ICCs were calculated using the method of Shrout and Fleiss [121]. Both rater consistency coefficients and rater agreement coefficients were calculated. Subjects were community-residing individuals enrolled at the New York University Aging and Dementia Research Center and ranged in severity from SCI [66] to moderately severe dementia (GDS stage 6).</td>
<td>Rater consistency coefficients for the 7 BEHAVE-AD categories were all statistically significant (p &lt; 0.01), and the correlations for these 7 categories ranged from 0.65 to 0.91. For the total BEHAVE-AD scores, the rater consistency ICC was 0.96 (p &lt; 0.01). Rater agreement coefficients for the 7 BEHAVE-AD categories were all statistically significant (p &lt; 0.01), and the correlations for these 7 categories ranged from 0.65 to 0.91. For the total BEHAVE-AD scores, the rater agreement ICC was 0.96 (p &lt; 0.01).</td>
</tr>
<tr>
<td>Sclan et al. [83], Study 2</td>
<td>Interrater reliability with 2 experienced, French-speaking geropsychiatrists trained in the procedures, using a French translation of the BEHAVE-AD. All subjects and informants were native-born French speakers, and the subjects were residents of a nursing home in a suburb of Paris, France, where the interviews were conducted. As in study 1, the designated interviewer elicited information relevant to the BEHAVE-AD, and both clinicians separately rated the responses. After the first rater completed the interview, the second rater had the opportunity to clarify any uncertainties, with the first rater still present to hear the responses. The designated interviewer was alternated for consecutive informant interviews, and the number of designated interviews completed by each geropsychiatrist was similar. As in study 1, two ICCs were calculated using the method of Shrout and Fleiss [121]. These were rater consistency coefficients and rater agreement coefficients. The severity level of the subjects ranged from mild (GDS stage 4) to severe dementia (GDS stage 7), and the MMSE scores ranged from 0 to 25 (mean ± SD = 11.4 ± 7.1). Subject diagnoses were probable AD (75%), multi-infarct dementia (10%), and dementia due to other causes (15%).</td>
<td>Rater consistency coefficients for the 7 BEHAVE-AD categories were all statistically significant (p &lt; 0.01), and the correlations for these 7 categories ranged from 0.80 to 0.99. For the total BEHAVE-AD scores, the rater consistency ICC was 0.96 (p &lt; 0.01). Rater agreement coefficients for the 7 BEHAVE-AD categories were all statistically significant (p &lt; 0.01), and the rater agreement ICCs for these 7 categories ranged from 0.78 to 0.99. For the total BEHAVE-AD scores, the rater agreement ICC was 0.95 (p &lt; 0.01).</td>
</tr>
</tbody>
</table>
Table 1 (continued)

<table>
<thead>
<tr>
<th>Authors [Ref.]</th>
<th>Methodology</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monteiro et al. [84]</td>
<td>Telephone evaluation of the BEHAVE-AD was compared with in-person assessment using the same rater and in-person assessment using a different rater. Evaluations were performed by experienced clinicians. ICCs were calculated to establish intrarater and interrater reliability. The two-way random-effect model recommended by Shrout and Fleiss [121] was used. The mean time interval between the telephone contact and the clinic evaluation was 6.2 days (SD = 5.6), with a range of 2–30 days. ICCs were obtained using BEHAVE-AD category sum scores and total scale scores. All subjects had knowledgeable informants. Severity ranged from MCI to moderately severe dementia (GDS stage 6). Two studies were conducted. In both studies, all evaluations were conducted blind to any prior information or data on the subjects. In study 1, the clinician who had performed the telephone interview (clinician 1) rated the subject again in person through an interview performed by a different clinician (clinician 2). In study 2, the ratings performed on the telephone by clinician 1 were compared with the ratings performed by clinician 2 in person. ICCs for 1 of the 7 BEHAVE-AD categories (Hallucinations) could not be computed because of absence of variance. In study 1, i.e., the telephone versus clinic same rater evaluation, the ICC for the 6 evaluable BEHAVE-AD categories ranged from 0.65 to 0.96 (all p values ≤0.001). For the BEHAVE-AD total scale scores for study 1, the ICC was 0.89 (p ≤ 0.001). In study 2, i.e., the telephone versus clinic interrater evaluation, the ICC for the 6 evaluable BEHAVE-AD categories ranged from 0.48 to 0.86 (p values for 5 of the 6 evaluable BEHAVE-AD categories were ≤0.001; for the remaining category Anxieties and Phobias, the p value was ≤0.05). For the BEHAVE-AD total scale scores in study 2, the ICC was 0.94 (p ≤ 0.001).</td>
<td></td>
</tr>
</tbody>
</table>

E  Criterion validity
1. Symptomatic responsivity to nonpharmacological intervention independent of the effects on cognition

<table>
<thead>
<tr>
<th>Authors [Ref.]</th>
<th>Methodology</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levy et al. [24]</td>
<td>Single-blind, dose escalation study of the efficacy of the non-benzodiazepine anxiolytic medication buspirone in the treatment of disruptive behavior in subjects with possible AD [32]. After a 2-week washout phase (4 weeks for chronic benzodiazepine users), subjects were given placebo during week 1.</td>
<td>All 7 BEHAVE-AD categorical symptomatic scores as well as the total BEHAVE-AD score and the BEHAVE-AD global score manifested decrements, indicating less symptomatology and/or disturbance in response to the placebo 'intervention'. Hence, each of the 7 BEHAVE-AD categorical elements responded in a syndromic manner to the placebo intervention. BEHAVE-AD total severity scores declined by 25% and BEHAVE-AD global scores decreased by 14% in response to the placebo intervention.</td>
</tr>
</tbody>
</table>
Table 1 (continued)

<table>
<thead>
<tr>
<th>Authors [Ref.]</th>
<th>Methodology</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Katz et al. [25]; Monteiro et al. [103]; Reisberg et al. [100]</td>
<td>Randomized, double-blind trial in institutionalized elderly with dementia and diagnoses of AD, vascular dementia, or both. Subjects were randomized to medication or placebo for 12 weeks. The study was completed by 70% of patients, and an endpoint analysis was conducted on 163 patients randomized to placebo.</td>
<td>All 7 BEHAVE-AD symptomatic categories showed significant ( (p &lt; 0.01) ) reductions in total scores in comparison with baseline with the placebo 'treatment'. For 5 of the symptomatic categories, the significance level was ( p &lt; 0.001 ). The magnitude of symptomatic reduction from baseline to endpoint in the placebo-treated subjects ranged from 18% (for Activity Disturbances) to 45% (for Anxieties and Phobias). Total BEHAVE-AD scores decreased approximately 25% in the placebo-treated patients. Hence, each of the 7 BEHAVE-AD symptomatic categories responded in a syndromic (running together) manner to the placebo intervention in this large multicenter trial with 40 separate sites throughout the United States.</td>
</tr>
</tbody>
</table>

2. Symptomatic responsivity to pharmacological intervention independent of the effects on cognition

<table>
<thead>
<tr>
<th>Authors [Ref.]</th>
<th>Methodology</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levy et al. [24]</td>
<td>See above. Buspirone was administered subsequent to the washout and week 1 placebo treatment condition in escalating doses: weeks 2 and 3, 15 mg daily; weeks 4 and 5, 30 mg daily; weeks 6 and 7, 45 mg daily; and weeks 8 and 9, 60 mg daily.</td>
<td>The optimal treatment response was observed with the 30-mg daily buspirone dosage. For this dosage and the 45-mg daily buspirone treatment, BEHAVE-AD global scores were significantly improved in comparison with the placebo treatment condition. Six of the 7 BEHAVE-AD symptomatic categories showed improvement in scores in comparison with the placebo condition. The remaining symptomatic category (Affective Disturbances) showed no change from the placebo condition. Hence, this anxiolytic treatment appears to have produced a syndromic incremental treatment response in 6 of the 7 BEHAVE-AD treatment categories in comparison with the placebo condition. The effects appear to have been independent of cognitive change. No significant changes on the MMSE were observed over the course of this study.</td>
</tr>
</tbody>
</table>
See above. Significant effects of the risperidone treatment in comparison with the placebo treatment were observed at two of the three risperidone medication dosages, i.e., 1 and 2 mg daily. Specifically, total BEHAVE-AD scores improved in these subjects at 1 (n = 148) and 2 mg daily (n = 165) at the study endpoint in comparison with the placebo-treated subjects (p = 0.001). Similar findings were noted for study completers at week 12. Decrement in comparison with the placebo treatment were noted for 6 of the 7 BEHAVE-AD symptomatic categories at both effective dosages at the study endpoint. Only one BEHAVE-AD symptomatic category (Anxieties and Phobias) showed little difference between the placebo and the 1- or 2-mg daily treatment groups. Hence, this study adds further support for the syndromic nature of the response to pharmacological treatment with the BEHAVE-AD symptomatic categories. This study also adds support for differential responsiveness of BEHAVE-AD symptomatic categories to medications with different modes of action. For example, whereas the BEHAVE-AD Anxieties and Phobias symptomatic category showed a significant response to the anxiolytic buspirone treatment in the study of Levy et al. [24], using a repeated-measures ANOVA (p < 0.05), no effect on the BEHAVE-AD Anxieties and Phobias symptomatic category was seen with the risperidone treatment. Also, the antipsychotic medication risperidone produced significant improvements in comparison with the placebo treatment on all individual items of the Paranoid and Delusional Ideation symptomatic category of the BEHAVE-AD [25, p. 109]. Finally, the BEHAVE-AD effects seen in the study of Katz et al. [25] appear to have been independent of cognitive effects or changes.
Other risperidone studies have also documented the efficacy of risperidone treatment in alleviating BPSD using the BEHAVE-AD. For example, De Deyn et al. [28] published a pooled data analysis of three 12-week, multicenter, randomized, double-blind, placebo-controlled, parallel group trials conducted in the United States, Australia, and Europe, respectively. The US trial was the study of Katz et al. [25] described above. The Australian trial was from Brodaty et al. [27] and the primarily European trial was published by De Deyn et al. [26].

The three trials examined were all of 12-week duration, and all were preceded by a 1-week washout period. Each of the trials was a multicenter, randomized, double-blind, placebo-controlled, parallel group phase III study, in which all other psychotropic medications were discontinued prior to the baseline visit. The total number of patients in the pooled efficacy analysis was 1,150. Of these, 722 received risperidone and 428 received placebo. The mean dose of risperidone was 1.0 mg/day. Analyses of efficacy were conducted at weeks 4, 8, and 12 and at the study endpoint.

At all evaluation points, i.e., at weeks 4, 8, and 12 and at the endpoint, scores on the BEHAVE-AD total scale were significantly (p < 0.001) more improved with risperidone than with placebo. Significant effects of risperidone treatment in comparison with the placebo treatment at the study endpoint were seen on the following BEHAVE-AD symptomatic categories: Paranoid and Delusional Ideation (p = 0.002); Activity Disturbances (p = 0.009); Aggressiveness (p < 0.001), and Anxieties and Phobias (p = 0.030). Nonsignificantly greater changes with risperidone treatment in comparison with the placebo treatment at the study endpoint were seen for the BEHAVE-AD symptomatic categories Hallucinations and Diurnal Rhythm Disturbances. No difference was seen in the BEHAVE-AD symptomatic category Affective Disturbances between the risperidone and the placebo treatment groups at the study endpoint. Hence, a clear conclusion from this pooled analysis study and the study of Katz et al. [25] is that Anxieties and Phobias, as measured by the BEHAVE-AD, are potentially responsive to antipsychotic medication treatment. Also, all BEHAVE-AD symptomatic categories are potentially more responsive to antipsychotic treatment than to placebo intervention. Hence, in terms of pharmacological treatment, the 7 BEHAVE-AD symptomatic categories constitute a pathologic syndrome.

### Utility

1. The BEHAVE-AD has been used as an outcome measure (generally the primary outcome measure) in pivotal trials associated with the approval of risperidone for the treatment of BPSD [e.g., 25–28] in more than 30 nations throughout the world including Australia, Finland, Germany, Ireland, The Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and Canada. Risperidone is today the most prescribed antipsychotic medication for the treatment of these symptoms in the United States and many other nations.

2. The BEHAVE-AD has been used as part of the NYU-CIBIC-Plus assessment [13, 14]. In this context, the BEHAVE-AD has been used as part of a primary outcome measure in the pivotal trials associated with the initial worldwide approvals of rivastigmine (Exelon [15, 16]) and memantine (Namenda, Ebixa, Axura [17]) for the treatment of AD. These medications are presently 2 of the 3 most widely prescribed medications for the treatment of AD in the United States and worldwide.

*Although the progressive worsening of cognitive and functional symptomatology is fundamental to the understanding and definitions of AD, one reference which is of relevance in the detailed description of the course of these symptoms is Reisberg et al. [9].

* Copyright © 2013 Barry Reisberg, MD. All rights reserved.
-2.1 vs. -1.3 for BEHAVE-AD psychotic symptom scores (p = 0.003). On the basis of these studies, risperidone was approved for the treatment of BPSD (or similarly designated entities) in more than 30 nations including Australia, Canada, Finland, Germany, Ireland, The Netherlands, Norway, Portugal, Spain, Sweden, and Switzerland.

Observations of Problems with Antipsychotic Medications in Dementia and Associated Prescribing Practices

An analysis of the results of the risperidone studies described above and an additional study by the manufacturer found an increased rate of cerebrovascular adverse events including strokes and transient ischemic attacks in the risperidone-treated subjects in comparison with placebo-treated patients. Cerebrovascular adverse events were twice as common with risperidone treatment as with placebo treatment (4 vs. 2%) [87, 88]. Subsequently, in the pooled analysis of the 3 studies described above, it was found that 1.6% of patients who received risperidone and 0.7% of placebo-treated patients had a cerebrovascular adverse event [28]. An analysis of 6 studies found a risk of cerebrovascular events in patients receiving risperidone of 3.3 versus 1.2% in placebo-randomized subjects [28].

At approximately the same time as the publication of the pooled analysis of the risperidone trials described above, a meta-analysis of randomized placebo-controlled trials with atypical antipsychotic medications found an increased mortality in patients randomized to the medication treatments (3.5 vs. 2.3%, with an odds ratio of 1.54) [89]. The overall relative risk of death was 1.65 for patients who received medications versus patients who received placebo. Although a sensitivity analysis did not show evidence of a differential risk for individual medications, the highest relative risk of death was associated with olanzapine treatment (RR = 2.31) and the lowest was associated with risperidone treatment (RR = 1.35). These findings were consistent with an FDA public health advisory issued at approximately the same time [90]. Shortly afterward, in the same year, 2005, it was found that conventional antipsychotic medications were associated with a significantly higher risk of mortality than were atypical antipsychotic medications [91]. This increased risk of death for conventional antipsychotic medications in comparison with atypical antipsychotic medications was observed at all intervals studied (up to 180 days) and in both dementia and nondementia subgroups as well as in both nursing home residents and non-nursing home residents. The greatest increases in risk occurred with higher dosages of antipsychotic medications [91].

In the year prior to the publication of these very dramatic and very disturbing findings with respect to the increased mortality with the use of antipsychotic medications in dementia patients, recommendations were published regarding the treatment of BPSD in dementia in general and AD in particular. The recommendations were published in the chapter on AD of the American Association for Geriatric Psychiatry’s (AAGP) Comprehensive Textbook of Geriatric Psychiatry – Third Edition [92]. These recommendations were summarized as follows:

’In terms of pharmacological treatment of BPSD symptomatology, the clinical adage “start low, go slow” applies.’ [92, p. 491]

These recommendations were translated very concretely in terms of treatment. With respect to antipsychotic medications for BPSD in patients with AD, it was noted:

’Although clinical circumstances dictate the schedule of dosage titration, an optimal clinical response is not achieved for many weeks on any particular dosage of medication. Therefore, ideally the clinician should endeavor to leave a patient on a particular dosage of medication for many weeks before further dosage adjustments.’ [92, p. 491, emphasis added]
A caveat was also noted with respect to medication prescribing for BPSD:

'The exigencies of particular situations, of course, will frequently not permit this time luxury in dose adjustments, and clinicians will frequently need to make rapid dose adjustments. However, the clinician should also be prepared to adjust medication dosage downward as well as upward in response to particular patient needs and emergence of side effects.' [92, p. 491, emphasis added]

The above general recommendations for BPSD treatment with antipsychotic medications were specified very concretely with the example of risperidone treatment (table 2):

1. 'For risperidone treatment, this rule [start low, go slow] translates into an optimal starting dose of 0.25 mg daily.' [92, p. 491]
2. 'Extrapyramidal side effects may not peak until a patient has been on a particular dosage of medication for as long as 6 months' [92, p. 491; see also 93]
3. 'The emergence of extrapyramidal symptoms including rigidity and akathisia, is a continuing concern with risperidone treatment of BPSD.' [92, p. 491]
4. 'After some months of treatment, a steady-state dosage of 0.25 to 1 mg of risperidone daily is frequently effective in controlling BPSD symptoms.' [92, p. 491]

This review is published nearly a decade after the AAGP textbook recommendations regarding the treatment of BPSD with risperidone were issued. At the present time, we would somewhat extend the above recommendations with respect to the treatment of BPSD in AD patients with risperidone. At the present time, the recommendation would note that a starting dose of 0.125 mg of risperidone (one half of the lowest dose manufactured in tablet form, obtained by using a pill cutter to divide the 0.25-mg tablets) is frequently both useful and sufficient in optimally controlling BPSD. In addition, most AD patients can be controlled with risperidone doses in the range from 0.125 to 0.75 mg daily. Finally, the balance between the optimal dosage of medication and the dosage of medication that produces side effects is very
delicate in terms of the ideal BPSD treatment of AD patients with risperidone. Therefore, it is presently recommended that the brand medication should be used if possible. This latter recommendation takes into consideration the very broad permissible variability for generic medications. For example, the US FDA permits up to a 45% difference in dosage for generic medications [94]. This means the generic dose may be 45% less or 45% more than the brand dose. Therefore, a single generic dose may be 90% more or 90% less than the prior dose. In the case of risperidone, with the delicate balance of efficacy on one side and the emergence of akathisia and other side effects on the other side, this swing and variability in dosing may not be clinically justifiable or desirable.

Other important recommendations with respect to the treatment of BPSD were published in the 2004 AAGP chapter on AD. One was that ‘pharmacologic treatment should always be given in conjunction with non-pharmacologic treatment approaches’. Another was that ‘treatment of BPSD symptomatology in … patients with AD with traditional, so-called typical neuroleptic agents such as haloperidol can cause severe problems’. These problems with haloperidol treatment were described in greater detail. Other more general problems with antipsychotic medication prescribing were also noted. One important example is akathisia (the inability to remain still): it was noted that, ’It is often impossible to distinguish akathisia as a medication side effect from the activity disturbances resulting from AD. Under these circumstances, the clinician frequently increases the medication dosage and thereby worsens the condition of the patient’ [92]. Also, very importantly, it was noted that falling is a major problem: ‘The more severely the patient is affected by AD, the greater the gait and balance disturbance of the patient’ [92, 95–97].

Therefore, it should be noted that the recommendations forwarded by Reisberg and Saeed [92] in 2004 were very different from the nature of the trials that were conducted for the approval of BPSD medications. In the case of risperidone, direct comparisons between the AAGP textbook recommendations and the pivotal trial procedures are possible (table 2).

The same patterns of the usage of very high dosages of antipsychotic medications and very rapid titrations exclusively upwards, in some cases to the maximally tolerated medication doses, applied to studies of treatment with antipsychotic medications other than risperidone. As already reviewed, the consequence, in part, may have been the finding in 2005 of the increases in death with antipsychotic medication usage in the collective pharmacological trials [89, 90]. Later that same year, Wang et al. [91], in their review of both atypical antipsychotic medications and so-called ‘typical’ traditional antipsychotic medications, found that the greatest risk of mortality occurred with higher dosages of antipsychotic medications. Nevertheless, antipsychotic medications continue to be prescribed with dosages and titration schedules that are much higher and much faster than those recommended in the 2004 AAGP textbook. The differences between prescribing practices and titration schedules commonly employed are even greater with respect to the updated recommendations which have been forwarded in this review.

For example, 2 years after the publication of the textbook recommendations [92] and 1 year after the publication of the increased mortality with antipsychotic medications in elderly patients with dementia [89] and elderly in general [91], Schneider et al. [98] published the CATIE-AD (Clinical Antipsychotic Trials of Intervention Effectiveness – Alzheimer’s Disease) study. This large 42-site double-blind placebo-controlled trial randomized predominantly community-residing patients with AD and psychosis, aggression, or agitation to receive Zyprexa (olanzapine), Seroquel (quetiapine), Risperdal (risperidone), or placebo. The dosages were adjusted upwards or downwards on an ‘as-needed’ basis. Subjects were followed for up to 36 weeks, and a main outcome analysis was done at 12 weeks. The subjects were given much higher medication doses, which were adjusted much more quickly than in the 2004 AAGP textbook recommendations [92]. For example, in the case of
Risperdal (risperidone), the 'low-dose' capsule was 0.5 mg. This 'low dose' was two times the starting dose recommended in the 2004 textbook [92]. The actual mean starting dose in the study was 0.7 mg of Risperdal, which is nearly three times the recommended starting dose in the 2004 textbook [92] and nearly six times greater than the starting dose being recommended herein. The mean last dose of Risperdal in the CATIE-AD study was 1.0 mg, which is identical to the mean dose given in the pooled analysis of 3 of the prior Risperdal trials published by De Deyn et al. [28] in 2005.

Schneider et al. [98] found in their CATIE study that many subjects in all treatment and placebo groups discontinued treatment 'for any reason'. Significantly more subjects were discontinued from the placebo treatment because of perceived lack of efficacy than subjects from the blinded Risperdal treatment group. On the other side of the 'equation', significantly more subjects discontinued the Risperdal treatment because of perceived side effects than subjects in the placebo group. At 12 weeks, there was little difference between the subjects in the Risperdal and the placebo groups on a main outcome measure, the Clinical Global Impression of Change (CGIC). However, only small percentages of either Risperdal-treated (28.6%) or placebo-randomized subjects (20.9%) reached this 12-week study evaluation point. Improvement was observed in 29% of the Risperdal-treated subjects and in 21% of the placebo-randomized subjects. The conclusion of this study was that 'adverse effects offset advantages... of atypical antipsychotic drugs for the treatment of psychosis, aggression, or agitation in patients with Alzheimer's disease' [98]. Herein, we suggest that much better outcomes can be obtained with antipsychotic medications if the 'start low, go slow' and other recommendations in the AAGP textbook published in 2004 [92] and the updated recommendations provided in this publication are followed.

The BEHAVE-AD and the Behavioral Pathological Syndrome of AD

The BEHAVE-AD was published in 1987 [1]. This scale was unique in its identification and characterization of the nature of the behavioral disturbance pathology in AD. The history of the development of the BEHAVE-AD has, to some extent, been discussed earlier in this perspective and review. Additionally, the history of the development of the BEHAVE-AD and some of its unique elements has been discussed in the literature [99]. One unique element of the BEHAVE-AD was that the scale separates cognitive and functional symptoms and changes that occur in AD from behavioral disturbance symptoms. Cognitive and functional symptoms in the context of AD all occur universally with the temporal progression of AD, and the losses in these domains tend to worsen progressively as the disease advances [19, 99]. In contrast, the behavioral disturbances identified with the BEHAVE-AD, although common and disturbing to caregivers or other informants, (1) do not occur universally in all AD patients and (2) all peak in manifestation and severity at some stage prior to the final GDS stage 7 of AD [22].

Additionally, the BEHAVE-AD identifies symptoms that are potentially remediable in the course of AD, both in terms of potential responsivity to neuroleptic (antipsychotic) medications and other psychotropic behavioral medication approaches, and also in the context of nonpharmacological management of AD patients. Finally, it should be noted that many of the symptoms identified with the BEHAVE-AD and its predecessor scale, the Symptoms of Psychosis in Alzheimer's Disease (SPAD) rating scale, had not been used in rating instruments prior to the advent of these assessment measures. As already reviewed in part, the reliability both in person [24, 82–84] and using a telephone interview format [84], the construct validity in terms of the separation of cognitive and functional symptoms of AD from the behavioral disturbance symptomatology [9, 19, 22], and the criterion validity in terms of responsivity to both nonpharmacological [e.g., 24; reviewed in 2] and pharmacological treatment with both
antipsychotic [25–28] and anxiolytic medication [24] have all been demonstrated for the BEHAVE-AD.

Furthermore, it has been demonstrated that the 7 behavioral categories assessed with the BEHAVE-AD constitute a distinct clinical syndrome occurring in the context of AD [100]. A clinical syndrome has been defined as ‘the aggregate of signs and symptoms associated with any morbid process’ [101] or as ‘a set of symptoms which occur together’ [102]. The word syndrome derives from Greek roots, with ‘syn’ meaning ‘together’ and ‘dromos’ meaning ‘a running’. Hence, literally, a clinical syndrome is a set of clinical symptoms that ‘run together’.

It has already been noted in this perspective and review that all of the 25 BPSD identified with the BEHAVE-AD differ from the cognitive and functional symptoms of AD in that all these BEHAVE-AD symptoms peak in occurrence at a stage prior to the final stage (i.e., GDS and FAST stage 7) of AD [22]. In addition, the cognitive and functional symptoms of AD differ from the BEHAVE-AD symptoms in that they continue to worsen throughout the course of AD including the final GDS and FAST stage 7 of AD [9, 19, 23]. Hence, from these perspectives, the BEHAVE-AD symptoms constitute a distinct clinical syndrome, distinguishable from the cognitive and functional symptoms of AD.

However, pharmacological treatment trials have provided much more specific and rigorous proof of the distinct syndromic nature of the BPSD syndrome identified with the 7 BEHAVE-AD symptomatic categories. Specifically, the 7 BEHAVE-AD symptomatic categories, in terms of the associated BEHAVE-AD severity scores, have been shown to respond in concert (running together) to the nonpharmacological supportive effects associated with the entry into a pharmacological trial and the resulting nonspecific additional support provided to the placebo-treated group. Furthermore, effective pharmacological treatment has been shown to affect all 7 BEHAVE-AD symptomatic categories to a greater extent than placebo treatment, with associated significant effects on some BEHAVE-AD categories as well as total BEHAVE-AD scores. Finally, these significant effects of pharmacological interventions on the BEHAVE-AD have been shown to be independent of medication effects on cognition.

For example, Levy et al. [24] conducted a single-blind dose escalation study of the efficacy of the non-benzodiazepine anxiolytic drug buspirone in the treatment of disruptive behavior in subjects with possible AD [32]. Data of the 12 patients who completed the 9 weeks of the dose escalation study were analyzed. The dose escalation protocol was as follows: week 1, placebo; weeks 2 and 3, buspirone 15 mg daily; weeks 4 and 5, buspirone 30 mg daily; weeks 6 and 7, buspirone 45 mg daily, and weeks 8 and 9, buspirone 60 mg daily. The best medication dosage, i.e., 30 mg per day, was compared with the placebo treatment for the primary outcome measure, the BEHAVE-AD.

In this study, the placebo treatment was associated with a reduction in the BEHAVE-AD severity scores for each of the 7 BEHAVE-AD symptomatic categories [24]. These results demonstrate the sensitivity of the BEHAVE-AD to nonpharmacological therapeutic approaches and the syndromic nature of this responsiveness. Specifically, all 7 BEHAVE-AD symptomatic categories responded with a decrease in scores to the nonpharmacological placebo intervention (i.e., the 7 BEHAVE-AD symptomatic categories ‘ran together’ in a syndromic manner). The BEHAVE-AD total severity scores declined by 25% and the BEHAVE-AD global scores decreased by 14% in response to the nonspecific placebo and enrollment interventions [24].

In this study of Levy et al. [24], the syndromic nature of the 7 BEHAVE-AD symptomatic categories was also observed with respect to the pharmacological treatment intervention. Data for the present syndromic analysis are reported for the most effective pharmacological treatment dose, i.e., 30 mg daily of buspirone. The total BEHAVE-AD severity scores decreased significantly on the 30-mg daily dosage of buspirone treatment condition in comparison with the placebo treatment condition (Scheffé’s pairwise post hoc comparison). The total BEHAVE-AD scores declined by 29% with 30 mg of buspirone daily in comparison with the placebo...
treatment. Also, a post hoc analysis showed that the BEHAVE-AD global score was significantly better on the 30-mg daily dosage of buspirone in comparison with placebo. The BEHAVE-AD global scores declined by 48% with 30 mg of buspirone daily in comparison with the placebo treatment. Six of the 7 BEHAVE-AD categories individually showed lower total severity scores with the 30-mg buspirone treatment in comparison with the placebo treatment. For the remaining symptomatic category Affective Disturbance, no change was noted between the placebo treatment and the 30-mg buspirone treatment. Therefore, the additional effect of the pharmacological treatment with 30 mg of buspirone was observed in a syndromic manner for 6 of the 7 BEHAVE-AD categories. For 1 of the 7 BEHAVE-AD symptomatic categories, the incremental effects of the 30-mg buspirone pharmacological treatment was statistically significant in comparison with the placebo treatment. Specifically, a significant incremental effect of the 30-mg buspirone treatment in comparison with the placebo treatment was observed for the BEHAVE-AD category D (Aggressiveness).

These syndromic effects seen in the study of Levy et al. [24] appear to have occurred independently of cognitive changes. The baseline MMSE score was 11.8 and the MMSE score at the end of the 9-week treatment trial was 11.6, showing a small, statistically nonsignificant difference.

An analysis of the syndromic nature of each of the 7 categories of symptoms assessed with the BEHAVE-AD, their sensitivity to the pharmacological intervention, and the independence from cognitive change was also conducted on the large multicenter study of Katz et al. [25]. This study was used as a pivotal trial associated with the approval of risperidone for the treatment of BPSD in many nations throughout the world. Forty sites in the United States participated in this study. All subjects were residents in chronic care settings. Subjects with AD and/or vascular dementia were eligible for participation in the study. Participants received placebo or a fixed daily dose of risperidone of 0.5, 1, or 2 mg for 12 weeks. The doses for subjects who were randomized to receive 1 or 2 mg/day were increased from 0.5 to 1 or 2 mg/day in increments of 0.5 mg/day every 2 days during the first week of the double-blind phase of the study until the participants reached their assigned dosage level. Katz et al. [25] conducted an endpoint analysis of all study subjects who received at least one dose of the study medication and had at least one post-baseline assessment. The study comprised 625 randomized subjects, and the BEHAVE-AD was the primary outcome measure. An analysis of the 7 individual BEHAVE-AD symptomatic categories was conducted, which was more detailed than the analysis published by Katz et al. [25, 100, 103].

All treatment conditions in the study of Katz et al. [25] resulted in significant reductions in each of the 7 BEHAVE-AD symptomatic categories compared with baseline as well as in reductions in the BEHAVE-AD total scores. The placebo treatment resulted in decrements in the BEHAVE-AD categorical symptoms from baseline to endpoint of 18% (for Activity Disturbances) to 45% (for Anxieties and Phobias). The total BEHAVE-AD scores decreased by approximately 25% in response to the placebo treatment in the study of Katz et al. [25]. It is interesting to note that this response to the placebo intervention was virtually identical to that observed by Levy et al. [24]. Hence, it is clear that, in terms of each of the 7 BEHAVE-AD symptomatic categories, the BEHAVE-AD was shown to have a syndromic response in the study of Katz et al. [25] to the nonspecific effects associated with the entry into a clinical trial and the associated increase in social and medical attention.

In this study of Katz et al. [25], differential sensitivity (efficacy) was also observed between the placebo treatment group and 2 of the medication treatment groups. Specifically, subjects who received 1 or 2 mg risperidone daily improved more than placebo-treated subjects in the BEHAVE-AD total scores and in the psychosis BEHAVE-AD categories (i.e., the combination of the BEHAVE-AD category A. Paranoid and Delusional Ideation and the BEHAVE-AD category B. Hallucinations) and in the BEHAVE-AD category D. Aggressiveness.
For the total BEHAVE-AD scores, the p values for these differences between the 2 mg daily risperidone treatment group and the placebo treatment group at the final study week (week 12) and at the endpoint were p < 0.001 and p = 0.001, respectively. For the psychosis subscale, the p values for the difference between the 2 mg daily risperidone treatment group and the placebo group at week 12 and at the endpoint were p = 0.002 and p = 0.01, respectively. For the BEHAVE-AD category Aggressiveness, the p values at week 12 and at the endpoint were both p < 0.001. Additionally, significant improvements in comparison with the placebo treatment were found on all of the individual items of the Paranoic and Delusional Ideation category of the BEHAVE-AD. These results support the criterion validity of the BEHAVE-AD with respect to antipsychotic medication treatment.

In addition, a detailed analysis of the response to the efficacious doses of risperidone, i.e., 1.0 and 2.0 mg daily, supports the syndromic nature of the response to risperidone on the BEHAVE-AD in comparison with placebo. Specifically, larger decrements in the BEHAVE-AD categorical scores were observed for 6 of the BEHAVE-AD categories in the two statistically significant risperidone medication dosage groups than in the placebo treatment group at the study endpoint [25, 100, 103]. The BEHAVE-AD categories A (Paranoid and Delusional Ideation), B (Hallucinations), C (Activity Disturbances), D (Aggressiveness), E (Diurnal Rhythm Disturbances), and F (Affective Disturbances) all decreased to a greater extent at the endpoint in the subjects treated with 1 or 2 mg/day risperidone in comparison with the placebo treatment subjects. The remaining BEHAVE-AD symptomatic category G (Anxieties and Phobias) showed very little difference (≤5% differences) between the placebo-treated subjects and the subjects treated with 1 or 2 mg/day risperidone. Hence, these results from the study of Katz et al. [25] support the syndromic nature of the BEHAVE-AD categorical response to antipsychotic medication intervention over and above the placebo treatment condition.

In the study of Katz et al. [25], the syndromic responses of the BEHAVE-AD, both in terms of total scores and in terms of individual BEHAVE-AD categorical responses to both nonspecific placebo treatment and also to specific treatment with antipsychotic medication, were independent of effects on cognition. The subjects tested by Katz et al. [25] were considerably more impaired than those studied by Levy et al. [24]. The mean baseline MMSE score of the subjects tested by Katz et al. [25] randomized to the double-blind treatment conditions (n = 625) was 6.6 ± 6.3. A further indication of the severity of these subjects is that 96% of the Katz et al. [25] subjects had a FAST stage [20, 78] ≥6a, indicating that they needed assistance with putting on their clothing properly and/or had more severe functional deficits.

Furthermore, 47% of the Katz et al. [25] subjects had a FAST stage ≥7a, indicating that at minimum, these subjects could no longer dress or bathe or toilet independently and were doubly incontinent, and that their speech had deteriorated to the extent that intelligible speech was limited to about a half dozen intelligible words in an average day or in the course of a clinical interview. Many of these subjects with a FAST stage ≥7a may have had even more severe deficits with speaking, ambulation, etc., in accordance with the FAST procedure. The MMSE scores decreased from baseline to follow-up in both of the treatment conditions associated with the greatest improvements on the BEHAVE-AD total and categorical scores, i.e., the 1- and 2-mg daily risperidone treatment groups. Specifically, the MMSE scores decreased in the subjects taking 1 mg risperidone per day by −0.20 and −0.74 at week 12 and at the endpoint, respectively. For the subjects taking 2 mg risperidone per day, the MMSE scores decreased by −0.14 and −0.64 at week 12 and at the endpoint, respectively. Hence, whereas cognitive scores worsened, the BEHAVE-AD total scores and the 7 BEHAVE-AD categorical scores all improved significantly in comparison with baseline values. These results strongly support the independence from cognition of the pathological syndrome assessed with the BEHAVE-AD and also of the independence from cognition of the 7 individual symptomatic
categories of the BEHAVE-AD. Whereas cognition decreased compared to baseline slightly, the BEHAVE-AD total scores and each of the 7 BEHAVE-AD categorical scores significantly improved compared to baseline.

It is instructive to compare the BEHAVE-AD syndromic response to anxiolytic medication treatment in the study of Levy et al. [24] with the BEHAVE-AD syndromic response to antipsychotic medication treatment observed in the study of Katz et al. [25]. Only the anxiolytic treatment showed a significant effect on the BEHAVE-AD symptomatic category G (Anxieties and Phobias). Although this significant effect on the Anxieties and Phobias BEHAVE-AD subscale was not observed for the 30-mg buspirone dose, which was the focus of the earlier discussion, buspirone in doses of 45 and 60 mg daily resulted in significant improvements in this BEHAVE-AD symptomatic category in comparison with the placebo treatment. It is also notable that this Anxieties and Phobias BEHAVE-AD symptomatic category was the least responsive to pharmacological intervention with the antipsychotic medication risperidone in comparison with the nonspecific effects of the placebo intervention. Hence, the BEHAVE-AD symptomatic categories appear to have differential responsivity and sensitivity to different classes of psychotropic medications.

The pooled analysis of multicenter randomized placebo-controlled trials of the efficacy and safety of risperidone for the treatment of BPSD in patients with dementia provides an additional view of the pathological behavioral syndrome of AD and its responsiveness to antipsychotic treatment [28]. This pooled analysis examined the results of 3 previously published studies, including the US multicenter trial of Katz et al. [25] described above. The other studies included in this pooled analysis were an Australian study [27] and a predominantly European study [26]. All of these studies were carried out in nursing home facilities. In addition, all of the studies had a 1-week washout period, at which time psychotropic medications were discontinued, followed by randomization to risperidone treatment or placebo.

Subjects from the pooled studies were evaluated with the BEHAVE-AD and other assessment scales at post-washout weeks 4, 8, and 12. Also, an endpoint analysis was conducted using the last observation carried forward procedure. The pooled study efficacy analysis included 1,150 patients, of which 722 received risperidone treatment and 428 received placebo. The mean MMSE score for the entire study group in the pooled efficacy analysis, including patients who were subsequently excluded from the efficacy analysis because of lack of post-baseline efficacy assessments (n = 5), was 6.79 (standard error, 0.19).

In this pooled analysis study, the mean dose of risperidone given across all of the trials was 1.0 mg/day (standard error, 0.02). The subjects receiving the risperidone treatment showed significantly improved BEHAVE-AD total scores in comparison with the placebo-treated subjects at all assessment points during the study and at the study endpoint (p < 0.001). Specifically, the BEHAVE-AD scores were significantly lower for the risperidone-treated subjects than for the placebo-treated subjects at the 4-, 8-, and 12-week treatment points as well as at the endpoint analysis (all p values <0.001). Therefore, the BEHAVE-AD indicated the efficacy of the risperidone antipsychotic treatment in treating the behavioral pathological syndrome of AD and related dementias. Furthermore, these effects were independent of changes in cognition. The publication noted that ‘no clinically or statistically significant differences were observed between risperidone (6.3 ± 0.3) and placebo (6.7 ± 0.4) with respect to their effect on MMSE scores at endpoint’ [28].

This pooled analysis also summarized the difference between the risperidone and placebo treatment groups on each of the 7 BEHAVE-AD subscales at the endpoint. Significant differences in favor of the risperidone treatment were seen for the BEHAVE-AD symptomatic categories A (Paranoid and Delusional Ideation; p = 0.002), C (Activity Disturbances; p = 0.009), D (Aggressiveness; p < 0.001), and G (Anxieties and Phobias; p = 0.030). Of the remaining 3 BEHAVE-AD symptomatic categories, 2 showed changes that were larger in the risperidone
treatment group than in the placebo group. These were categories B (Hallucinations) and E (Diurnal Rhythm Disturbances). One BEHAVE-AD symptomatic category, i.e., category F (Affective Disturbances), did not show any difference in BEHAVE-AD change scores between the placebo and the risperidone treatment groups.

Taken together, the study of Katz et al. [25] and the pooled analysis of De Deyn et al. [28] lend further support to the BEHAVE-AD categorical syndromic response with antipsychotic medication treatment in comparison to placebo treatment. The one BEHAVE-AD category in the study of Katz et al. [25] which did not show a favorable response to risperidone treatment in comparison with placebo treatment, i.e., the BEHAVE-AD category G (Anxieties and Phobias), showed a significant response (p = 0.03) in the pooled analysis of De Deyn et al. [28]. Also, it should be noted that all of the antipsychotic risperidone studies included relatively severe subjects, and some of the less responsive or inconsistently responsive BEHAVE-AD symptomatic categories have been shown to peak at earlier stages in the progression of AD dementia. For example, the BEHAVE-AD category E (Diurnal Rhythm Disturbance) was found to peak in occurrence at GDS stage 5 of AD dementia, corresponding to moderate AD [22]. Another example is the BEHAVE-AD category G (Anxieties and Phobias). This symptomatic category was also found to peak in occurrence at GDS stage 5 of AD dementia. As noted in the preceding discussion, this symptomatic category showed a significant therapeutic response to the risperidone treatment in the pooled analysis [28], but not in the study of Katz et al. [25]. A partial explanation for this disparity might be the somewhat less cognitively impaired composition of the subjects in the pooled analysis, with a mean MMSE score of 6.8, in comparison with the subjects in the study of Katz et al. [25] who had a mean MMSE score of 6.6. Apart from the differences in the mechanism of action of the antipsychotic medication risperidone and the anxiolytic medication buspirone, this severity difference may also explain part of the significant effect observed on the BEHAVE-AD Anxieties and Phobias symptomatic category with the buspirone treatment. The buspirone subjects had a mean MMSE score of 11.8 at baseline, which was considerably higher than that in the study of Katz et al. [25] or in the pooled analysis of the risperidone studies [28].

**Use of the BEHAVE-AD in Worldwide Studies of Medications Approved for the Treatment of Cognition in AD Patients**

The BEHAVE-AD has shown widespread utility in worldwide studies, both in the approval of BPSD medications [104], as discussed above, and also in the worldwide approvals of cognitively acting medications for the treatment of AD. With respect to the approvals of cognitively acting medications, the BEHAVE-AD has been employed as a component of the NYU-CIBIC-Plus scale [13, 14] (table 3). In this context, the BEHAVE-AD has been used as part of a primary outcome measure in the pivotal trials associated with the initial worldwide approvals of both cholinesterase inhibitor medication and NMDA (N-methyl-D-aspartate) receptor antagonist medication for the treatment of AD. Specifically, the NYU-CIBIC-Plus assessment, which is comprised in part of the BEHAVE-AD, was used in the initial pivotal trials for the approval of both rivastigmine (Exelon) [15, 16] and memantine (Namenda, Ebixa, Axura) [17] for the treatment of AD.

The approvals of these medications in this context might indicate that informant-reported and direct subject observation-based behavioral disturbances either decreased or remained the same, or, alternatively, did not outweigh the positive cognitive and/or functional effects of these medications in terms of their overall positive effects on AD patients. However, these rivastigmine and memantine studies did not analyze the sources of the positive effects on the NYU-CIBIC-Plus assessment beyond the final CIBIC-Plus global score. Hence, no information
### Table 3. The Clinician’s Interview Based Impression of Change with Caregiver Input (CIBIC-Plus) and the NYU-CIBIC-Plus Assessment*

**A** Wording used in the memantine pivotal trial for the NYU-CIBIC-Plus final assessment [17]

<table>
<thead>
<tr>
<th>Is the subject?</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Markedly improved</td>
<td>1</td>
</tr>
<tr>
<td>Moderately improved</td>
<td>2</td>
</tr>
<tr>
<td>Minimally improved</td>
<td>3</td>
</tr>
<tr>
<td>Unchanged</td>
<td>4</td>
</tr>
<tr>
<td>Minimally worse</td>
<td>5</td>
</tr>
<tr>
<td>Moderately worse</td>
<td>6</td>
</tr>
<tr>
<td>Markedly worse</td>
<td>7</td>
</tr>
</tbody>
</table>

**B** Procedures for the assessment of the NYU-CIBIC-Plus

1. Must be obtained by a clinician independent of other study data
2. The CIBIC rater must be blind to other study data
3. Uses semi-structured ascertainment methodologies to arrive at final severity assignments
4. Contains both direct patient assessment and informant assessment procedures
5. Systematically surveys cognition, function, and behavior

**C** The NYU-CIBIC-Plus is the most detailed and comprehensive of the CIBIC-Plus procedures. These NYU-CIBIC-Plus procedures include the following:

1. All CIBIC-Plus procedures described above in A and B.
2. Part 1: A subject (patient) assessment using validated semi-structured interviews assessing cognition and behavior. The subject-based assessment includes:
   (a) A cognitive component. This systematically surveys the following areas:
       I. Concentration
       II. Recent memory
       III. Remote memory
       IV. Orientation
       Based upon previously validated procedures described in Reisberg and Ferris [53].
   (b) A behavioral component. This systematically surveys the following areas:
       I. Paranoid and Delusional Ideation
       II. Hallucinations
       III. Activity Disturbances
       IV. Agitation
       V. Affective Disturbances
       VI. Anxiety and Phobias
       Based upon previously validated procedures described in Auer et al. [12].
3. Part 2: An informant (caregiver) assessment using a validated semi-structured interview assessing functioning and behavior. The informant-based assessment includes:
   (a) A functional component. This systematically assesses the level and occurrence of the functional disability using validated procedures described in Sclan and Reisberg [20].
   (b) A behavioral component. This systematically assesses behavioral disturbances in the following areas:
       I. Paranoid and Delusional Ideation
       II. Hallucinations
       III. Activity Disturbances
       IV. Aggressiveness
       V. Diurnal Rhythm Disturbances
       VI. Affective Disturbances
       VII. Anxieties and Phobias
       Based upon procedures validated in Reisberg et al. [1] and Sclan et al. [83].
4. Procedures for combining this comprehensive information towards ascertainment of a CIBIC-Plus severity rating [13].

* Copyright © 2007 Barry Reisberg, MD. All rights reserved.
regarding the behavioral effects of rivastigmine or memantine beyond speculations such as the ones above can be derived from the NYU-CIBIC-Plus scores in these approvals.

However, a recent study of the efficacy of a Comprehensive, Individualized Person-Centered Management (CI-PCM) program in addition to memantine treatment versus memantine treatment alone illustrates the role of the BEHAVE-AD as an element of the NYU-CIBIC-Plus assessment [105, 106]. The essential elements of the NYU-CIBIC-Plus assessment have been described previously [14] and are shown in table 3. The use of global scales in AD assessment and the history of these scales in the context of the development of the NYU-CIBIC-Plus assessment have been extensively reviewed [14, 107] and are described very briefly below.

The CIBIC scales were developed out of a concern by the US FDA that medications for the treatment of AD not be approved based upon small effects that are not clinically significant. Therefore, it was concluded that AD studies should contain an assessment of change which a clinician can observe in studies of AD medications for regulatory body approval. However, since AD studies involve assessments which are typically separated by an interval of weeks or longer, and since even the same clinician may assess subjects differently at different times, there was a need to translate this FDA mandate into relevant measures' [14]. 'There are 3 categories of clinician interview-based global scales: (1) anchor-based, (2) unanchored, and (3) guideline-based' scales [107]. In terms of guideline-based scales for AD studies, an early instrument which was used was called the CIBI (Clinician Interview Based Impression). The CIBI assessed the patient only, with selected areas of deficit recorded at baseline [108]. In November 1991, the US FDA sent a letter to sponsors of AD pharmaceutical trials proposing a new measure called the CIBIC (Clinician's Interview Based Impression of Change). Although the 1991 letter was never published, the essential elements of the new measure were subsequently described in the literature [107, 109]. ‘The underlying premise [of the CIBIC] was that if an experienced clinician can perceive clinical change in an interview, then such change is likely to be clinically meaningful’ [107]. A major aspect of the CIBIC assessments was that the evaluating clinician assesses the patient without reference to any additional information which might bias the assessment. Shortly after the designation of the CIBIC, a second measure was described and developed: the CIBIC-Plus. The CIBIC-Plus assessment measure incorporates information from an informant as well as information from the patient. If more than a single informant is used to obtain information regarding the patient, then the assessment is termed a Clinical Global Consensus Assessment. Examples of scales which have been developed and used in large-scale pharmacological trials based on the FDA guidelines for the CIBIC and CIBIC-Plus scales are the Alzheimer's Disease Cooperative Studies Unit's Clinician's Global Impression of Change (ADCS-CGIC) and the NYU-CIBIC-Plus.

The ADCS-CGIC was developed as part of the ADCS Instrument Development Project under the direction of Steven Ferris [110]. An initial publication on this assessment was from Olin et al. [111] in 1996 and described some of the procedures for the construction of the instrument. Subsequently, Schneider et al. [112] described the validity and reliability of the ADCS-CGIC. In 2006, self-rated and study partner-rated versions of the ADCS-CGIC were published [113]. The ADCS-CGIC was used successfully in the pivotal trials that resulted in the worldwide approvals of donepezil for the treatment of AD [114, 115] and has been widely used subsequently.

The NYU-CIBIC-Plus is the most detailed of the CIBIC-Plus ratings. The elements of this assessment are described in table 3. Table 3 shows an element of all CIBIC-Plus assessments. This is that at the conclusion of the CIBIC-Plus interview, there is a determination of a final global assessment of change (table 3A). An overview of the CIBIC-Plus procedures used in obtaining the NYU-CIBIC-Plus is provided in table 3B. These procedures include: (1) That the interviews must be conducted by a clinician who is not involved in obtaining any of the other
data in the study. (2) That the ratings obtained by the CIBIC-Plus interviews must be performed without reference to any other data obtained in the study. In other words, the CIBIC-Plus rater must be blind to any other data in the study. (3) The CIBIC-Plus uses semi-structured procedures to arrive at the final decision with respect to change on the CIBIC-Plus. (4) As a CIBIC-Plus assessment, rather than simply a CIBIC assessment procedure, the NYU-CIBIC-Plus contains both a direct patient (subject) assessment and an assessment from a knowledgeable informant, such as a spouse or a caregiver. Finally, the NYU-CIBIC-Plus systematically surveys three behavioral domains to assist the clinician in formulating final assessments of change from baseline on the NYU-CIBIC-Plus. These domains, surveyed by the NYU-CIBIC-Plus clinician, are cognition, functioning, and behavior.

The NYU-CIBIC-Plus assessment has two major parts. The first part is a patient (subject) interview-based assessment, and the second part is an informant (caregiver) interview-based assessment. Each part has two components that are outlined in Table 3C.

As mentioned above, part 1 of the NYU-CIBIC-Plus is a subject-(patient)-based assessment. The components of the subject-based assessment are: (a) a cognitive component and (b) a behavioral component. As is true of all elements of the NYU-CIBIC-Plus assessment, these subject-based assessments are comprised of semi-structured procedures that had previously been used, validated, and shown to be reliable in studies which had been conducted prior to the component's incorporation into the CIBIC-Plus rating instrument.

The NYU-CIBIC-Plus part 1, subject-based assessment cognitive component, incorporates procedures assessing the subject's abilities in 4 areas: (I) concentration, (II) recent memory, (III) remote memory, and (IV) orientation. The semi-structured procedures employed in these assessments are from the previously developed and validated BCRS axes I–IV [53] and have been described in detail in that context.

The behavioral component of the NYU-CIBIC-Plus part 1 is used to directly assess the subject in 6 different areas of potential behavioral disturbance. The behavioral disturbance areas assessed are: (I) Paranoid and Delusional Ideation, (II) Hallucinations, (III) Activity Disturbances, (IV) Agitation, (V) Affective Disturbances, and (VI) Anxieties and Phobias. The procedures used are derived from those employed in the BEHAVE-AD [1] and were subsequently modified and validated as a separate, potentially independent assessment: the E-BEHAVE-AD [12].

Part 2 of the NYU-CIBIC-Plus is an informant-based assessment. The components of the informant-based assessment are: (a) a functional component and (b) a behavioral component. These informant-based assessments are comprised of semi-structured procedures which had previously been used and validated and shown to be reliable in studies which had been conducted prior to the component's incorporation into the CIBIC-Plus rating instrument.

The functional component of the NYU-CIBIC-Plus part 2 is based upon the previously validated FAST scale [20] and describes functional deficits present in individuals with SCI, deficits present at an executive level, deficits in the instrumental activities of daily life, incipient basic activity of daily life deficits, progressive basic activities of daily life deficits, and progressive functional deficits that occur after individuals can no longer perform any basic activities of daily life. Procedures for scoring the levels of deficit and any excess (nonordinal) functional deficits for this functional component have been published [14].

The behavioral component of the NYU-CIBIC-Plus part 2 is based upon the BEHAVE-AD [1, 83]. This behavioral component systematically assesses informant-based behavioral disturbances in the following areas: (I) Paranoid and Delusional Ideation, (II) Hallucinations, (III) Activity Disturbances, (IV) Aggressiveness, (V) Diurnal Rhythm Disturbances, (VI) Affective Disturbances, and (VII) Anxieties and Phobias.

As noted earlier in this section, the recent studies of the benefits of a CI-PCM program in addition to memantine treatment, in comparison with memantine treatment alone, illustrate...
the potential utility of the detailed and comprehensive NYU-CIBIC-Plus assessments in distinguishing not only clinically observable changes in an AD trial, but also in sensitively assessing the sources of these observable changes.

In the study comparing the CI-PCM program plus memantine versus memantine treatment alone [105], significant improvements were observed in the NYU-CIBIC-Plus assessment in the subject group receiving the management program in comparison with the subject group which did not receive this program. More specifically, with only 10 subjects per group, significant positive effects were observed at all time points studied with the CI-PCM program (p < 0.01 at weeks 4 and 12, and p < 0.001 at week 28) [105]. Since in this CI-PCM program study, the same NYU-CIBIC-Plus measure was used as in the pivotal trial associated with the initial worldwide approval of memantine [17] and since the criteria for subject selection were very similar, it was possible to compare the magnitude of differences between these two studies which were conducted 10 years apart. In the memantine versus placebo pivotal trial, for the observed cases at the 28-week study endpoint, a 0.3-point difference in favor of the memantine treatment condition was observed on the NYU-CIBIC-Plus assessment, which was a statistically significant difference in this study (p = 0.03) [17]. For the study published in 2013 comparing the CI-PCM program plus memantine versus memantine treatment alone, the magnitude of the difference in the NYU-CIBIC-Plus score at the end of 28 weeks of treatment was 2.7 [106]. Hence, the CI-PCM program appears to have added a 900% benefit over and above the benefit previously demonstrated with memantine treatment alone [17, 106].

The above result is very dramatic and exciting. However, it immediately raises the question about the origin of the effect of the CI-PCM program. Is it the result of an effect of the management program on cognition? Is the dramatic benefit the result of the effect of the management program on functioning? Is the dramatic benefit the result of the effect of the management program on BPSD and associated disturbances? If the dramatic benefit is the result of the effect of the management program on BPSD-related behavioral disturbances, are they behavioral disturbances which are observable in the context of a subject interview, or are they behavioral disturbances which are reported by informants? Naturally, the positive effects may have been the result of a combination of two or more of the above factors. If so, what is that combination of dramatic positive benefits?

Since the NYU-CIBIC-Plus assessment is, in part, the product of 4 separate, individually validated assessments, each with detailed semi-structured interviews, [1, 12, 20, 53, 83], it is possible to address the above questions. An analysis indicated that the CI-PCM program effects were not the result of changes in cognition. No significant differences were observed between the CI-PCM subject group and the group which did not receive the management program on the cognitive component of the NYU-CIBIC-Plus. Also, the other component of part 1 of the NYU-CIBIC-Plus, the direct subject assessment behavioral component, did not show significant changes between the management and non-management groups. However, both components of part 2 of the NYU-CIBIC-Plus, the informant based assessments, showed significant changes on the measure. Specifically, the CI-PCM subjects showed a significant benefit on the functional component in comparison with the non-management subjects (p = 0.01). Further, the informant-based behavioral component (based on the BEHAVE-AD) showed significant positive effects of the management program both in terms of total scores (p < 0.01) and in terms of the global behavioral disturbance rating (p < 0.01).

One question, which naturally arises from the above findings, is why the significant positive effects of the NYU-CIBIC-Plus behavioral disturbance components were observed on the basis of informant reports, but not on the basis of the direct subject assessment. In this case, the study investigators believe that several factors explain the disparity. One is that the study was conducted in community-residing individuals. Particularly, the non-management subjects were so disturbed (as well as impaired in other ways) that it was very difficult to
succeed in getting the informants to bring the subjects to the research center for an evaluation. Therefore, it was necessary to employ some of the management group techniques in order to be able to evaluate the subjects at the research center, which, necessarily, temporarily mitigated the subjects’ distress. The investigators also believe that the individual attention that is provided to subjects in the context of a direct subject evaluation may also temporarily mitigate observed behavioral disturbances.

The overall results of the CI-PCM study clearly indicate the sensitivity of the NYU-CIBIC-Plus assessment to change in response to nonpharmacological interventions as well as the ability of this detailed scale to separate and study clinical domains of origin of observed changes in patient behavior as well as changes in other capacities.

The BEHAVE-AD-FW

As reviewed in this publication, the BEHAVE-AD has demonstrated:

(1) Construct validity in terms of the separation of the BEHAVE-AD total scores and the 7 individual BEHAVE-AD categorical scores from the cognitive and functional symptoms of AD.

(2) Reliability in terms of BEHAVE-AD total scores and the 7 BEHAVE-AD categorical scores, even when administered using a telephone interview format.

(3) Criterion validity in terms of sensitivity of the BEHAVE-AD total scale scores and each of its elements to nonpharmacological intervention; for example, sensitivity to the added attention and other services associated with the entry into a clinical trial.

(4) Criterion validity in terms of sensitivity to pharmacological intervention over and above the effects of nonpharmacological placebo and related clinical trial services, with treatment with two classes of pharmacological agents (anxiolytic and antipsychotic medications). Furthermore, this criterion validity has also been demonstrated for both of these medication classes independent of effects of the medication on cognition.

(5) As will be discussed subsequently in this publication, criterion validity for the BEHAVE-AD has additionally been demonstrated in terms of the finding of a significant relationship between BEHAVE-AD scores based upon informant reports of behavior over the preceding 2-week interval and the actual observation of behavioral disturbances in the same subjects by independent observers over an interval of approximately 20 min [12].

Additionally, as already reviewed in this publication, the BEHAVE-AD has been very successful in the context of the scale’s usage in the approval of antipsychotic medication for the treatment of BPSD or related conditions in many nations around the world. Furthermore, as already noted, the BEHAVE-AD has been very successful in that it has been incorporated as an element of the NYU-CIBIC-Plus assessment. In this latter context, the BEHAVE-AD has been an element of a primary outcome measure in pivotal trials associated with the initial worldwide approvals of 2 of the 3 most widely used pharmaceutical agents for the treatment of cognition-related symptoms in AD in the world at this time. These medications, rivastigmine (Exelon) and memantine (Namenda, Ebixa, etc.), represent both classes of widely approved medications having a cognitive effect, i.e., the cholinesterase inhibitors and the NMDA receptor antagonist acting medication, respectively.

Despite the success of the BEHAVE-AD, because of the importance of BPSD symptomatology in terms of disturbance to persons with dementia and their support providers [116] as well as the association of these symptoms with more rapid institutionalization [117–119] and increased mortality [89–91, 120], it is desirable to obtain as much information on behavioral symptoms as possible and practical, for example, in the context of clinical trials, especially those which focus on behavioral symptoms. The BEHAVE-AD is a severity scale and, as
already noted, assesses 25 characteristic AD-related behavioral and psychological symptoms and their associated disturbances in 7 major categories. Each of these symptoms is scored on a 4-point severity scale. The severity scale used with the BEHAVE-AD incorporates the following guidelines for symptomatology over the prior 2-week interval: 0 signifies the absence of the symptom; 1 indicates the presence of the symptom; 2 indicates the presence of the symptom, generally with an emotional component, and 3 indicates the presence of the symptom, generally with an emotional and physical component.

The more recently developed BEHAVE-AD-FW incorporates all elements of the traditional BEHAVE-AD [1, 2]. The 2-week time frame of symptomatic assessment used for the BEHAVE-AD is also maintained. Two weeks is an adequate time frame for active symptomatic assessment. It is also a proper time frame for accurate symptomatic assessment. Two weeks is a sufficiently brief period for accurate retrospective reflection and recall of intercurrent events. A longer period, such as 1 month, is a much more difficult interval in terms of accurate behavioral retrospective reflection.

The methodology and scoring for the BEHAVE-AD-FW are as follows: (1) The severity rating is established for each symptomatic item; (2) the frequency at the severity rating for the symptomatic item is then scored, and (3) the frequency scale is as follows: 1 = once, 2 = every several days, 3 = daily, and 4 = more than once daily.

As noted above, the preceding 2-week interval is assessed for each item. For frequency levels 3 and 4, the level is scored if the behavior at the assessed severity level occurs on the majority of the days over the prior 2 weeks. For one symptomatic item, ‘Day/Night disturbance’, a 3-point frequency rating is used, since ‘complete disturbance of diurnal rhythm’ cannot occur more frequently than ‘daily’.

The BEHAVE-AD-FW is scored by multiplying the severity score by the frequency score for each symptomatic item. The categorical and total scale scores are obtained by adding the frequency-weighted severity scores for the symptomatic items. The global rating for the BEHAVE-AD-FW is scored in the same manner as for the traditional BEHAVE-AD severity scale.

A reliability study was conducted on the BEHAVE-AD-FW with respect to the added elements and in comparison with the traditional BEHAVE-AD [2]. Details of this study were published in Monteiro et al. [2], and a summary is provided herein. The reliability study was conducted on regularly scheduled baseline or follow-up visits of subjects at the Aging and Dementia Research Center of the New York University School of Medicine.

Information was obtained on 28 subjects [14 men and 14 women, mean age 73.5 ± 7.9 (standard deviation, SD) years]. Subject severity ranged from mildly impaired (GDS stage 3) to severe dementia (GDS stage 7). The MMSE scores ranged from 28 to 0 (mean = 18.8, SD = 7.8). Five subjects had MCI, 22 subjects had probable AD, and 1 subject had possible AD [32]. Information on the subjects was obtained from informants who accompanied the subjects. These informants were spouses (n = 12), adult children (n = 10), close friends (n = 3), or paid professional caregivers (n = 3).

Two experienced physicians performed the evaluations. One physician interviewed the informant(s), and both clinicians independently rated the responses. In general, the order of the interviewers was alternated. At the conclusion of the interview, the observing clinician had the opportunity to clarify residual queries. The interviewing clinician and the observing clinician were not permitted to discuss any scoring-related matters with each other, and all scoring was performed independently.

ICCs were calculated for 3 measures using the method of Shrout and Fleiss [121]. These measures were: (1) unweighted severity scores (i.e., the traditional BEHAVE-AD severity scores); (2) the absolute frequency scores for each BEHAVE-AD item, and (3) the frequency-weighted scores for each item (i.e., the severity score for the item multiplied by the absolute frequency score for the item).
For the traditional BEHAVE-AD severity scores in this study, the ICCs ranged from 0.74 (for BEHAVE-AD category E – Diurnal Rhythm Disturbances) to 0.97 (for BEHAVE-AD category B – Hallucinations). All ICCs for the 7 categories were significant (all p values <0.001). For the total BEHAVE-AD severity scores, the ICC was 0.90 (p < 0.001).

For the unweighted, absolute frequency values using the 4-point (or, in the case of day/night disturbances, 3-point) scoring system described earlier in this section, the ICCs for the 7 BEHAVE-AD categories ranged from 0.86 (for BEHAVE-AD category C – Activity Disturbances) to 0.97 (for BEHAVE-AD category F – Affective Disturbances). All ICCs for the 7 BEHAVE-AD categories were significant (all p values <0.001). For the total BEHAVE-AD unweighted frequency score, the ICC was 0.96 (p < 0.001).

For the frequency-weighted severity scores in this study (i.e., the values obtained by multiplying the absolute frequency score by the magnitude of the severity scores for each item of the BEHAVE-AD), the ICCs for the 7 BEHAVE-AD categories ranged from 0.69 (for BEHAVE-AD category E – Diurnal Rhythm Disturbances) to 0.98 (for BEHAVE-AD category B – Hallucinations). All BEHAVE-AD frequency-weighted severity scores for the 7 BEHAVE-AD categories were significant (all p values <0.001). For the total BEHAVE-AD, the ICC for the frequency-weighted severity score was 0.91 (p < 0.001).

Hence, the study of Monteiro et al. [2] clearly indicates that the BEHAVE-AD-FW is a very reliable instrument. The reliability for the additional frequency-weighting absolute values for the total of the 25 BEHAVE-AD items is excellent, as is the reliability of the total BEHAVE-AD-FW frequency-weighted scale scores. These excellent reliabilities are comparable to those of the traditional BEHAVE-AD.

The BEHAVE-AD-FW is of clear relevance since clinicians who try to treat behavioral disturbances endeavor to decrease both the severity of the disturbance and the frequency of the disturbance. Additionally, for the purpose of pharmacological trials, the BEHAVE-AD-FW is likely to have approximately 2.5 times the sensitivity of the traditional BEHAVE-AD. This estimate is based upon the observation from the BEHAVE-AD-FW reliability study that the total BEHAVE-AD-FW mean scores were 2.5 times those of the mean BEHAVE-AD severity scores for the 28 study subjects [2]. Hence, we believe that the BEHAVE-AD-FW can be very useful for both pharmacological and nonpharmacological behavioral disturbance remediation studies in patients with AD and related dementias (see table 4 for a succinct review of the characteristics, current usage, and continuing promise of the BEHAVE-AD-FW assessment instrument).

The E-BEHAVE-AD

From a clinician’s perspective, it is desirable not to solely rely upon informant reports in the context of the treatment of behavioral pathological symptoms in AD. For example, caregiver reports can be influenced by various factors. These include:

1. Emotional problems directly relating to the caregiving process including burden and guilt.
2. The stresses of caregiving.
3. The nature of coping styles on the part of the caregiver.
4. In the case of family caregivers, lifelong relationship patterns.
5. Pathological behaviors in the caregivers, evoked by the patient’s illness, such as caregiver depression, caregiver aggression, and caregiver abuse [122–124].
6. Differential concerns with respect to caregiver issues. For example, nursing staff has been found to be relatively concerned with care-related matters such as patients resisting...
Table 4. The BEHAVE-AD-FW: conceptual framework, methodology, construct validity, reliability, criterion validity, and utility *

A Conceptual framework

The BEHAVE-AD has been demonstrated to be a valid, reliable, and sensitive scale. The BEHAVE-AD has demonstrated sensitivity to both nonspecific psychological interventions and further sensitivity to the additional effects of pharmacological intervention with two classes of pharmacological medications, i.e., anxiolytic medication and antipsychotic medication. Despite the success of the BEHAVE-AD and its clear utility, there is a need for increased sensitivity of the measure, if possible. We therefore developed a frequency-weighted scale dimension for concurrent assessment with the BEHAVE-AD.

B Methodology

The elements of the previously developed BEHAVE-AD assessment are used in their entirety. These elements include the 2-week time frame of symptomatic assessment. They also include the 25 BEHAVE-AD questions assessed in the 7 symptomatic categories on the 4-point severity scale. Additionally, a frequency-weighted score is added for each item of the traditional BEHAVE-AD. For 24 of the 25 BEHAVE-AD symptomatic items, the frequency dimension is assessed on a 4-point scale. One symptomatic item, 'day/night disturbance', is assessed on a 3-point frequency score. The frequency ratings are all referenced to the prior 2-week interval, and the information is obtained from the same knowledgeable informant from whom the severity information is obtained, immediately after the severity level is established. The frequency is referenced to the symptomatic item severity level. The frequency scoring is: 1 = once (over the preceding 2-week interval); 2 = every several days; 3 = daily, and 4 = more than once daily. Conceptually and procedurally, the frequency level '4' is not applicable for day/night disturbance, hence this item is scored on a 3-point frequency scale. The frequency-weighted score for each item is obtained by multiplying the severity item score by the frequency score. The frequency-weighted score for each BEHAVE-AD-FW category is obtained by adding the frequency-weighted scores for the category items. Similarly, the total scale frequency-weighted score is the sum of the frequency-weighted scores of the 25 BEHAVE-AD items.

C Construct validity

For spouses, professional caregivers, friends, and others in contact with individuals with dementia or with the antecedents of dementia, such as MCI, the frequency of a behavioral disturbance adds to the severity of the disturbance in terms of the actual burden of their social contact. Therefore, a reliably obtained frequency dimension should add to the severity assessment in terms of the full measure of social burden (disturbance). For treatment trials, frequency-weighted symptomatic assessments together with severity symptomatic assessments should potentially provide greater sensitivity to therapeutic efficacy.

D Reliability

<table>
<thead>
<tr>
<th>Authors [Ref.]</th>
<th>Methodology</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monteiro et al. [2]</td>
<td>Informants of 28 subjects with MCI (n = 5) or dementia (n = 23) were interviewed. The dementia subjects had probable AD (n = 22) or possible AD (n = 1). The subjects consisted of 14 men and 14 women, with a mean age of 73.5 (SD = 7.9) years. The MMSE scores ranged from 28 to 0 (mean 18.8 ± 7.8). The informants were from one of the following categories: spouse (n = 12), one or more adult children (n = 10), a close friend of the subject (n = 3), or a paid professional caregiver (n = 3). Subjects were baseline or follow-up participants in longitudinal studies at the New York University (NYU) School of Medicine’s Aging and Dementia Research Center and the NYU US National Institute on Aging (NIA)-supported Alzheimer’s Disease Center. Two experienced clinicians interviewed the informants, generally on an alternating basis.</td>
<td>For all 3 methodologies studied, all 7 symptomatic categories and total scores showed significant correlations between the scores for the 2 clinicians (all p values &lt;0.001). For the traditional BEHAVE-AD severity scores, the correlations between the ratings of the 2 clinicians for the 7 BEHAVE-AD symptomatic categories ranged from 0.74 (for Diurnal Rhythm Disturbances) to 0.97 (for Hallucinations) (all p values &lt;0.001). For the total BEHAVE-AD severity scores, the ICC for the scores of the 2 raters was 0.90 (p &lt; 0.001). For the unweighted frequency scores, the correlation between the ratings of the 2 clinicians for the 7 symptomatic categories ranged from 0.86 (for Activity Disturbances) to 0.97 (for Affective Disturbances). For the total BEHAVE-AD-FW absolute frequency scores, the ICC for the 2 raters was 0.96 (p &lt; 0.001).</td>
</tr>
</tbody>
</table>
care or bathing. In contrast, it has been found that family members report more affective symptoms such as depression [125].

(7) Differential observations and experiences of symptomatology. For example, dementia patient reports resulted in a DSM-III diagnosis of major depression in 14% of patients based on the patients’ own reports and in 50% based on family reports [126]. Similar findings have been observed by other investigators [127].

Therefore, from the perspective of both clinically based treatment needs and associated behavioral treatment research, it was desirable to have a rating scale, which could, in a relatively brief period, directly assess pharmacologically remediable behavioral and

Table 4 (continued)

<table>
<thead>
<tr>
<th>Authors [Ref.]</th>
<th>Methodology</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monteiro et al. [2]</td>
<td>Both clinicians were present for all informant interviews and independently rated the responses based upon the information obtained by the designated interviewer. The observing clinician had an opportunity to clarify responses from the informant at the conclusion of the designated clinician interview. No discussion of scoring was permitted between the clinicians. The reliability study examined ICCs with respect to 3 components of the BEHAVE-AD-FW scoring: (1) traditional BEHAVE-AD severity scores, (2) absolute frequency values, and (3) frequency-weighted severity scores for the BEHAVE-AD-FW. Finally, for the frequency-weighted severity scores, the correlation for the 7 symptomatic categories of the scores of the 2 clinicians ranged from 0.69 (for Diurnal Rhythm Disturbances) to 0.98 (for Hallucinations). For the entire BEHAVE-AD-FW, the correlation between the scores of the 2 raters was 0.91 (p &lt; 0.001). Hence, in absolute terms, the reliability of the BEHAVE-AD-FW was found to be even greater than that of the traditional BEHAVE-AD. Furthermore, since the BEHAVE-AD-FW adds an additional dimension of observation and results to the total scores, which are approximately 2.5 times greater than those of the traditional BEHAVE-AD, we hypothesize that the BEHAVE-AD-FW is approximately 2.5 times as sensitive to therapeutic effects of interventions as the traditional, BEHAVE-AD without the frequency dimension.</td>
<td></td>
</tr>
</tbody>
</table>

E Criteria validity

The BEHAVE-AD-FW assessment instrument contains all elements of the traditional BEHAVE-AD assessment. Therefore, all studies which have demonstrated the criterion validity of the traditional BEHAVE-AD instrument in terms of sensitivity to both nonpharmacological (e.g., placebo related) and pharmacological intervention [e.g., 24–28, 100] apply to the BEHAVE-AD-FW. Additionally, the BEHAVE-AD-FW adds a frequency rating dimension which has been demonstrated to be reliable [2] and which should add an approximately 2.5 times greater sensitivity to the traditional BEHAVE-AD [2].

F Utility

The BEHAVE-AD-FW has been used to identify the relationship between an individual delusional symptom and the cognitive course of AD [5]. The BEHAVE-AD-FW has also been used to uncover the relationship between a particular delusional symptom in AD and changes in blood flow in a brain region associated with emotionality [6]. Hence, the BEHAVE-AD-FW assessment is beginning to fulfill the goals and promise of this instrument in the sensitive identification of individual delusional symptoms in AD. However, perhaps the greatest promise of the BEHAVE-AD-FW assessment is the more sensitive identification of treatment responsiveness in AD. The BEHAVE-AD-FW has apparently not yet been utilized for this important purpose, which can serve to advance treatment possibilities for AD.
psychological symptoms and associated disturbances in AD patients. Additionally, since the behavioral and psychological symptoms in AD differ from those in other disorders such as schizophrenia or major depression, it was necessary to have a direct patient observational instrument which evaluates the pharmacologically remediable behavioral and psychological symptoms of AD. We therefore developed such an instrument, i.e., the E-BEHAVE-AD [12].

The E-BEHAVE-AD was developed primarily using the caregiver informant-based BEHAVE-AD as a symptomatic model. Additionally, 3 of the 12 symptomatic examples in the E-BEHAVE-AD Aggressiveness category and 4 of the 9 symptomatic examples from the E-BEHAVE-AD Affective Disturbance category come from Teri et al. [128]. Six of the 7 informant-based BEHAVE-AD symptomatic categories are assessed with the E-BEHAVE-AD. Specifically, the E-BEHAVE-AD assesses: (A) Paranoid and Delusional Ideation, (B) Hallucinations, (C) Activity Disturbance, (D) Aggressiveness, (E) Affective Disturbance, and (F) Anxieties and Phobias. The informant-based BEHAVE-AD symptomatic category which is not assessed with the E-BEHAVE-AD is Diurnal Rhythm Disturbances (BEHAVE-AD category E), since it is not possible to directly observe this symptom during the brief direct observational time frame of the E-BEHAVE-AD assessment. Within these 6 psychological dimensions, the E-BEHAVE-AD assesses 12 symptoms. Each symptom is assessed on a 4-point severity scale where ‘0’ signifies the absence of the symptom during the observation period, and ‘3’ indicates the presence of the symptom with a severe magnitude.

Each of the 12 E-BEHAVE-AD symptomatic questions includes several symptomatic examples. The great majority of these symptomatic examples come from the traditional informant-based BEHAVE-AD. Two of the 12 E-BEHAVE-AD items have 8 symptomatic examples for each item, including the ‘other’ exemplar; 2 E-BEHAVE-AD symptomatic items have 7 exemplars; 1 E-BEHAVE-AD symptomatic item has 5 exemplars; and 6 items have 4 exemplars. The remaining E-BEHAVE-AD symptomatic item, i.e., ‘tearfulness or crying (with or without tears)’, has the single exemplar of ‘other’. The total time taken to administer the E-BEHAVE-AD instrument is approximately 20 min.

Two studies were conducted in association with the initial publication of the E-BEHAVE-AD. One was a reliability study, and the other was a study of the relationship between the direct observationally based E-BEHAVE-AD assessment and the traditional, informant-based BEHAVE-AD assessment.

The E-BEHAVE-AD reliability study included 20 dementia subjects (9 men, 11 women, mean age 73.2 ± 8.0 years). The diagnostic distribution of the subjects was as follows: AD [32], n = 15; dementia associated with cerebrovascular factors, n = 1, and dementia associated with diverse disturbances and/or etiologies, n = 4. On the basis of the GDS, the severity distribution was as follows: mild dementia (GDS stage 4), n = 3; moderate dementia (GDS stage 5), n = 3; moderately severe dementia (GDS stage 6), n = 13, and severe dementia (GDS stage 7), n = 1. The mean MMSE scores were 11.6 (SD = 9.4).

The procedures for the reliability study included the following: (1) all subjects were participants in US National Institute on Aging-sponsored investigations at the New York University School of Medicine; (2) evaluations were conducted by experienced clinicians; (3) one rater independently interviewed the subject while the second rater observed; (4) the interviewing rater was alternated randomly; (5) there was no communication between the raters with respect to the scoring, and (6) all scoring was carried out by each rater independently. Intraclass reliability correlations were used to determine the interrater reliabilities of the raters for each individual E-BEHAVE-AD item and for the total E-BEHAVE-AD scores. Two models were examined: raters-fixed ICCs and raters-random ICCs.

The results of the E-BEHAVE-AD reliability study indicated that the E-BEHAVE-AD is a very reliable measure. One of the 12 E-BEHAVE-AD items could not be evaluated because of
absence of variance (specifically, none of the subjects were observed to have hallucinations).
The ICCs for the 11 remaining items were all significant \( (p < 0.01) \) using both models. Using
the raters-fixed model, the ICCs ranged from 0.76 to 1.00 \( (\text{for 3 items}) \). For the entire
E-BEHAVE-AD scale score, including the 12 E-BEHAVE-AD items, the ICC using the raters-
fixed model was 0.97 \( (p < 0.01) \). Very similar results were obtained using the raters-random
model: for the 11 evaluable individual items, the ICCs ranged from 0.84 to 1.00 \( (\text{for 3 items}) \).
For the entire E-BEHAVE-AD scale scores, the intrarater reliability for the total scale scores
using the raters-random model was 0.97 \( (p < 0.01) \). Hence, it is clear that the E-BEHAVE-AD
is a very reliable instrument for the direct measurement of behavioral disturbances based
upon direct subject evaluation.

The second study reported in association with the initial public publication describing the
E-BEHAVE-AD assessment examined the relationship between the E-BEHAVE-AD direct
assessment of subject symptomatology and the BEHAVE-AD informant-based assessment of
dementia and cognitive decline-associated behavioral and psychological symptom-based
disturbances \[12\]. For this comparative study, 49 subjects and their informants were assessed.
Five subjects were clinically normal \( (\text{GDS stage 1 or 2}) \), 5 subjects had MCI \( (\text{GDS stage 3}) \),
and the remaining 39 subjects had dementia \( (\text{GDS stages } \geq 4) \). Of the 39 dementia subjects, 27
subjects had probable AD \[32\], 6 had dementia associated with cerebrovascular disease, 1
had dementia associated in part with a prolonged vitamin B12 deficiency, and 5 subjects had
dementia associated with diverse possibly etiologically relevant processes. The subject
severity range in this study was from stage 1 \( (\text{no objective or subjective cognitive impairment}) \)
to stage 7 \( (\text{severe dementia}) \) on the GDS. The mean MMSE score was 18.3 \( (SD = 9.7) \). During
a single subject and informant visit, subjects were evaluated on the E-BEHAVE-AD assessment,
and their informants provided information for the BEHAVE-AD assessment. During this visit,
different raters obtained information with respect to the observer-based E-BEHAVE-AD and
the informant-based BEHAVE-AD.

In order to evaluate the relationship between the observer-rated E-BEHAVE-AD and the
informant-based BEHAVE-AD, which have different numbers of items assessing 6 compa-
rable categories of symptoms, the scales were compared with respect to the presence or
absence of symptomatology in each of the 6 comparable symptom categories and with
respect to the presence or absence of symptoms in the total of the 6 comparable symptom
categories. The Pearson correlation coefficients were computed for the presence versus
absence scores of the E-BEHAVE-AD and BEHAVE-AD analogous symptomatic categories.

The results of these comparisons were as follows. For the Hallucinations symptomatic
category, the correlation coefficient could not be calculated since no hallucinations were
observed using the E-BEHAVE-AD, although hallucinations were reported by informants
using the BEHAVE-AD. For the remaining 5 symptom categories, significant relationships
were observed between the observer-based E-BEHAVE-AD symptoms and the informant-
based BEHAVE-AD symptoms. Specifically, the Pearson correlations ranged from 0.32 \( (p <
0.05) \) for the Affective Disturbances symptom category to 0.44 \( (p < 0.01) \) for the Aggres-
sivity symptom category. The correlations between the observer-rated and the informant-
based symptoms for the remaining symptomatic categories \( (\text{Paranoid and Delusional}
Ideation, Activity Disturbances, and Anxiety and Phobias) \) ranged from 0.37 to 0.41 and were
also significant \( (p < 0.01) \). For the total observer-based, E-BEHAVE-AD symptomatic presence
or absence scores and the total informant-based, BEHAVE-AD symptomatic presence or
absence scores for the 6 comparable symptom categories, the Pearson correlation was
0.51 \( (p < 0.01) \).

It is, of course, both clinically interesting and important from research and other perspec-
tives that observations of behavioral disturbances assessed over a period of approximately
20 min correlate very significantly with informant reports of symptomatology over a
preceding interval of 2 weeks. It is also interesting and important from both clinical and research perspectives that the majority of variance between these disparate symptomatic assessment modalities is non-overlapping. In introducing the rationale for the observer-rated E-BEHAVE-AD assessments, we discussed some of the reasons for examining behavioral symptoms from these different perspectives. The results of this comparative study [12] also suggest additional interesting and important perspectives regarding this variance. For example, the strongest categorical symptomatic relationship between the observer-based E-BEHAVE-AD assessments and the informant-based BEHAVE-AD assessments was seen for the Aggressivity symptomatic category. This means that aggressiveness seen by informants over the preceding 2 weeks is frequently observed over a brief 20-min observation period. However, most of the variance in this symptomatic expression is non-overlapping. Part of the difference, of course, may be the observational time differences. Nevertheless, it is also possible that aggressiveness is only expressed when the subject is called upon to perform activities of daily life, such as dressing or bathing, which would not be observed in a clinic setting. It is also possible that agitation is only observed when the subject is incontinent. Additionally, the more formal clinic setting might itself be a deterrent to aggressiveness which the subjects might express more openly when they are at home.

The weakest categorical symptomatic relationship between the approximately 20-min observer-based E-BEHAVE-AD assessment and the informant-based report of symptomatology over the preceding 2 weeks was seen for the Hallucinations category of behavioral symptoms. Hallucinations were not observed in the approximately 20-min E-BEHAVE-AD observation period; however, they were reported by the informants. A variety of explanations for these discrepancies are possible. For example, subjects with dementia may go downstairs in their home and see their spouse. Upon returning upstairs, they might report to their professional caregiver that they saw their father downstairs. The caregiver might interpret this report as a hallucination, whereas a clinician who understood the whole situation might interpret the subject report as a misidentification, rather than a hallucination. Therefore, there appear to be clear reasons for comprehensively assessing BPSD using both observer- and informant-based evaluation modalities.

A recent study of the utility of observer- and informant-based scales in the context of pharmacological treatment of agitation and psychosis in AD and related dementias provides further data with respect to the robust effects seen with the brief clinician interview-based E-BEHAVE-AD [129]. This study used a standard of improvement as the Clinical Global Impression of Severity (CGI-S) scale [130]. The CGI-S scale is an observer-rated instrument in which a clinician assesses the overall severity of the patient’s illness, analogous in many ways to the NYU-CIBIC-Plus and other global severity instruments discussed earlier in this review. However, the CGI-S scale is much less structured than the CIBIC assessment instruments, being only a simple score of the severity of illness on a 1- to 7-point scale. The E-BEHAVE-AD assessment was compared with the Neurobehavioral Rating Scale (NBRS) [131], a 27-item observer-rated instrument which is based on the BPRS assessment [132] mentioned briefly earlier in this review [49]. The E-BEHAVE-AD assessment was also compared with the NPI [29], discussed in greater detail in the subsequent section of this review. The NPI scale version used in this study was the NPI-12 [30]. The NPI assesses symptomatology based upon informant information over a preceding interval of 1 month. This study concluded that ‘In the ROC analysis for overall clinical improvement in response to treatment, there were no differences in the areas under the correlated curves for the three instruments but they demonstrated different sensitivity and specificity at different cut-off points for target symptoms reduction. The E-BEHAVE-AD performed best at a cut-point of 30% target symptom reduction and the NBRS and NPI both performed best at 50%.’ Hence, it appears from this study of Ismail et al. [129] that the approximately 20-min observer-based
E-BEHAVE-AD assessment obtained as much information in terms of the study outcome measures as the informant-based rating instrument (i.e., the NPI-12), examining a time window of 1 month. However, methodological questions, such as the use of an observer-based outcome measure standard in the study of Ismail et al. [129], should be noted as caveats to the aforementioned conclusion, and these issues should be examined in future studies with less methodologically biased outcome standards.

As noted earlier, the E-BEHAVE-AD is a component of the comprehensive NYU-CIBIC-Plus global assessment measure [13, 14]. More specifically, the E-BEHAVE-AD is incorporated in part 1, the subject- (patient-) based assessment portion of the NYU-CIBIC-Plus assessment as the behavioral component. In this context, the E-BEHAVE-AD has been used as part of a primary outcome measure in the pivotal trials associated with the initial approvals of 2 of the 3 most widely used medications for the treatment of AD at the present time, namely, rivastigmine (Exelon) [15, 16] and memantine (Namenda, Ebixa, Axura) [17]. Since these medications have been approved primarily because of their effects on cognition and functioning in AD and since there was no subanalysis of the NYU-CIBIC-Plus assessments, it cannot be assumed that there was a change in the E-BEHAVE-AD assessment in these studies.

In a recent study where the NYU-CIBIC-Plus component measures were separately examined in relationship to therapeutic effects of an intervention, the value of the observer-based E-BEHAVE-AD in providing a separate therapeutic perspective was demonstrated [106]. Specifically, in this study of a CI-PCM program, very robust and strongly significant effects were observed on the informant-based BEHAVE-AD assessment (p < 0.01). However, no significant effects were observed with the direct observation E-BEHAVE-AD measure. The possible reasons for this disparity between these two different views of behavior have been reviewed earlier in this section. The results of this study clearly indicate the scientific and clinical virtues of behavioral assessment from both the perspective of the informant and the direct observation of the subject in the context of providing a comprehensive understanding of the nature and value of a therapeutic intervention [106]. Table 5 provides a succinct review of the E-BEHAVE-AD assessment measure, including the conceptual framework, methodology, construct validity, reliability, criterion validity, and utility of this rating scale.

### Differences between the NPI and the BEHAVE-AD and the BEHAVE-AD-FW

In recent years, the NPI [30] has been widely used. Hence, it is important to highlight differences between the NPI and the BEHAVE-AD and BEHAVE-AD-FW measures.

#### Assessment of Individual Behavioral Symptoms with the BEHAVE-AD and BEHAVE-AD-FW and Assessment of Symptomatic Categories with the NPI

The NPI differs from the BEHAVE-AD and the BEHAVE-AD-FW in several important respects. First, ‘the NPI uses a screening methodology. If the screening questions are positive, more extended questioning is pursued; if they are negative, the domain is not explored in depth’ [133]. However, when the domain is explored in depth, the individual symptoms are not measured with the NPI, only the impact of the symptomatic category as a whole is assessed. These procedures are very different from those of the BEHAVE-AD and the BEHAVE-AD-FW. The BEHAVE-AD and the BEHAVE-AD-FW are specifically designed to assess potentially remediable (from both nonpharmacological and pharmacological intervention perspectives) behavioral symptoms in AD and closely related dementias. As described below, the great majority of these BEHAVE-AD and BEHAVE-AD-FW symptoms are relatively common in the context of AD. With the traditional BEHAVE-AD, each of these symptoms is separately assessed in terms of severity, and with the BEHAVE-AD-FW, these symptoms are separately assessed
A Conceptual framework

For numerous reasons, it is desirable not to rely solely on caregivers for the assessment of behavioral disturbances. Caregiver reports may be influenced by the emotional stress of caregiving, by the needs of the institution, by the burdens of caregiving, by the nature of the relationship with the subject, etc. Therefore, for both clinical and research purposes, it is desirable to have a direct assessment of behavioral disturbances in individuals with brain aging and the progression of AD.

In addition, since it has been demonstrated that behavioral symptoms in patients with AD and, to some extent, with MCI and even SCI differ from those of other conditions, such as major depression or schizophrenia [22], it is desirable to have a rating instrument which can directly assess behavioral symptomatology in the brain aging to AD continuum.

B Methodology

Using the informant-based BEHAVE-AD as a model [1], an assessment instrument was developed which could be used for the direct evaluation of the subject. Six of the 7 BEHAVE-AD symptomatic categories were successfully adapted for the direct observation format. One category from the BEHAVE-AD, Diurnal Rhythm Disturbances, was not adapted because it was not possible to observe these symptoms in the context of a brief clinical interview. Twelve symptoms in the 6 behavioral symptomatic categories were identified. Up to 8 exemplars for each symptom are provided. The great majority of symptom exemplars are derived from the BEHAVE-AD. Seven of the 60 exemplars are derived from Teri et al. [128].

The 12 E-BEHAVE-AD items are scored on a 4-point severity scale. The general criteria for this scale are similar to that employed for the informant-based BEHAVE-AD assessment. Specifically, an item is scored as ‘0’ if no symptomatology with respect to the item is observed, as ‘1’ if the symptom is present in a mild form, as ‘2’ if the symptom is present in moderate form, such as with an emotional component, and as ‘3’ if the symptom is present in a severe form, such as with an emotional and physical component. Specific definitions appropriate to the individual items are provided.

In addition to a total score for the 12 items, the E-BEHAVE-AD has a global score assessment. This global score rating is also on a 4-point scale, which is similar to that employed for the individual E-BEHAVE-AD items. For example, a score of ‘0’ indicates no observable behavioral problems seen during the interview and ‘an absence of problems which are likely to be disturbing to the patient’s caregivers and/or disturbing or dangerous to the patient’. Similarly, the ‘1’ = mild, ‘2’ = moderate, and ‘3’ = severe global scores are anchored to the criteria of disturbance to the caregivers and disturbance or danger to the subject (patient).

The E-BEHAVE-AD also contains instructions for the methodology of the assessment. For example, the interview should be conducted only with the rating clinician and the subject present, in the absence of caregivers or other informants. Also, the interview should be conducted as a conversation with the subject (patient). The conversation can be a dialogue and should include more general topics such as the subject’s activities or hobbies. The colloquy should last approximately 20 min.

C Construct validity

Since the E-BEHAVE-AD symptoms and scoring methodology are derived from the informant-based BEHAVE-AD, the construct validity features described for the BEHAVE-AD are also applicable for the E-BEHAVE-AD. These include the differentiating features from the cognitive and functional symptoms of AD. In addition, a criterion validity study was published describing the relationship between the observer-rated E-BEHAVE-AD symptoms and the informant-based BEHAVE-AD symptoms in 49 subjects. This study is described below.
Table 5 (continued)

D  Reliability

<table>
<thead>
<tr>
<th>Authors [Ref.]</th>
<th>Methodology</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auer et al. [12]</td>
<td>Twenty community-residing patients were evaluated by 2 raters. The sample comprised 11 women and 9 men, with a mean age of 73.2 years (SD = 8.0). Fifteen patients had AD, 1 had dementia associated with cerebrovascular disease, and 4 had dementia associated with diverse disturbances and/or etiologies. The mean MMSE score was 11.6 (SD = 9.4). On the GDS, subjects had the following severity distribution: GDS stage 4 (mild dementia), n = 3; GDS stage 5 (moderate dementia), n = 3; GDS stage 6 (moderately severe dementia), n = 13, and GDS stage 7 (severe dementia), n = 1. The 2 raters independently evaluated the patients on the E-BEHAVE-AD. One rater performed the interview while the other rater observed. The order of interviewer and rater was alternated. There were no communications between the raters with respect to scoring, and all scoring was done independently. Intraclass correlation was used to determine the interrater reliability between the raters.</td>
<td></td>
</tr>
</tbody>
</table>

For the total E-BEHAVE-AD, the ICC for the ratings of the 2 examiners was 0.97 (p < 0.01) using both raters-fixed and raters-random methods. For the individual BEHAVE-AD items, the ICCs ranged from 0.76 to 1.00 (there was perfect agreement for 3 of the 11 items for which a correlation coefficient could be computed) using the raters-fixed procedure. For one item, 'hallucinations', the correlation could not be computed because of absence of variance on the item since the symptom was not observed in the 20 subjects evaluated. All correlations for the 11 observed E-BEHAVE-AD symptoms were significant (p < 0.01). Using the raters-random procedure, the ICCs ranged from 0.73 to 1.00 for the 11 items for which correlations could be calculated. Also, using the raters-random procedure, all correlations were significant for the 11 E-BEHAVE-AD symptoms observed (p < 0.01). Hence, the E-BEHAVE-AD showed excellent interrater reliability. |

E  Criterion validity

<table>
<thead>
<tr>
<th>Authors [Ref.]</th>
<th>Methodology</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auer et al. [12]</td>
<td>The BEHAVE-AD is an informant-based rating scale, whereas the E-BEHAVE-AD is a direct observation-based rating instrument. Depending in part on the perspective, including the needs of the subject (patient) and the needs of the persons who live with, or work with, or interact with the subject, either an informant-rated or a subject-based assessment instrument can be considered the standard for criterion validity. In addition, for other reasons, it was important to know the relationship between the observer-based subject symptomatology and the informant-based symptomatology. Therefore, Auer et al. [12] conducted a study of the relationship between symptomatic assessments using the informant-based BEHAVE-AD [1] and the direct observer-based E-BEHAVE-AD [12]. In accord with the standard rating scale procedures, information regarding the BEHAVE-AD was obtained from a knowledgeable informant with respect to the prior 2-week period, whereas information with respect to the E-BEHAVE-AD was obtained by a clinician observer in the course of a colloquy with the subject lasting approximately 20 min. Pearson correlation coefficients were used to determine the agreement between the informant reports of symptomatology and the clinicians’ observation of symptomatology in each of the 6 symptomatic areas. Overall, for the 6 analogous symptomatic categories, there was a 0.51 correlation between the informant-based BEHAVE-AD symptomatic reports and the E-BEHAVE-AD symptomatic observations (p &lt; 0.01). For one of the symptomatic categories, Hallucinations, correlations could not be computed since these symptoms were never observed by the clinician. For the remaining 5 symptomatic categories, significant (p &lt; 0.01) relationships were observed for all symptomatic areas. The Pearson correlations ranged from 0.32 for Affective Disturbances to 0.44 for Aggressiveness.</td>
<td></td>
</tr>
</tbody>
</table>
Auer et al. [12] A total of 49 subjects and informants (mean age, 72.5 years, SD = 8.4; 22 women and 27 men) were studied. All subjects were participants in research longitudinal and pharmacologic studies conducted at the New York University Medical Center’s Alzheimer’s Disease Center. Subjects were healthy apart from the presence of MCI (n = 5) or dementia (n = 39). The remaining subjects had no cognitive impairment or SCI (n = 5). Of the dementia subjects, 27 had probable AD [32], 6 had dementia associated with cerebrovascular disease, and 6 had dementia associated with diverse etiologies or processes. The mean MMSE score of the subject population was 18.3 (SD = 9.7). All ratings were obtained in the course of a single study visit by 2 different raters. One rater assessed the BEHAVE-AD based upon the informant reports, while the other rater assessed the E-BEHAVE-AD based upon subject observation. Each rater was unaware of the other rater’s evaluations. The 6 comparable BEHAVE-AD and E-BEHAVE-AD symptomatic categories were examined. Because the Diurnal Rhythm Disturbances symptomatic category of the BEHAVE-AD is not assessed with the E-BEHAVE-AD, this symptomatic category was excluded from the analyses. In addition, because the number of items assessed is not comparable in the BEHAVE-AD and the E-BEHAVE-AD symptomatic categories, the presence or absence of symptoms in each of the 6 comparable symptom categories was compared with the 2 assessment procedures.

In conclusion, there are significant relationships between behavioral disturbances observed in brain aging, MCI, and dementia (primarily AD and closely related conditions) using an approximately 20-min direct subject evaluation (the E-BEHAVE-AD) and an informant-based report of symptomatology over the prior 2 weeks. However, these methodologies are also, to some extent, measuring different constructs.

**Utility**

The E-BEHAVE-AD can be very useful in providing an independent view of behavioral symptoms of dementia patients for clinical and research purposes. The major usage in worldwide studies of the E-BEHAVE-AD assessment until the present time has been in the context of the role of the E-BEHAVE-AD as part of the NYU-CIBIC-Plus assessment [13, 14]. In this context, the E-BEHAVE-AD has been used as part of a primary outcome measure in the pivotal trials associated with the initial approvals of 2 of the 3 most widely used medications for the treatment of AD, namely, rivastigmine (Exelon) [15, 16] and memantine (Namenda, Ebixa, Axura) [17]. However, there was no separate analysis of the CIBIC-Plus constituents in these studies. A recent study of a management program intervention examined the NYU-CIBIC-Plus component measures in terms of their contribution to efficacy [106]. The informant-based BEHAVE-AD component of the NYU-CIBIC-Plus was very sensitive to the intervention and showed very robust results. This observation of robust results in the BEHAVE-AD, albeit much greater in magnitude than in previous studies, is consistent with observations in previous studies which have found robust results on the BEHAVE-AD with nonspecific placebo interventions in clinical trials. However, the E-BEHAVE-AD assessment component of the NYU-CIBIC-Plus did not show a response to the management program. Hence, the E-BEHAVE-AD assessment is potentially a very useful independent source of relevant information with respect to the treatment of behavioral symptoms in dementia patients.

* Copyright © 2013 Barry Reisberg, MD. All rights reserved.
in terms of both severity and frequency on 4-point severity and 4-point frequency scales (with the exception of one item, which is assessed on a 3-point frequency scale).

For example, for the important question with respect to delusions, the NPI has a screening question about false beliefs, people trying to steal from or harm the subject, and family members not being who they say they are. If the screening question is not endorsed, this symptomatic domain is not pursued further. If it is endorsed, a series of yes or no questions with respect to specific symptoms are presented. However, these symptoms are not assessed for frequency or severity. Only the entire symptomatic domain is assessed for frequency and severity.

In the BEHAVE-AD, each of the 7 important delusional symptoms is separately assessed on a 4-point severity scale, and in the BEHAVE-AD-FW, the 7 delusional symptoms are assessed on 4-point severity and 4-point frequency scales. The importance of these individual assessments is illustrated by current studies. For example, Katz et al. [25], using the traditional BEHAVE-AD in a large US multicenter study of the efficacy of risperidone in comparison with placebo treatment, found significant beneficial effects of the antipsychotic medication in each of the 7 individual items in the Paranoid and Delusional Ideation symptomatic category of the BEHAVE-AD. In addition, in this study, the effects of the risperidone versus the placebo treatment were independent of effects on cognition. Hence, the separate assessment of paranoid and delusional symptoms in this study clearly adds important additional information with respect to antipsychotic treatment. Also for the proper study of antipsychotic or related medications, the separate assessment of the important delusional symptoms adds additional power in terms of the assessment of the important delusional category of antipsychotic medication assessment.

As noted in the introduction to this review, the study and assessment of the individual delusions commonly occurring in the course of AD also permits the identification of important insights with respect to these symptoms. As these delusions have diverse origins, which include the cognitive, neurochemical as well as neurophysiological changes occurring in the course of AD, treatments may vary depending upon when in the disease the specific delusions occur. For example, Nakatsuka et al. [5] have found that the specific delusion ‘residence is not home’ peaks in frequency-weighted severity scores in AD patients with MMSE scores of 0–9. In contrast, an earlier study, which used the traditional BEHAVE-AD, found that the commonly occurring delusion ‘people are stealing things’ from the AD patient peaks in individuals in the moderate stage of AD (GDS stage 5) with a mean MMSE score of 14 [22]. Since both cognitively and neurochemically, AD persons in these different severity levels differ, depending upon when in the course of AD the delusions occur, the nonpharmacological and pharmacological treatments for these disparate delusions occurring at different stages in the disease may also differ in various ways.

Furthermore, Nakatsuka et al. [6], by specifically studying the ‘residence is not one’s home’ delusion using the BEHAVE-AD-FW, uncovered a potentially important physiologic finding. They found a correlation between the occurrence of this delusion and decreased cerebral blood flow in the right amygdala using SPECT scanning [6]. Since the amygdala is known to be involved in emotionality, the decreased blood flow in this region might be indicative of injury to the amygdala, which is predisposing to the emergence of the delusion. This finding might ultimately greatly enrich the understanding of pathophysiological origins of specific delusional and perhaps other behavioral symptoms in AD.
Assessment of Symptoms of Relevance for AD with the BEHAVE-AD and the BEHAVE-AD-FW and Assessment of Symptoms of Relevance Mainly for Mania, Affective Disorders, and Other Conditions with the NPI

Another major and very important difference between the NPI assessment and the BEHAVE-AD and BEHAVE-AD-FW assessments is that the BEHAVE-AD specifically assesses symptoms which: (1) occur in AD, (2) for the most part, are relatively common in the course of AD, and (3) are potentially remediable in the course of AD with nonpharmacological interventions such as management and pharmacological interventions such as antipsychotic medications. In marked contrast, ‘the NPI has scales for behaviors common to Alzheimer’s disease, but also includes scales for behavior characteristic of frontotemporal degenerations and other dementias’ [133]. However, not only does the NPI have scales for ‘other dementias’, it also has scales that clearly appear to relate mainly to affective disorders such as mania and depression.

For example, the NPI has a scale (‘F’) for ‘Elation/Euphoria’. This includes the overview scale question: ‘Does the patient seem too cheerful or too happy for no reason?’ Subscale questions include: ‘1. Does the patient appear to feel too good or to be too happy, different from his/her usual self?’; ’6. Does the patient “talk big” or claim to have more abilities or wealth than is true?,’ and ‘Does the patient show any other signs of feeling too good or being too happy?’ These questions appear to relate to mania, bipolar disorder associated with mania, or other mania-like illnesses. These questions have nothing to do with the behavioral disturbances of AD. AD patients are distressed, disturbed, or concerned by the changes in their cognitive and functional capacities. Unfortunately, this never produces a state of being ‘too happy’. Apart from clinical observation, evidence that elation/euphoria as described by the NPI has no relationship to AD symptomatology is supported by current studies. For example, Lyketsos et al. [31] studied 214 community-residing AD persons at all levels of severity and found that elation/euphoria as assessed with the NPI was absent in 99.5% of the subjects. This occurrence of elation/euphoria of 0.5% with the NPI assessment was very similar to the 0.3% occurrence observed in 673 community-residing elderly individuals who did not have dementia [31].

Another example is that the NPI has a scale for changes in appetite. This includes questions such as: ‘does the patient have a poor appetite?’; ‘has the patient had an increase in appetite?’; ‘has the patient had a loss of weight?’, and ‘has the patient gained weight?’ These changes do not occur in association with behavioral disturbances in AD. In AD, as the disease evolves, especially in the final GDS stage 7 of AD when the MMSE score is typically zero, patients do become progressively emaciated due to progressive losses in motor capacities [95–97] (which begin much earlier in the illness process [96, 97]) and associated musculoskeletal wasting [134]. However, appetite changes do not occur in AD. Under these circumstances, in the context of the treatment of the behavioral disturbances of AD with medications, the changes on this NPI scale are much more likely to pick up side effects of pharmacotherapies, including both experimental and concomitant medications, rather than any possible therapeutic effects in terms of the treatment of behavioral symptoms. The problem is that the side effects of behaviorally relevant medications for AD will be confused with therapeutic effects of the medications and hence lead to confounding or misinterpretation of possible true therapeutic effects of the medications.

In marked contrast to the NPI, the BEHAVE-AD and the BEHAVE-AD-FW assess symptoms which are characteristic of AD and, for the most part, common in AD. For example, in one study [22], 17 of the 25 symptoms identified with the BEHAVE-AD and the BEHAVE-AD-FW occurred in 25% or more of the AD persons at one or more dementia stages of the AD process, and 20 of the 25 BEHAVE-AD symptoms occurred in at least 15% of the subjects at one or more of the GDS dementia stages (i.e., GDS stage ≥4). As already extensively discussed, all 7
BEHAVE-AD and the BEHAVE-AD-FW symptomatic categories have been shown to respond both to nonpharmacological interventions, such as those associated with keeping subjects in the placebo arm of a clinical trial, and, to a greater extent, to pharmacotherapeutic interventions, such as those associated with antipsychotic and anxiolytic medications.

Assessment of Behavioral Disturbance Symptoms of AD with the BEHAVE-AD and the BEHAVE-AD-FW Independent of Cognitive and Functional Symptoms, and Mixing of Behavioral Disturbance Symptoms with Cognitive and Functional Symptoms with the NPI

Yet another major difference between the BEHAVE-AD and the BEHAVE-AD-FW assessments and the NPI measures is the separation of behavioral symptoms of AD from cognitive and functional symptoms of AD with the BEHAVE-AD and the BEHAVE-AD-FW, and the mixing of behavioral symptoms of AD with cognitive and functionally based symptoms with the NPI.

As discussed extensively in the course of this review, the BEHAVE-AD and the BEHAVE-AD-FW consist of symptoms that differ from cognitive and functional symptoms of AD in fundamental ways. These include: (1) the 7 BEHAVE-AD and BEHAVE-AD-FW symptomatic categories all respond to both nonspecific management-type interventions associated with the placebo arm of a clinical trial in the absence of effects on cognition; (2) the BEHAVE-AD and BEHAVE-AD-FW categories all respond further to pharmacological interventions, such as antipsychotic medication or anxiolytic medication, in the absence of effects on cognition; (3) unlike the cognitive and functional symptoms of AD, all symptoms in the BEHAVE-AD and the BEHAVE-AD-FW peak in occurrence at some stage prior to the final stage of AD [9], and (4) unlike the cognitive and functional symptoms of AD, which occur universally and progressively throughout the AD process, the 25 BEHAVE-AD and BEHAVE-AD-FW symptoms, albeit for the most part common, do not occur in all subjects with AD at any stage of the disease process.

In marked contrast, the NPI scale contains a symptomatic category, called ‘Apathy/Indifference’, which very strongly relates to the progressive cognitive and functional stages of AD. This includes questions such as ‘has the patient lost interest in the world around him/her?’, ‘is he/she more difficult to engage in conversation or doing chores?’, and ‘does the patient contribute less to household chores?’ Of course, these capacities will change inexorably and continuously with the advance of cognitive and functional decline in AD [59, 135, 136]. A medication which improves cognition and functioning in AD, such as a cholinesterase inhibitor or an NMDA receptor antagonist, would be expected to improve symptomatology on the Apathy/Indifference subscale of the NPI. However, such a medication should not be misinterpreted as improving behavioral disturbances in AD.

Summary of Differences between the NPI and the BEHAVE-AD and BEHAVE-AD-FW Measures

In summary, the BEHAVE-AD and the BEHAVE-AD-FW assessments differ from the NPI in very important ways. These include: (1) the individual assessment of symptoms with the BEHAVE-AD and the BEHAVE-AD-FW and the assessment of symptomatic categories with the NPI; (2) the assessment of symptoms of relevance for AD with the BEHAVE-AD and the BEHAVE-AD-FW and the assessment of symptoms of relevance for various disorders, including mania, hypomania, and affective disorders, as well as AD, with the NPI, and (3) the assessment of behavioral disturbance symptoms which differ from the cognitive and functional symptoms of AD with the BEHAVE-AD and BEHAVE-AD-FW assessments, in contrast to the NPI, which includes at least one symptomatic category which is primarily cognitively and functionally based. These differences between the BEHAVE-AD and the BEHAVE-AD-FW and the NPI assessment are summarized in table 6.
Table 6. Differences between the BEHAVE-AD [1] and the BEHAVE-AD-FW [2] and the NPI [29, 30, 133] in the context of the assessment of treatment of BPSD-related disturbances in AD*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Twenty-five, generally commonly occurring behavioral symptoms in AD, falling within 7 symptomatic categories are assessed. Each symptom is individually assessed on a 4-point severity scale for the BEHAVE-AD or on a 4-point severity scale and a 4-point frequency scale (with the exception of a 3-point frequency scale for 1 of the 25 symptomatic items) for the BEHAVE-AD-FW.</td>
<td></td>
<td>Surveys symptoms in 12 symptomatic categories. These include categories which are relevant for behavioral disturbances in AD and other categories which are not relevant for AD but applicable for mania or hypomania, affective disorder, or other conditions. Symptoms in the 12 categories are surveyed. Each category is scored on a 4-point frequency scale and a 3-point severity scale. Individual symptoms are not scored.</td>
</tr>
<tr>
<td>All 7 symptomatic categories contain symptoms which are commonly occurring at one or more dementia stages of AD. These range from a prevalence of &gt;20% for Hallucinatory Disturbances in mild dementia and moderately severe dementia (GDS stages 4 and 6 [7]) to &gt;60% prevalence at one or more dementia stages for the Paranoid and Delusional Ideation, Activity Disturbances, Aggressiveness, Affective Disturbances, and Anxieties and Phobias symptomatic categories [22]. Therefore, 5 of the 7 BEHAVE-AD symptomatic categories have a prevalence of &gt;60% at one or more of the 4 GDS dementia stages.</td>
<td></td>
<td>Surveys 12 symptomatic categories, some of which have no relevance for AD or have more relevance for medication side effects than for the symptomatology of AD. For example, the NPI symptomatic category ‘Elation/Euphoria’, a symptom of mania and related conditions, was found to occur in only 1 of 214 AD persons at all levels of severity in a community survey [31]. This was similar to the prevalence in nondemented persons in the community. Therefore, Elation/Euphoria is not a symptom of AD. Another example is the NPI symptomatic category ‘Appetite/Eating Disorders’, which includes symptoms such as increase in appetite, gaining weight, and poor appetite. These symptoms are not part of the behavioral disturbances of AD and are generally not part of AD in any context. Some of these symptoms occur in depressive disorder and they can occur as a result of side effects of putative AD therapies or be confused with the musculoskeletal wasting in the final GDS stage 7 of AD. Mixing symptoms which are not part of the behavioral pathologic syndrome of AD but which can occur as a result of medication side effects with symptoms which are behavioral disturbances can cloud the assessment of risk/benefit ratios of putative therapies for AD behavioral disturbances.</td>
</tr>
</tbody>
</table>

Separation of cognitive and functional changes in AD from the behavioral disturbances of AD versus mixing of cognitive and functional symptoms of AD with behavioral disturbances of AD

The purpose of a rating scale for the assessment of behavioral disturbances in AD is to assess those symptoms for treatment which may be applicable for these disturbances. These treatments include both pharmacological and nonpharmacological treatments. The BEHAVE-AD and the BEHAVE-AD-FW contain 7 symptomatic categories. All these symptomatic categories have demonstrated significant (p < 0.01) positive responses to the nonpharmacological effects of entry into the placebo arm of a clinical trial, and these effects appear to have been independent of effects on cognition [e.g., 25]. Furthermore, total BEHAVE-AD scores showed a significant additional response above that of the placebo condition to antipsychotic medication pharmacotherapy in this trial (p = 0.001). Positive responses were seen for 6 of the 7 BEHAVE-AD symptomatic categories. These effects were independent of cognition [25]. The one category of the BEHAVE-AD which did not show an additional positive response above that of the placebo condition to the pharmacotherapy in the antipsychotic medication trial of Katz et al. [25], i.e., Anxieties and Phobias, has shown a significant positive response to pharmacotherapy, above that of placebo, in an antianxiety medication trial [24]. In this study, once again, total BEHAVE-AD scores, BEHAVE-AD global scores, and each of the BEHAVE-AD symptomatic categories showed improvements on the nonspecific placebo intervention. An anxiolytic medication produced additional improvements above the placebo condition on 6 of the 7 BEHAVE-AD symptomatic categories. The one symptomatic category which showed no change in the Katz et al. [25] antipsychotic medication trial showed a significant added benefit, over and above that of the placebo condition, with the anxiolytic medication (buspirone) [24]. These effects in the Levy et al. [24] trial appear to have been independent of effects on cognition (i.e., no significant changes on the MMSE were seen). Therefore, all BEHAVE-AD symptomatic categories have been demonstrated to be potentially responsive to pharmacotherapies independent of the effects on cognition. These observations are compatible with the design of the BEHAVE-AD, which contains cognition-independent symptoms and symptomatic categories which, although related to AD, do not correlate with cognition or functional changes in AD. Another way of explaining the above statement is that all BEHAVE-AD and BEHAVE-AD-FW symptoms and symptomatic categories peak at some stage prior to the final GDS stage 7 of AD [7, 22]. Another qualitative difference of the BEHAVE-AD and BEHAVE-AD-FW symptomatic categories and the individual BEHAVE-AD and BEHAVE-AD-FW symptoms is that all these symptoms, and symptomatic categories, although generally commonly occurring at various stages, do not occur universally at any stage of AD. This prevalence is fundamentally different from cognitive and functional changes in AD, which become increasingly universal as the disease progresses.

The NPI contains a symptomatic category termed ‘Apathy/Indifference’, which, as defined, tracks very closely with the progressive cognitive and functional changes in AD. For example, as defined in the NPI, this category includes symptoms such as ‘is he/she more difficult to engage in... doing chores?’ Also, ‘does the patient contribute less to household chores?’ and ‘is it more difficult to keep a conversation going?’ These changes track very closely with the cognitive and functional progression of AD. For example, it has been demonstrated that each GDS stage of AD from MCI to mild dementia, to moderate dementia, is associated with progressive changes (losses) in self-care, household activities, general activities, food preparation, etc [136]. Additionally, these progressive AD cognition and functional stage-related changes have been demonstrated for conversation and social functioning [136]. The problem with including a subscale which is primarily cognitively and functionally based in a behavioral inventory is that when applied to medications with positive effects on cognition and functioning, these positive effects can be misinterpreted as positive effects on behavioral disturbances.

© 2013 Barry Reisberg, MD. All rights reserved.
Conclusion

Recent findings with respect to the value of investigating the specific delusions which occur in the course of AD [5] and the possibly characteristic brain changes associated with one of these delusions [6] have served as an impetus for this comprehensive review of and commentary on BPSD. The history of the development of the informant-based BEHAVE-AD [1] and the more detailed BEHAVE-AD-FW [2] has been discussed. We have also discussed the development of the direct subject observation-based E-BEHAVE-AD [12]. The symptoms assessed with these instruments differ fundamentally from cognition- and function-based symptoms of AD. For example, unlike the cognitive and functional symptoms of AD which occur universally and progressively with the advance of AD, the behavioral symptoms of AD are not universal in the course of AD and, in the context of the entire course of AD, including the frequently ignored final GDS stage 7 of AD, the behavioral symptoms described with the BEHAVE-AD, the BEHAVE-AD-FW, and the E-BEHAVE-AD are not progressive in the context of the entire cognitive, functional, and temporal course of AD, i.e., they all peak at some stage prior to the final GDS stage 7.

The symptoms assessed with the BEHAVE-AD, the BEHAVE-AD-FW, and the E-BEHAVE-AD have been shown, in the context of total scale scores, to respond to pharmacotherapy with psychotropic medications [25–28, 129]. For the BEHAVE-AD, these total scale responses have been studied in considerably greater detail. All 7 BEHAVE-AD symptomatic categories have shown significant improvements in response to the management-type effects associated with the entry into the placebo arm of a clinical trial (p < 0.01) [25]. Furthermore, all 7 BEHAVE-AD symptomatic categories have shown additional improvements in response to psychotropic pharmacotherapies [24, 25, 28]. Additionally, all of the 7 individual Paranoid and Delusional Ideation categorical symptoms assessed with the BEHAVE-AD have shown significant improvements in response to the antipsychotic medication risperidone in comparison with the placebo (management) treatment [25].

The BEHAVE-AD-FW assessment instrument has all of the elements of the BEHAVE-AD assessment plus a frequency-weighted dimension. It is estimated that the BEHAVE-AD-FW adds approximately 2.5 times the sensitivity to the BEHAVE-AD and hence can be very useful for pharmacological trials.

The direct subject evaluation-based E-BEHAVE-AD was developed not as a replacement for the informant-based BEHAVE-AD and BEHAVE-AD-FW assessments, but to provide an additional dimension of the systematic observation and evaluation of behavioral disturbances, i.e., from direct subject evaluation. This direct observation of behavior, lasting approximately 20 min, has been found to correlate significantly with the BEHAVE-AD, based on informant observation over the prior 2-week interval. However, most of the variance between informant reports with the BEHAVE-AD and the direct observation of symptoms with the E-BEHAVE-AD is non-overlapping. Hence, we believe these instruments assess, in large part, separate useful clinical and research information. For example, both the E-BEHAVE-AD and the BEHAVE-AD have been used as part of a comprehensive clinical trial assessment measure, the NYU-CIBIC-Plus, which endeavors to assess overall clinical improvements including behavior and other domains [13, 14]. In a recent study using the NYU-CIBIC-Plus and employing a management intervention, the informant-based BEHAVE-AD component showed significant large improvements, whereas the observer-based E-BEHAVE-AD scale component of the comprehensive NYU-CIBIC-Plus assessment showed no changes [106].

The treatment of behavioral disturbances assessed with the BEHAVE-AD, the BEHAVE-AD-FW, the E-BEHAVE-AD, and related measures is important in alleviating stress in the caregivers and other social contacts of AD patients as well as in postponing negative outcomes for patients with AD including institutionalization and demise. Nevertheless, the usage of antipsy-
Chotic medications for treating behavioral disturbances and investigations of antipsychotic medications for treating these behavioral disturbances have been severely constrained by findings of increased cerebrovascular adverse events and increased mortality with these medications. We have noted that these negative outcomes have likely been the result of the failure of clinical trials to follow the ‘start low, go slow’ adage [92]. This failure to follow the ‘start low, go slow’ recommendation also extends to clinical settings as well as clinical trials. For example, we have recommended previously and also in this publication typical starting doses for risperidone of 0.25 mg or half of a 0.25-mg risperidone tablet and subsequent dosage adjustments after weeks on the low starting dosage [92]. The dosages given in clinical trials and the titration periods for antipsychotic usage have been in marked variance with these recommendations. Specifically, doses have been much higher and titration periods much shorter than our recommendations. Hopefully, these errors will be rectified in future antipsychotic medication studies, which might then result not only in improvements in behavioral disturbance assessment measures, but also in an absence of increased morbidity and mortality.

Finally, in recent years, the NPI has been widely used in research for the treatment of behavioral symptoms of AD and other dementias. Major differences between the BEHAVE-AD and the BEHAVE-AD-FW and the NPI have been reviewed herein. These include the following: (1) The BEHAVE-AD and the BEHAVE-AD-FW assess symptom categories which are of relevance for AD-related behavioral and psychological symptoms and associated disturbances and which occur in AD. In contrast, the NPI assesses some symptomatic categories and symptoms, such as the Elation/Euphoria symptomatic category, which do not occur in AD [31]. (2) The BEHAVE-AD and the BEHAVE-AD-FW assess individual behavioral symptoms which are potentially remediable in the course of AD. For example the BEHAVE-AD and the BEHAVE-AD-FW assess 7 individual Paranoid and Delusional Ideation symptoms, each of which has shown significant improvements with antipsychotic medication treatment [25]. Each symptom is assessed with severity and frequency measures with the BEHAVE-AD-FW assessment instrument. The NPI surveys symptoms categorically but not individually. For example, in the NPI category termed ‘Delusions’, the NPI does not assess individual symptoms – only the Delusions category is assessed on severity and frequency measures. (3) The BEHAVE-AD and the BEHAVE-AD-FW assess symptoms and symptomatic categories which are of relevance for the behavioral disturbances of AD and which occur in AD and which do not relate primarily to possible side effects of medications used to treat behavioral disturbances in AD. The NPI includes symptoms, such as weight gain and increase in appetite, which do not occur as a result of AD and which are likely to be the result of medication side effects when they do occur in AD. (4) The BEHAVE-AD and the BEHAVE-AD-FW assess symptomatic categories which are responsive to treatment, for example to treatment with antipsychotic medications or with anxiolytic medication, and which are independent of cognition, as demonstrated in clinical trials [25, 24]. The NPI includes a symptom category, ‘Apathy/Indifference’, which as defined in the NPI, for example, by less contribution to household chores and by the patient being less conversational, is primarily cognitive [59, 135, 136]. This is problematic because, in the context of a behavioral disturbance scale, it can result in confusion between the cognitive effects of a potentially useful medication and a medication which is effective for behavioral disturbances.

In summary, progress is continuously being made in the understanding, assessment, and treatment of behavioral disturbances in AD. We have discussed selected aspects of this continuing progress in this historical review, perspective and commentary. Our goal is that this selective review will assist clinicians and scientists in achieving further gains in this very important aspect of the clinical treatment of AD and its antecedents.
Acknowledgements

This perspective, commentary, and historical review and the studies described was supported in part by United States Department of Health and Human Services (DHHS) grants P30 AG08051, AG03051, AG09127, and AG11505, from the National Institute on Aging and by grant MH43486 from the National Institute of Mental Health of the U.S. National Institutes of Health, by grants 90AZ2791, 90AM2552, and 90AR2160 from the U.S. DHSS Administration on Aging, by a Senior Clinical Research Fellowship provided by Forest Laboratories, Inc., by the Fisher Center for Alzheimer’s Research Foundation, and by grants from Mr. Zachary Fisher and Mrs. Elizabeth M. Fisher, Mr. William Silverstein, and Mr. Leonard Litwin. Additional support is acknowledged from Mr. Michael Stern, The Estate of Rosalind Cherry, The Hagedorn Fund, The Louis J. Kay Foundation, The Stringer Foundation, and The Forest Research Institute.

Disclosure Statement

Dr. Barry Reisberg is the developer and copyright holder, individually or with others, of many of the assessment instruments discussed in this publication. These include the following: The Clinical Rating Scale for Symptoms of Psychosis in Alzheimer’s Disease (SPAD), the Brief Cognitive Rating Scale (BCRS), the Functional Assessment Staging (FAST) procedure, the Behavioral Pathology in Alzheimer’s Disease Rating Scale (BEHAVE-AD) measure, the Empirical BEHAVE-AD Rating Scale (E-BEHAVE-AD), the Behavioral Pathology in Alzheimer’s Disease Frequency Weighted Severity Scale (BEHAVE-AD-FW), the Activities of Daily Living International Scale (ADL-IS), and the Global Deterioration Scale (GDS). Dr. Stefanie Auer is a co-developer and co-copyright holder of the Empirical BEHAVE-AD Rating Scale (E-BEHAVE-AD).

References


104 Janssen Pharmaceuticals: Personal communication to B. Reisberg, circa 2004 as follows: the Behavioral Pathology in Alzheimer’s Disease [BEHAVE-AD] assessment has been used in pivotal trials (generally as the pivotal trial primary outcome measure) for the approval of medication for the treatment of Behavioral and Psychological Symptoms of Dementia (BPSD) in worldwide studies which have resulted in the approval of medication to treat these symptoms, risperidone [Risperdal], in more than 30 nations, including Australia, Canada, Finland, Germany, Ireland, The Netherlands, Norway, Portugal, Spain, Sweden, and Switzerland.


