Published in final edited form as:

Am J Geriatr Psychiatry. 2013 January ; 21(1): 78–87. doi:10.1016/j.jagp.2012.10.013.

A Comparison of the E-BEHAVE-AD, NBRS and NPI in Quantifying Clinical Improvement in the Treatment of Agitation and Psychosis Associated with Dementia

Zahinoor Ismail, BMs MD FRCPC^{1,2,3}, Chetachi A Emeremni, MA⁴, Patricia R. Houck, MSH⁴, Sati Mazumdar, PhD⁴, Jules Rosen, MD⁵, Tarek K. Rajji, MD^{1,2}, Bruce G. Pollock, MD, PhD, FRCPC^{1,2,5}, and Benoit H. Mulsant, MD, MS, FRCPC^{1,2,5,*}

¹Centre for Addiction and Mental Health, Geriatric Mental Health Program, Toronto, Ontario, Canada

²Department of Psychiatry, University of Toronto, Toronto Ontario, Canada

³Department of Psychiatry, University of Calgary, Calgary Alberta, Canada

⁴Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, USA

⁵Department of Psychiatry, University of Pittsburgh, Pittsburgh, USA

Abstract

Objectives—The aim of this study is to compare the Empirical Behavioral Rating Scale (E-BEHAVE AD), Neurobehavioral Rating Scale (NBRS), and Neuropsychiatric Interview (NPI) in detecting behavioral disturbance and psychotic symptoms in dementia and characterizing changes in response to treatment.

Design—87 subjects in the randomized controlled trial “Continuation Pharmacotherapy for Agitation of Dementia” (CPAD) were included in this analysis. We compared the detection and changes over 12 weeks of both agitation and psychosis using these 3 instruments. A receiver operating characteristic (ROC) analysis was performed to compare the performance of the three instruments in detecting global improvement.

Results—The instruments were equally likely to detect agitation. The NBRS was most likely to detect psychosis. While the NPI best detected improvement in agitation, the instruments were equal for detecting improvement in psychosis. 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 symptom reduction. The E-BEHAVE-AD performed best at a cutpoint of 30% target symptom reduction and the NBRS and NPI both performed best at 50%.

Conclusion—The E-BEHAVE-AD, NBRS, and NPI were more similar than different in characterizing symptoms but differed in detecting response to treatment. Differences in sensitivity and specificity may lead clinicians to prefer a specific instrument, depending on their goal and the expected magnitude of response to any specific intervention.

*Corresponding author: Geriatric Mental Health Program, Centre for Addiction and Mental Health, 1001 Queen Street West, Toronto, Ontario, Canada M6J 1H4. Phone: +1 416 535 8501x4749. Fax: +1 416 583 1307. Benoit_Mulsant@camh.net.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Alzheimer's Disease; pharmacotherapy; clinical trials; BPSD; NPS; neuropsychiatric symptoms; dementia; agitation; psychosis; NPI; NBRs; E-BEHAVE-AD; rating scales; Behavioural Disturbance in Alzheimer's Disease (BDAD)

Introduction

Behavioral and psychological symptoms of dementia are common and result in significant morbidity. These neuropsychiatric symptoms have been the target of many intervention studies, both pharmacological and non-pharmacological (1–5). Interventions have often been found to be either ineffective or to have a modest effect (6–8). However, there is no gold standard in assessing for the presence of these symptoms or their response to intervention, and many instruments are used (9). The aim of this study is to compare the Empirical Behavioral Rating Scale (E-BEHAVE-AD), Neurobehavioral Rating Scale (NBRs), and Neuropsychiatric Interview (NPI) in detecting neuropsychiatric symptoms and characterizing changes in response to treatment utilizing outcome data from a published randomized clinical trial, The Continuation Pharmacotherapy for Agitation of Dementia (CPAD) study (2, 10).

Methods

The CPAD study is a 12-week randomized double blind trial that compared the effectiveness of citalopram and risperidone in patients suffering from agitation or psychotic symptoms in dementia. Subjects were consecutively recruited on an inpatient unit and were eligible if they had Alzheimer's dementia (AD), vascular dementia, dementia with Lewy bodies (DLB), mixed dementia or dementia not otherwise specified. Target symptoms had to be of moderate or higher severity as evidenced by the need for hospitalization and a rating of 3 or higher (moderate to severe) on at least one of the agitation items (aggression, agitation, hostility) or psychosis items (suspiciousness, hallucinations, or delusions) of the NBRs. Patients were excluded if they had a current or past diagnosis of schizophrenia, schizoaffective disorder, delusional disorder, psychotic disorder not otherwise specified, bipolar disorder, mental retardation, cognitive deficits following head trauma, or a current diagnosis of delirium, substance-induced persisting dementia, Parkinson disease, drug/alcohol abuse, or dependence. Other exclusion criteria were a major depressive episode within the past 6 months or clinically significant depressive symptoms with a rating of 12 or higher on the Cornell Scale for Depression in Dementia (CSDD) (11); unstable physical illness; abnormal laboratory findings; treatment with a depot neuroleptic drug within 2 months of screening or fluoxetine within 4 weeks of screening; or a history of allergy or intolerance to citalopram or risperidone. CPAD reported that changes in the primary outcome measures—the NBRs agitation or psychosis scores—did not differ significantly between the two medication groups. Within the medication groups, the decrease in the NBRs agitation score was significant with citalopram (–12.5%) but not risperidone (–8.2%). The decreases in NBRs psychosis score were significant both with citalopram (–32.3%) and risperidone (–35.2%). CPAD also reported similar results with the NPI as a secondary outcome measure. (2).

The two aims of this analysis were to compare the ability of the E-BEHAVE-AD, NBRs, and NPI to detect behavioral disturbance and psychotic symptoms and to characterize changes of these symptoms in response to treatment. In most trials of behavioral disturbances in dementia, a total score cut-off determines eligibility, and outcomes are typically based on relative or absolute score changes (12, 13). This approach has been

criticized because small score change can be statistically but not clinically significant. Comparing effect sizes addresses in part this problem. However, the measure that may be the most relevant to the patients, their families and professional caregivers is whether or not the symptoms that required an intervention (“target symptoms”) have resolved or not. Thus, we focused our analysis on the detection of target symptoms at baseline and their resolution associated with treatment. In the absence of generally accepted methods, for each instrument, we used the following approach to define the presence of target symptoms and their resolution:

E-BEHAVE-AD

The E-BEHAVE-AD is a reliable and valid clinician interview rated 12-item instrument developed to assess behavioral pathology in Alzheimer’s disease and related dementia (14). Analogous to the caregiver rated Behavior Pathology in Alzheimer’s Disease (BEHAVE-AD) (15), the empirically rated E-BEHAVE-AD rates 12 symptoms as 0 (not present), 1 (vague, not clearly observable), 2 (clearly defined) and 3 (severely present) resulting in a total scores of 0–36.

Based on a factor analysis of the BEHAVE-AD (16) and approaches validated for the NBRS (see below and (17, 18)), we identified four E-BEHAVE-AD target symptoms: paranoid/delusional ideation, hallucinations, activity disturbance, and aggressivity. A target symptom was considered to be present if its item score was 2 or 3 at baseline and to have resolved if the score decreased to 1 or 0.

NBRS

The NBRS is a multidimensional 27-item observer-rated instrument that measures behavioral disturbances commonly seen in dementia (19) and has been validated for content validity in Alzheimer’s and vascular dementias (20, 21). Based on the Brief Psychiatric Rating Scale (22), symptoms are rated from 0 to 6 where 0 indicates the absence of a symptom, 1 a very mild, 2 a mild, 3 a moderate, 4 a moderately-severe, 5 a severe and 6 an extremely severe symptom.

Based on previous studies that have shown that this approach can differentiate changes associated with active pharmacotherapy vs. placebo (3, 17, 18), we identified six NBRS target symptoms: aggression, agitation, hostility, delusions, hallucinations and suspiciousness. A target symptom was considered to be present if its item score was 3 or higher at baseline and to have resolved if the score decreased to 2 or less.

NPI

The NPI is a short, informant rating instrument developed to assess neuropsychiatric psychopathology in dementia (23). The NPI-12 is a 12-item instrument assessing delusions, hallucinations, agitation/aggression, dysphoria/depression, anxiety, apathy, irritability, euphoria, disinhibition, aberrant motor behavior, nighttime behavior disturbance and appetite and eating abnormalities (24). Each of the 12 items is scored based on the product of its frequency (1–4; occasionally, often, frequently and very frequently) by its severity (1–3; mild, moderate and marked) resulting in total scores up to 144.

In keeping with findings from NPI factor analyses (25–32) and similar to the approach with the E-BEHAVE-AD and NBRS, we identified six NPI target symptoms: delusions, hallucinations, agitation/aggression, disinhibition, irritability and aberrant motor behavior. A target symptom was considered to be present if its item score was 4 or higher at baseline and to have resolved if the score decreased to 3 or less.

Clinical Global Impression of Severity (CGI-S)

The CGI-S is a one-item observer-rated instrument that measures the severity of the patient's illness at the time of assessment, relative to the clinician's past experience with patients who have the same diagnosis (33). Scores range from 1 to 7 with 1 corresponding to being normal, not at all ill; 2, borderline mentally ill; 3, mildly ill; 4, moderately ill; 5, markedly ill; 6, severely ill; and 7, extremely ill. For this analysis, clinical improvement was defined as a reduction in score of at least 1 point and a final score of less than 4.

Data Analysis

In the CPAD study 103 subjects were randomized, and administered the CGI-S, E-BEHAVE-AD, NBRs, and NPI, at baseline and completion. All randomized subjects had at least one NBRs target symptom at baseline since inclusion criteria included the presence of a target symptom of moderate or higher severity on the NBRs. However, 15 subjects did not have an E-BEHAVE-AD or NPI target symptom at baseline; they were excluded as we could not determine clinical improvement with these two instruments in this group. One additional subject was excluded as he received only one dose of risperidone and his treatment response could not be determined. Thus 87 subjects were included in this analysis: 81 met diagnostic criteria for a probable ($N = 68$) or possible ($N = 13$) Alzheimer's dementia; 27 met criteria for probable ($N = 8$) or possible ($N = 19$) DLB (with 21 meeting criteria for both possible Alzheimer's dementia and possible DLB). Table 1 presents the demographic and clinical characteristics of these 87 subjects.

First, the presence of target symptoms was tabulated for all 3 instruments. Then, to compare the three instruments, since they did not include the same target symptoms, we aggregated target symptoms into an agitation and a psychosis syndrome for each instrument. For the E-BEHAVE-AD, the agitation syndrome comprises the activity disturbance and aggressivity items while the psychosis syndrome comprises the paranoid/delusional ideation and the hallucinations items. For the NBRs, the agitation syndrome comprises the aggression, agitation and hostility items, while the psychosis syndrome comprises the delusions, hallucinations, and suspiciousness items. For the NPI, the agitation syndrome comprises the agitation/aggression, disinhibition, irritability, and aberrant motor behavior while the psychosis syndrome comprises the delusions and hallucinations items.

Presence of agitation or psychosis (defined as the presence of at least one target symptom) was then determined for each instrument. The significance of the differences among the correlated proportions was assessed using the Cochran Q statistics which takes into account the correlation arising from studying the same subjects under different conditions (34, 35). Resolution of each individual target symptom and clinical improvement in agitation or psychosis (defined as the resolution of at least one target agitation or psychosis symptom) was then assessed for each instrument. Overall clinical improvement (defined as the resolution of at least one target symptom) was assessed for each instrument and compared with clinical improvement as determined with the CGI-S using a McNemar χ^2 statistics.

Finally, we classified the presence or absence of clinical improvement according to the CGI-S (i.e., a reduction in score of at least 1 point and a final score of less than 4) and used this classification as the gold standard in the ROC analyses of the E-BEHAVE-AD, NBRs, and NPI. ROC curves were generated comparing the presence or absence of clinical improvement with the proportion of symptoms resolved for each instrument. For example, a cut-off of 50% means that a subject is considered clinically improved if he or she has experienced a 50% or more reduction in the total number of his or her target symptoms (e.g., resolution of one target symptom if the subject presented with one target symptom; resolution of 3 if he or she presented with 5). For each instrument and for each of these cut-

offs, we plot the proportion of subjects correctly classified as improved (sensitivity, Y axis) and correctly classified as not improved (specificity, X axis) according to the classification based on the CGI-S described above. For instance, with a cut-off of 100% reduction in target symptom, the sensitivity is 0% (no subject who has improved according to the CGI-S is correctly classified as having improved by the instrument) and the specificity is 100% (all subjects classified as not having improved by the CGI-S are correctly classified as not having improved by the instruments). The areas under the correlated ROC curves were compared using the Mann-Whitney Statistics of the non-parametric method of Delong, Delong and Clarke (36). Specificity and sensitivity in detecting clinical improvement according to the CGI-S was determined for all three instruments at the following cutpoints for the proportion of target symptoms resolved: 20%, 30%, 40%, 50%, 60%, 70% and 80%.

Results

The number of subjects presenting with specific target symptoms and syndromes as well as rates of clinical improvement on the E-BEHAVE-AD, NBRs, or NPI are presented in Table 2. All 87 (100.0%) subjects were classified as having the agitation syndrome with the E-BEHAVE-AD, as compared to 85/87 (97.7%) with the NBRs, and 83/87 (95.4%) with the NPI (Cochran's $Q=4.0$, $df=2$, $p=0.14$). The E-BEHAVE-AD classified 58/87 (66.7%) subjects as having the psychotic syndrome, compared to 62/87 (71.3%) with the NBRs, and 46/87 (52.9%) with the NPI (Cochran's $Q=20.8$, $df=2$, $p<0.0001$). Rates of improvement were statistically different between instruments for agitation (E-BEHAVE-AD: 24/87, 27.6%; NBRs: 42/85, 49.4%; NPI: 52/83, 62.7%; Cochran's $Q=33.84$, $df=2$, $p<0.0001$, $n=81$) but not for psychosis (E-BEHAVE-AD: 30/58, 51.7%; NBRs: 34/62, 54.8%; NPI: 23/46, 50.0%; Cochran's $Q=1.85$, $df=2$, $p=0.40$, $n=43$).

Figure 1 presents the rates of overall clinical improvement for all four instruments. These rates were significantly different (CGI-S: 28/87; 32%; E-BEHAVE-AD: 41/87, 47%; NBRs: 57/87, 66%; NPI: 58/87, 67%; Cochran's $Q=45.5$, $df=3$, $p<0.0001$).

The ROC curves generated for the E-BEHAVE-AD, NBRs, and NPI are presented in Figure 2. Areas under the curve (AUC, SE) for the E-BEHAVE-AD, NBRs and NPI were 0.822 (0.050), 0.900 (0.037) and 0.839 (0.044) respectively. Paired comparisons of the correlated AUCs showed no statistical differences between them (E-BEHAVE-AD vs. NBRs, $\chi^2=2.06$; $df=1$; $p=0.15$; E-BEHAVE-AD vs. NPI, $\chi^2=0.08$; $df=1$; $p=0.78$; NBRs vs. NPI, $\chi^2=1.33$; $df=1$; $p=0.25$).

Sensitivity and specificity for each instrument at the various cut-off points for the proportion of symptoms resolved are presented in Figure 3. Optimal cut-off points that maximize sensitivity and specificity for each instrument are as follows: the optimal relative reduction in target symptoms is 30.0% for the E-BEHAVE AD (sensitivity 0.79, specificity 0.73), 50.0% for the NBRs, (sensitivity 0.89, specificity 0.85), and 50.0% for the NPI score (sensitivity 0.86, specificity 0.76).

Discussion

Given the variability in presentation of neuropsychiatric symptoms in dementia, it is important to consistently characterize and monitor these symptoms. Only then can treatment responses truly be quantified. Numerous rating instruments have been developed in order to do this but they differ markedly. For instance, they cover different symptoms and scores are based on frequency or severity or a combination of the two. Scoring based on frequency may be subject to recall bias and may be less sensitive than scoring based on magnitude or severity. However, although severity may be of greater clinical relevance, it is harder to

quantify. Moreover, instruments elicit information in different ways. This analysis aimed at: (1) determining how three instruments that have been used in major clinical trials --the E-BEHAVE-AD, NBRs and NPI-- compared in capturing neuropsychiatric phenomenology; and (2) comparing the performance of these instruments in characterizing treatment response (clinical improvement) versus each other and the CGI-S. CPAD reported a percentage score reduction for both agitation and psychosis syndromes for both treatment arms. The relevance of this study is that the ROC curves generated in this secondary analysis give context to the score reductions in CPAD and compare them to the CGI-S, which captures global clinical improvement.

Of the 87 subjects included in this analysis, agitation was detected by the E-BEHAVE-AD in all, while the NBRs and NPI detected agitation in 85 and 83 subjects respectively. There was more variability in capturing psychosis with the E-BEHAVE-AD, NBRs and NPI detecting psychosis in 58, 62 and 46 subjects respectively. Presence of psychosis or agitation as per the NBRs was an inclusion criteria in the CPAD study and this may explain why more subjects had psychosis according to the NBRs compared to the E-BEHAVE-AD or NPI. However, the discrepancy may be a result of the difficulty in identifying delusions in the context of dementia. Misplacing objects and then assuming that they were stolen can present as suspiciousness or as forgetfulness/confabulation, the former being classified as delusional, the later as not delusional. The NBRs includes three psychotic items --delusions, hallucinations and suspiciousness-- while the E-BEHAVE-AD and NPI include only delusions and hallucinations. The suspiciousness item may account for the increased sensitivity of the NBRs in detecting psychosis. Furthermore, while caregivers or staff can often overtly detect agitation, psychosis can have a more covert presentation and be more difficult to detect on a consistent basis.

Instruments did not perform equally in detecting clinical improvement for agitation. The NPI detected improvement in 63% of subjects, the NBRs in 49%, and the E-BEHAVE-AD in 28%. The NPI includes four agitation items (agitation, disinhibition, irritability and aberrant motor behavior), the NBRs includes three (agitation, aggression and hostility), and the E-BEHAVE-AD includes only two (aggression and activity disturbance). While the number of items may correlate with increased sensitivity to treatment response, the items are broken down into component behaviors probed by specific questions. The E-BEHAVE-AD activity disturbance item covers pacing and wandering, repetitive activities, and inappropriate activities; its aggression item covers verbal outbursts, physical threats and violence or other behaviors indicating agitation. Severity is scored from 0–4 (absent to severe) on the E-BEHAVE-AD while the NBRs is scored from 0–6 and the NPI is scored on the product of frequency (1–4) and severity (1–3) resulting in a 12-point scale (however, it is not possible to score 5, 7, 10 or 11 on this instrument). These different rating schemes may also result in different sensitivity when measuring treatment response: the scoring systems rather than the numbers of items may contribute to differences in detecting change in agitation. The three instruments detected improvement in psychosis in comparable proportions of subjects (NBRs: 55%; E-BEHAVE-AD: 52%; NPI: 50%). The NBRs includes three psychosis items (hallucinations, delusions, and suspiciousness), while the E-BEHAVE-AD includes two items. However, these two items are queried based on 16 questions: the questions about delusions probe more persecutory than misidentification delusions and hallucinations are queried across all sensory modalities. The NPI delusion item is rated based on both persecutory and misidentification delusions (using 9 questions) and the hallucination item is rated across all sensory modalities using 7 questions. Despite the more detailed probing of the E-BEHAVE-AD and NPI, the inclusion of a suspiciousness item in the NBRs may account for its comparability to the E-BEHAVE-AD and the NPI.

In the absence of a true “gold standard” we used the CGI-S as our measure to compare the 3 instruments and assess their ability to quantify symptoms and response to treatment. In clinical practice, gross detection of abnormal and troublesome behavior is generally adequate for a clinician to consider an intervention for that behavior. Similarly, an overall improvement in that behavior is adequate in clinical practice to consider the intervention successful. The use of the CGI-S captures qualitatively these clinically meaningful changes and reflects the “gestalt” approach widely used by clinicians. The AUCs in the ROC analysis did not differ significantly among the 3 instruments. However, the shape of the curves describe differences in utility and performance: the NPI appeared more likely to detect clinical improvement and less likely to detect absence of clinical improvement (i.e., more sensitive and less specific). Conversely, the E-BEHAVE-AD appears less likely to detect clinical improvement and more likely to detect absence of clinical improvement (i.e., less sensitive and more specific). The NBRs appears to perform “in the middle”. Similarly, cut-off points for the proportion of target symptoms resolved that maximize sensitivity and specificity are quite different for the three instruments attesting to differences in sensitivity and specificity. For the E-BEHAVE-AD, a 30% reduction in the number of target symptoms was the optimal cutoff, with a sensitivity of 0.79 and specificity of 0.73. For the NBRs, a 50% reduction in the number of target symptoms yielded a sensitivity of 0.89 and specificity of 0.85. Finally, for the NPI a 50% reduction in the number of target symptoms yielded a sensitivity of 0.86 and specificity of 0.76. Again, in the context of the excluded subjects, these differences reflect the different structures of the instruments discussed above (i.e., number of items corresponding to target symptoms and scoring schemes). They also highlight how one of the three instruments could be selected in clinical practice or for an RCT depending on the goal. For instance, one would favor the NPI when assessing the outcome of a low-risk intervention (i.e., in a situation when false positives are less problematic than false negatives and thus one would favor sensitivity over specificity). Conversely, one would favor the E-BEHAVE-AD when assessing the outcome of a high-risk intervention.

There are several limitations of this study to consider. First, inclusion into CPAD was based on the NBRs criteria. Of the 103 subjects enrolled in CPAD based on these NBRs criteria, 15 subjects did not have target symptoms on the E-BEHAVE-AD or the NPI and they had to be excluded from the analysis because treatment response could not be determined with these instruments. While this may suggest the NBRs better detects behavioral disturbance in dementia, we do not know whether there were patients who would have had target symptoms with the E-BEHAVE-AD or the NPI and not the NBRs nor do we know if these differences reflect acute fluctuations in target symptoms. However, this limitation is also a strength: the exclusion of the 15 subjects resulted in a sample who presented with baseline symptoms on all three instruments, received the same intervention, and served as their own control when comparing the ability of the 3 instruments to detect changes and also in comparison to a clinical global impression. Also, most of our subjects had moderate to severe dementia (i.e., they were at the stage when severe behavioral or psychotic symptoms are most likely to be present). Thus, our results may not be generalizable to patients with mild dementia. Another limitation to this study is the inherent difficulty in interpreting differences among the scores given the instrument differences with respect to domains, number of symptoms, specific probing questions for symptoms of agitation or psychosis and the scoring schemes. However, we believe that highlighting these differences provides valuable insight into the use of these instruments for clinical and research purposes. As there is no gold-standard for assessing treatment response, the ROC analysis was based on the CGI-S. One could argue that CGI-S ratings are impressionistic and are more subject to bias than the three instruments we were comparing. Furthermore, if a subject presented with several target symptoms and only one resolved, he or she would be classified as having improved with the three instruments but not necessarily with the CGI-S because the severity

of his symptoms may not be rated as “mild” (or better) on the CGI-S. As a global qualitative measure, the CGI-S captures the clinical relevance of symptoms since intrusive and bothersome symptoms of psychosis or agitation would be reflected by worse scores on this instrument. Similarly, if a score reduction on a specific instrument doesn’t correspond to a global change captured by the CGI, then regardless of the statistical significance or the effect size of that score change, this score change may not reflect clinical improvement. One could argue whether overall clinical status or number of target symptoms is more meaningful clinically when assessing the outcome of an intervention. Still, judging improvement based on global clinical status may lead to reject as inefficacious an intervention that works only for a specific symptom (e.g., a medication that treats delusions but not agitation and hallucinations). Given our current state of knowledge the use of complementary approaches when determining clinical response to interventions for neuropsychiatric disturbances in dementia appears prudent.

Conclusion

Overall, the E-BEHAVE-D, NBRS and NPI performed satisfactorily when detecting agitation and psychosis associated with dementia and assessing their improvement in response to a pharmacologic intervention. However, there were some statistically significant differences between instruments that may be clinically relevant. The NPI was the most likely to classify agitation as having improved, followed by the NBRS and the E-BEHAVE-AD. Instruments were equally likely to detect improvement in psychosis. The proportion of target symptoms having resolved that optimized the assessment of global clinical improvement also differed among the three instruments: 30% for the E-BEHAVE-AD, 50% for the NBRS, and 50% for the NPI. Thus, the nature of the intervention being assessed and the a-priori expected outcome should lead to the selection of instruments. Given their complementary nature, it is prudent to use more than one instrument and the CGI-S when possible.

References

1. Ballard C, Brown R, Fossey J, et al. Brief psychosocial therapy for the treatment of agitation in Alzheimer disease (the CALM-AD trial). *Am J Geriatr Psychiatry*. 2009; 17:726–733. [PubMed: 19700946]
2. Pollock BG, Mulsant BH, Rosen J, et al. A double-blind comparison of citalopram and risperidone for the treatment of behavioral and psychotic symptoms associated with dementia. *Am J Geriatr Psychiatry*. 2007; 15:942–952. [PubMed: 17846102]
3. Pollock BG, Mulsant BH, Rosen J, et al. Comparison of citalopram, perphenazine, and placebo for the acute treatment of psychosis and behavioral disturbances in hospitalized, demented patients. *The American journal of psychiatry*. 2002; 159:460–465. [PubMed: 11870012]
4. Ujkaj M, Davidoff DA, Seiner SJ, et al. Safety and Efficacy of Electroconvulsive Therapy for the Treatment of Agitation and Aggression in Patients with Dementia. *Am J Geriatr Psychiatry*. 2011 (epub ahead of print).
5. Wang LY, Shofer JB, Rohde K, et al. Prazosin for the treatment of behavioral symptoms in patients with Alzheimer disease with agitation and aggression. *Am J Geriatr Psychiatry*. 2009; 17:744–751. [PubMed: 19700947]
6. Ballard CG, Gauthier S, Cummings JL, et al. Management of agitation and aggression associated with Alzheimer disease. *Nature reviews*. 2009; 5:245–255.
7. Gauthier S, Cummings J, Ballard C, et al. Management of behavioral problems in Alzheimer's disease. *International psychogeriatrics / IPA*. 2010; 22:346–372. [PubMed: 20096151]
8. Schneider LS, Tariot PN, Dagerman KS, et al. Effectiveness of atypical antipsychotic drugs in patients with Alzheimer's disease. *The New England journal of medicine*. 2006; 355:1525–1538. [PubMed: 17035647]

9. Jeon Y, Sansoni J, Low L, et al. Recommended Measures for the Assessment of Behavioral Disturbances Associated With Dementia. *Am J Geriatr Psychiatry*. 2010 (epub ahead of print).
10. Culo S, Mulsant BH, Rosen J, et al. Treating Neuropsychiatric Symptoms in Dementia With Lewy Bodies: A Randomized Controlled-trial. *Alzheimer Dis Assoc Disord*. 2010; 24:360–364.
11. Alexopoulos GS, Abrams RC, Young RC, et al. Cornell Scale for Depression in Dementia. *Biological psychiatry*. 1988; 23:271–284. [PubMed: 3337862]
12. Lee PE, Gill SS, Freedman M, et al. Atypical antipsychotic drugs in the treatment of behavioural and psychological symptoms of dementia: systematic review. *BMJ (Clinical research ed)*. 2004; 329:75.
13. Rodda J, Morgan S, Walker Z. Are cholinesterase inhibitors effective in the management of the behavioral and psychological symptoms of dementia in Alzheimer's disease? A systematic review of randomized, placebo-controlled trials of donepezil, rivastigmine and galantamine. *International psychogeriatrics / IPA*. 2009; 21:813–824. [PubMed: 19538824]
14. Auer SR, Monteiro IM, Reisberg B. The Empirical Behavioral Pathology in Alzheimer's Disease (E-BEHAVE-AD) Rating Scale. *International psychogeriatrics / IPA*. 1996; 8:247–266. [PubMed: 8994895]
15. Reisberg B, Auer SR, Monteiro IM. Behavioral pathology in Alzheimer's disease (BEHAVE-AD) rating scale. *International psychogeriatrics / IPA*. 1996; (8 Suppl 3):301–308. discussion 351–304. [PubMed: 9154579]
16. Harwood DG, Ownby RL, Barker WW, et al. The behavioral pathology in Alzheimer's Disease Scale (BEHAVE-AD): factor structure among community-dwelling Alzheimer's disease patients. *International journal of geriatric psychiatry*. 1998; 13:793–800. [PubMed: 9850876]
17. Mulsant BH, Mazumdar S, Pollock BG, et al. Methodological issues in characterizing treatment response in demented patients with behavioral disturbances. *International journal of geriatric psychiatry*. 1997; 12:537–547. [PubMed: 9193962]
18. Kastango KB, Kim Y, Dew MA, et al. Verification of scale sub-domains in elderly patients with dementia: a confirmatory factor-analytic approach. *Am J Geriatr Psychiatry*. 2002; 10:706–714. [PubMed: 12427579]
19. Levin HS, High WM, Goethe KE, et al. The neurobehavioural rating scale: assessment of the behavioural sequelae of head injury by the clinician. *Journal of neurology, neurosurgery, and psychiatry*. 1987; 50:183–193.
20. Sultzer DL, Levin HS, Mahler ME, et al. Assessment of cognitive, psychiatric, and behavioral disturbances in patients with dementia: the Neurobehavioral Rating Scale. *Journal of the American Geriatrics Society*. 1992; 40:549–555. [PubMed: 1587970]
21. Sultzer DL, Levin HS, Mahler ME, et al. A comparison of psychiatric symptoms in vascular dementia and Alzheimer's disease. *The American journal of psychiatry*. 1993; 150:1806–1812. [PubMed: 8238634]
22. Overall JE, Gorham DR. The brief psychiatric rating scale. *Psychological Reports*. 1962; 62:799–812.
23. Cummings JL, Mega M, Gray K, et al. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. *Neurology*. 1994; 44:2308–2314. [PubMed: 7991117]
24. Cummings JL. The Neuropsychiatric Inventory: assessing psychopathology in dementia patients. *Neurology*. 1997; 48:S10–S16. [PubMed: 9153155]
25. Aalten P, de Vugt ME, Lousberg R, et al. Behavioral problems in dementia: a factor analysis of the neuropsychiatric inventory. *Dementia and geriatric cognitive disorders*. 2003; 15:99–105. [PubMed: 12566599]
26. Frisoni GB, Rozzini L, Gozzetti A, et al. Behavioral syndromes in Alzheimer's disease: description and correlates. *Dementia and geriatric cognitive disorders*. 1999; 10:130–138. [PubMed: 10026387]
27. Fuh JL, Liu CK, Mega MS, et al. Behavioral disorders and caregivers' reaction in Taiwanese patients with Alzheimer's disease. *International psychogeriatrics / IPA*. 2001; 13:121–128. [PubMed: 11352329]
28. Gauthier S, Wirth Y, Mobius HJ. Effects of memantine on behavioural symptoms in Alzheimer's disease patients: an analysis of the Neuropsychiatric Inventory (NPI) data of two randomised,

- controlled studies. *International journal of geriatric psychiatry*. 2005; 20:459–464. [PubMed: 15852444]
29. Hollingworth P, Hamshere ML, Moskvina V, et al. Four components describe behavioral symptoms in 1,120 individuals with late-onset Alzheimer's disease. *Journal of the American Geriatrics Society*. 2006; 54:1348–1354. [PubMed: 16970641]
30. Mirakhor A, Craig D, Hart DJ, et al. Behavioural and psychological syndromes in Alzheimer's disease. *International journal of geriatric psychiatry*. 2004; 19:1035–1039. [PubMed: 15481075]
31. Spalletta G, Musicco M, Padovani A, et al. Neuropsychiatric symptoms and syndromes in a large cohort of newly diagnosed, untreated patients with Alzheimer disease. *Am J Geriatr Psychiatry*. 2010; 18:1026–1035. [PubMed: 20808086]
32. Vilalta-Franch J, Lopez-Pousa S, Turon-Estrada A, et al. Syndromic association of behavioral and psychological symptoms of dementia in Alzheimer disease and patient classification. *Am J Geriatr Psychiatry*. 2010; 18:421–432. [PubMed: 20220583]
33. Guy, W. ECDEU Assessment Manual for Psychopharmacology —Revised (DHEW Publ No ADM 76-338). Rockville, MD, U.S.: 1976.
34. Cochran WG. The comparison of percentages in matched samples. *Biometrika*. 1950; 37:256–266. [PubMed: 14801052]
35. Fleiss, JL.; Levin, B.; Paik, MC. *Statistical Methods for Rates and Proportions*, Third. Hoboken New Jersey: Wiley; 2003.
36. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988; 44:837–845. [PubMed: 3203132]

Clinical Improvement (%)

Figure 1.
Overall Clinical Improvement for Each Instrument

Figure 2.

ROC Analysis Comparing Proportion of Target Symptoms Resolved* on the E-BEHAVE-AD, NBRs, and NPI versus the Overall Improvement of CGI-S

E-BEHAVE-AD: Empirical Behavioral Rating Scale; NBRs: Neurobehavioral Rating Scale; NPI: Neuropsychiatric Interview (NPI)

*Each point on the curves corresponds to a specific cut-off for the proportion of target symptoms resolved: 100%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, 0%. For example, a cut-off of 50% means that a subject is considered clinically improved if he or she has experienced a 50% or more reduction in the total number of his or her target symptoms (e.g., resolution of one target symptom if the subject presented with one target symptom);

resolution of 3 if he or she presented with 5). For each instrument and for each of these cut-offs, we plot the proportion of subjects correctly classified as improved (sensitivity, Y axis) and correctly classified as not improved (specificity, X axis) according to the CGI-S. For instance, with a cut-off of 100% reduction in target symptom, the sensitivity is 0% (no subject who has improved according to the CGI-S is correctly classified as having improved by the instrument) and the specificity is 100% (all subjects classified as not having improved by the CGI-S are correctly classified as not having improved by the instruments). With a cut-off score of 50% reduction in target symptoms, the sensitivity for the E-BEHAVE-AD is 67.9% (67.9% of the subjects who have improved according to the CGI-S are correctly classified as having improved) and the specificity is 88.1% (88.1% of subjects classified as not having improved by the CGI-S are correctly classified as not having improved). With a cut-off of 0% reduction in target symptom, the sensitivity is 100% (all subjects who have improved according to the CGI-S are correctly classified as having improved by the instrument) and the specificity is 0% (no subject classified as not having improved by the CGI-S is correctly classified as not having improved by the instruments).

Instrument='E-BEHAVE-AD'

Instrument='NBRS'

Figure 3.
Sensitivity and Specificity for the Relative Decrease in Number of Target Symptoms on the E-BEHAVE-AD, NBRBS, and NPI versus the Overall Improvement of CGI

Table 1**Baseline Subject Characteristics**

Data are reported as mean (SD) unless otherwise noted

N	87
Age	81.2 (8.5)
Sex (N -- % Male)	36 -- 41.4%
Race (N -- % White)	69 -- 79.3%
Mini Mental State Examination (N=62)	9.7 (7.9)
Severe Impairment Battery (N=45)	63.5 (33.9)
NBRS total score	59.9 (15.7)
NPI total Score	43.4 (19.0)
E-BEHAVE-AD total score	15.5 (6.0)

Table 2
Specific Target Symptoms and Agitation or Psychosis Syndromes at Baseline and Rates of Clinical Improvement

	E-BEHAVE-AD		NBRs		NPI	
	Present at Baseline N = 87 (100%)	Rate of Clinical Improvement	Present at Baseline N = 87 (100%)	Rate of Clinical Improvement	Present at Baseline N = 87 (100%)	Rate of Clinical Improvement
Agitation	N/A	N/A	63 (72%)	25/63 (40%)	69 (79%)	21/69 (30%)
Disinhibition	N/A	N/A	N/A	N/A	50 (57%)	24/50 (48%)
Aggression	85 (98%)	18/85 (21%)	64 (74%)	18/64 (28%)	N/A	N/A
Hostility	N/A	N/A	75 (86%)	25/75 (33%)	N/A	N/A
Irritability	N/A	N/A	N/A	N/A	65 (75%)	25/65 (38%)
Aberrant Motor Behavior	N/A	N/A	N/A	N/A	48 (55%)	18/48 (38%)
Activity Disturbance	79 (91%)	16/79 (20%)	N/A	N/A	N/A	N/A
Delusions	48 (55%)	21/48 (44%)	54 (62%)	20/54 (37%)	38 (44%)	17/38 (45%)
Hallucinations	24 (28%)	15/24 (63%)	26 (30%)	16/26 (62%)	18 (21%)	11/18 (61%)
Suspiciousness	N/A	N/A	23/87 (26%)	13/23 (57%)	N/A	N/A
Syndrome						
Agitation	87 (100%)	24/87 (28%)	85 (98%)	42/85 (49%)	83 (95%)	52/83 (63%)
Psychosis	58 (67%)	30/58 (52%)	62 (71%)	34/62 (55 %)	46 (53%)	23/46 (50%)

N/A: Not Applicable

For target symptoms, rates of clinical improvement are shown as the number of subjects in whom the target symptom has resolved over the number in whom it was present at baseline. For syndromes, rates of clinical improvement are shown as the number of subjects in whom at least one target symptom has resolved over the number in whom the syndrome was present at baseline.