Identifying cognitive impairment during the Annual Wellness Visit: Who can you trust?

This study found that patients, family members, and even physicians have trouble detecting cognitive impairment. A better bet: Routinely administer an objective cognition test.

Abstract

Purpose ▶ Assessing for cognitive impairment is now mandated as part of the Medicare Annual Wellness Visit. This offers an unparalleled opportunity for early detection and treatment of dementia. However, physician observation supplemented by reports of patients and informants may be less effective than an objective screening test to achieve this goal.

Methods ▶ We used visual analog cognition scales (VACS) to quantify patient and informant subjective impressions of cognitive ability and compared these scales with the Folstein Mini-Mental State Exam (MMSE) and the Memory Orientation Screening Test (MOST) on a sample of 201 elderly patients seen for neuropsychological evaluation in a tertiary memory evaluation center. Outcome measures included dementia severity and scores from 3 standardized memory tests. Depression was also considered.

Results ▶ Patients were unable to judge their own cognition. Family informants rated only slightly better. Both screening tests outperformed patients and informants. The MOST was significantly better than the MMSE for determining dementia severity and memory for the total sample, as well as a subsample of patients who were less impaired and more typical of independent community-dwelling elders. Depression did not influence the test relationships.

Conclusions ▶ Neither patient nor informant subjective reports of cognition should be relied on to identify cognitive impairment within the Annual Wellness Visit. Providers would be best served by using a valid and reliable screening test for dementia.

Abstract

Purpose ▶ Assessing for cognitive impairment is now mandated as part of the Medicare Annual Wellness Visit. This offers an unparalleled opportunity for early detection and treatment of dementia. However, physician observation supplemented by reports of patients and informants may be less effective than an objective screening test to achieve this goal.

Methods ▶ We used visual analog cognition scales (VACS) to quantify patient and informant subjective impressions of cognitive ability and compared these scales with the Folstein Mini-Mental State Exam (MMSE) and the Memory Orientation Screening Test (MOST) on a sample of 201 elderly patients seen for neuropsychological evaluation in a tertiary memory evaluation center. Outcome measures included dementia severity and scores from 3 standardized memory tests. Depression was also considered.

Results ▶ Patients were unable to judge their own cognition. Family informants rated only slightly better. Both screening tests outperformed patients and informants. The MOST was significantly better than the MMSE for determining dementia severity and memory for the total sample, as well as a subsample of patients who were less impaired and more typical of independent community-dwelling elders. Depression did not influence the test relationships.

Conclusions ▶ Neither patient nor informant subjective reports of cognition should be relied on to identify cognitive impairment within the Annual Wellness Visit. Providers would be best served by using a valid and reliable screening test for dementia.

As of January 2011, physicians are required to include detection of cognitive impairment as part of their health risk assessment in the Medicare Annual Wellness Visit. The Centers for Medicare and Medicaid Services (CMS) specifically mandate an “assessment of an individual’s cognitive function by direct observation, with due consideration of information obtained by way of patient report, concerns raised by family members, friends, caretakers, or others.” Unfortunately, these means of assessment may be unreliable.

Why observation alone won’t work.

Physicians often fail to identify cognitive impairment until it becomes quite severe. This failure to diagnose may be due to time constraints, a focus on other health measures, or the lack of appropriate and usable tools. Reliance on patient self-report is also likely to be a flawed approach. A recent study found that most patients with dementia in a community sample denied they had memory problems. This is consistent with
“I remember everything I need to remember” is a common response to the question, “How is your memory?”

During the comment period preceding implementation of the CMS regulation, 38 national organizations comprising the Partnership to Fight Chronic Disease argued that reliance on subjective measures alone is inadequate to achieve the stated goal of the legislation. We share this concern.

Improving cognition assessment. Although family complaints have been viewed as valid in at least 1 commonly used screening instrument, the AD8 (with more than 2 of 8 complaints likely to aid in dementia detection) does not reflect severity of impairment, nor does it provide a score to follow a patient’s course over time.

To better quantify the subjective perceptions of cognition by patients and their families, we developed the Visual Analog Cognition Scale (VACS)—which we’ll describe in a bit—and added it to our protocol of neuropsychological tests for dementia. Visual analog scales are well-accepted measures for a variety of subjective phenomena, including pain, treatment response, sleep, affective states, and quality of life. We designed this current study to delineate the degree to which patient or informant perspective could assist physicians in the identification process.

We examined VACS responses from a consecutive sample of patients seen in our practice from July through December 2010. Our goal was to quantify the perceptions of patients and their informants regarding patients’ cognitive states across 5 important areas and to determine the relationship between these ratings and the objective results of neuropsychological evaluation. We also wanted to measure the relative accuracy of such subjective ratings with that of 2 validated screening tools, the Folstein Mini-Mental State Exam (MMSE) and the recently published Memory Orientation Screening Test (MOST), which we developed.

METHODS

Subjects
We administered the VACS to 201 patients as part of a 4-hour comprehensive neuropsychological evaluation. Patients were referred by community-based physicians, typically in primary care, neurology, or psychiatry. The sample was 66% female (n=133), with an average age of 78.5 (±6.8) years and an average education of 13.2 (±3.2) years. Of the 201 patients, 7 could not complete the VACS because of confusion or visual impairment; 20 had no accompanying informant. Of the 181 accompanied patients, 89 informants were grown children (49%), 64 were spouses (35%), 12 were siblings (7%), and 16 were friends or paid caregivers (9%).

Procedure
An administrative assistant handed each patient and informant the VACS as they checked in at the front desk. We asked them to fill out the questionnaire in the waiting room and advised them not to discuss their ratings with each other. We then conducted a comprehensive neuropsychological evaluation of the patient while another clinician separately interviewed the informant regarding the patient’s current health, cognitive and emotional symptoms, and daily function.

Instruments
The VACS is a 5-item, visual analog scale with parallel forms for patients (VACS-P) and informants (VACS-I). The form instructs the user to “Rate yourself (or the patient with whom you came) in each of these 5 areas by circling a number that best describes how you (they) are doing.” The 5 areas and their descriptions are:

- Attention: Keeping focused, avoiding being distracted, completing tasks
- Initiation: Starting tasks, following through, staying busy and active
- Judgment: Figuring things out and making good decisions
- Memory: Remembering new information and how to do things
- Self-care: Dressing, bathing, preparing food.

Each area has a visual analog scale of 1 to 10 below it, with each number occupying...
a box in a continuous sequence. Words appear above some of the numbers to help anchor the ratings in a systematic way: 1=very poor; 4=fair; 7=good; 10=very good.

The MMSE and its properties are well known. The MOST is a 29-point scale comprising 3-word recall, orientation to 6 date- and-time items, unforewarned recall of 12 pictured household items, and an 8-point clock drawing score. The validation study, using a total sample exceeding 1000 patients, demonstrated the MOST correlated highly and significantly (Pearson’s correlation coefficient $r=0.81$; $P<.001$) with dementia severity and 3 standardized memory tests. At a cutoff score of 18 points, it produced a 0.90 area under the curve (AUC) (95% confidence interval [CI], 0.87-0.94), with a sensitivity of 0.85 and specificity of 0.76, correctly classifying 83% of patients. Test-retest reliability was $r>0.90$; $P<.001$ for both shorter (average 2-month) and longer (average 9-month) intervals.

With each patient, we conducted a diagnostic interview and administered a battery of standardized neuropsychological tests to assess intelligence, attention, executive function, language, and memory. The measures of primary interest for this investigation were the MOST, MMSE, delayed story memory (Wechsler Memory Scale-Revised [WMS-R] Logical Memory-II, or LM-II), delayed visual memory (WMS-R Visual Recall-II, or VR-II), delayed recall of a 12-item repeated presentation list of common grocery store items (Shopping List Test-Recall, or SLT-R), and the 15-item Geriatric Depression Scale (GDS-15). Additionally, each psychologist made a clinical diagnosis, according to Diagnostic and Statistical Manual of Mental Disorders [Fourth Edition] (DSM-IV) criteria and rated the patient’s dementia severity (DS) on a 0-to-3 Clinical Dementia Rating-type scale. We based diagnoses and severity ratings on age- and education-adjusted neuropsychological test scores, medical and psychiatric history, patient interview, and separate interview with a family informant.

**Statistical methods**

We calculated VACS totals for each patient and informant. Total VACS scores ranged from 5 to 50. MOST scores, comprising 3-word recall, 6-item orientation, 12-item list memory, and an 8-point clock drawing score, ranged from 0 to 29. We used the MMSE in the traditional method, counting the first error in spelling WORLD backwards, yielding a result of 0 to 30. The GDS score, 0 to 15, reflected the number of items indicating depression. We computed neuropsychological tests using standard scoring techniques. We rated dementia severity as: 0=normal cognition; 0.5=mild cognitive impairment; 1.0=mild dementia; 2.0=moderate dementia; and 3.0=severe dementia. We also assigned half-point ratings from 1 to 3.

We compared MOST, MMSE, VACS-P, and VACS-I scores with dementia severity and the 3 neuropsychological tests of delayed memory and the GDS-15. We computed Pearson’s correlation coefficients and their levels of significance vs 0. Tests of significant differences between correlations used Fisher’s $z$-transformation and tested the normalized difference vs 0.

**RESULTS**

Diagnoses and dementia severity levels are listed in [TABLE 1.]

**TABLE 2** presents the mean scores for predictor and outcome variables. Correlations and significance ratings between the VACS-P, VACS-I, MOST, and MMSE with the criterion variables of Dementia Severity Rating, LM-II, VR-II, SLT-R, and GDS-15 are shown in [TABLE 3]. Patients, on average, rated themselves as having “good” cognition overall. There was no difference in patient self-ratings between the top quartile of dementia severity (mean=34.6; SD= 8.6) and those in the lowest quartile (mean=36.4; SD=9.0). Informants rated the patients, on average, as having only “fair” cognition. Objective neuropsychological tests, however, found the patients, on average, to be mild to moderately demented and to have mild to moderate impairment on objective memory tests. Most patients were not depressed, with an average GDS score well below the clinical cutoff of 7 or more items. However, 30 of the 194 (15.5%) who completed the VACS-P fell into the clinical range for depression.
Cognition diagnoses and severity levels in 201 consecutively evaluated elderly patients

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal cognition</td>
<td>8 (4.0)</td>
</tr>
<tr>
<td>Mild cognitive impairment</td>
<td>32 (15.9)</td>
</tr>
<tr>
<td>Dementia of all types</td>
<td>161 (80.1)</td>
</tr>
<tr>
<td>– Alzheimer’s disease</td>
<td>90 (55.9)</td>
</tr>
<tr>
<td>– Vascular dementia</td>
<td>62 (38.5)</td>
</tr>
<tr>
<td>– Frontotemporal dementia</td>
<td>4 (2.5)</td>
</tr>
<tr>
<td>– Other dementia</td>
<td>5 (3.1)</td>
</tr>
</tbody>
</table>

Dementia severity rating

<table>
<thead>
<tr>
<th>Rating</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (0)</td>
<td>8 (4.0)</td>
</tr>
<tr>
<td>Mild cognitive impairment (0.5)</td>
<td>32 (15.9)</td>
</tr>
<tr>
<td>Mild dementia (1.0)</td>
<td>42 (20.9)</td>
</tr>
<tr>
<td>Mild-moderate dementia (1.5)</td>
<td>45 (22.4)</td>
</tr>
<tr>
<td>Moderate dementia (2.0)</td>
<td>38 (18.9)</td>
</tr>
<tr>
<td>Moderate–severe dementia (2.5)</td>
<td>27 (13.4)</td>
</tr>
<tr>
<td>Severe dementia (3.0)</td>
<td>9 (4.5)</td>
</tr>
</tbody>
</table>

Patient self-ratings did not correlate ($r=0.02$) with dementia severity or with any of the 3 standardized memory tests. Informant scores correlated modestly with dementia severity and memory tests, but were significantly higher ($P<.001$) than those of the patients. Both the MOST ($r=–0.86$) and the MMSE ($r=–0.76$) had much stronger and highly significant ($P<.001$) correlations with dementia severity and with the memory measures ($r=0.49–0.70$). In addition, the MOST and MMSE were significantly ($P<.001$) better correlated with dementia severity and objective memory scores than were the informant ratings. Only the MMSE correlation with visual recall ($P=.06$) did not surpass that of the informant.

The MOST had a significantly higher correlation than the MMSE with dementia severity ($P<.01$) and with each of the 3 memory tests ($P<.05$). The MOST and MMSE scores were not related to level of depression ($r=–0.01$ and $–0.03$). Patient reports correlated significantly with depression level ($r=–0.40$; $P<.001$) as did those of the informants ($r=–0.22$; $P<.01$). Nevertheless, depression did not appear to be responsible for the limited relationship between patient self-ratings and objective test scores for cognition. When clinically depressed (GDS≥7) patients were removed from the analysis (remaining n=166), there was no significant improvement in the correlation between subjective ratings and objective scores.

We conducted a secondary analysis of patients whose cognition ranged between normal and mild-to-moderate dementia to see if more cognitively intact individuals would be more accurate at self-rating. In this subsample (n=127; mean age=77.3 years; 57% females), patient self-reports again did not correlate significantly ($r=0.05$) with dementia severity. Informant ratings remained modest, but significant ($r=–0.25$; $P=.004$) and statistically better ($P<.05$) than those of the patients. The MOST ($r=–0.69$; $P<.001$) and the MMSE ($r=–0.54$; $P<.001$) remained well-correlated with dementia severity and with the memory measures ($r=0.49–0.70$). In addition, the MOST and MMSE were significantly ($P<.001$) better correlated with dementia severity and objective memory scores than were the informant ratings. Only the MMSE correlation with visual recall ($P=.06$) did not surpass that of the informant.

The MOST had a significantly higher correlation than the MMSE with dementia severity ($P<.01$) and with each of the 3 memory tests ($P<.05$). The MOST and MMSE scores were not related to level of depression ($r=–0.01$ and $–0.03$). Patient reports correlated significantly with depression level ($r=–0.40$; $P<.001$) as did those of the informants ($r=–0.22$; $P<.01$). Nevertheless, depression did not appear to be responsible for the limited relationship between patient self-ratings and objective test scores for cognition. When clinically depressed (GDS≥7) patients were removed from the analysis (remaining n=166), there was no significant improvement in the correlation between subjective ratings and objective scores.

We conducted a secondary analysis of patients whose cognition ranged between normal and mild-to-moderate dementia to see if more cognitively intact individuals would be more accurate at self-rating. In this subsample (n=127; mean age=77.3 years; 57% females), patient self-reports again did not correlate significantly ($r=0.05$) with dementia severity. Informant ratings remained modest, but significant ($r=–0.25$; $P=.004$) and statistically better ($P<.05$) than those of the patients. The MOST ($r=–0.69$; $P<.001$) and the MMSE ($r=–0.54$; $P<.001$) remained well-correlated with dementia severity and with the memory measures ($r=0.49–0.70$). In addition, the MOST and MMSE were significantly ($P<.001$) better correlated with dementia severity and objective memory scores than were the informant ratings. Only the MMSE correlation with visual recall ($P=.06$) did not surpass that of the informant.

The MOST had a significantly higher correlation than the MMSE with dementia severity ($P<.01$) and with each of the 3 memory tests ($P<.05$). The MOST and MMSE scores were not related to level of depression ($r=–0.01$ and $–0.03$). Patient reports correlated significantly with depression level ($r=–0.40$; $P<.001$) as did those of the informants ($r=–0.22$; $P<.01$). Nevertheless, depression did not appear to be responsible for the limited relationship between patient self-ratings and objective test scores for cognition. When clinically depressed (GDS≥7) patients were removed from the analysis (remaining n=166), there was no significant improvement in the correlation between subjective ratings and objective scores.

We conducted a secondary analysis of patients whose cognition ranged between normal and mild-to-moderate dementia to see if more cognitively intact individuals would be more accurate at self-rating. In this subsample (n=127; mean age=77.3 years; 57% females), patient self-reports again did not correlate significantly ($r=0.05$) with dementia severity. Informant ratings remained modest, but significant ($r=–0.25$; $P=.004$) and statistically better ($P<.05$) than those of the patients. The MOST ($r=–0.69$; $P<.001$) and the MMSE ($r=–0.54$; $P<.001$) remained well-correlated with dementia severity and with the memory measures ($r=0.49–0.70$). In addition, the MOST and MMSE were significantly ($P<.001$) better correlated with dementia severity and objective memory scores than were the informant ratings. Only the MMSE correlation with visual recall ($P=.06$) did not surpass that of the informant.

While informants possessed some knowledge about a patient’s cognitive status, their ratings correlated only modestly with objectively measured cognition.

### TABLE 2
Mean test scores for predictor and outcome variables

<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>Mean (SD)</th>
<th>Outcome variables</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOST</td>
<td>15.5 (5.7)</td>
<td>Dementia Severity Rating</td>
<td>1.5 (0.8)</td>
</tr>
<tr>
<td>MMSE</td>
<td>23.2 (4.7)</td>
<td>LM-II</td>
<td>6.4 (8.2)</td>
</tr>
<tr>
<td>VACS-P</td>
<td>35.6 (8.4)</td>
<td>VR-II</td>
<td>5.4 (7.7)</td>
</tr>
<tr>
<td>VACS-I</td>
<td>27.6 (10.2)</td>
<td>SLT-R</td>
<td>4.3 (3.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GDS-15</td>
<td>3.3 (3.3)</td>
</tr>
</tbody>
</table>

Identifying Cognitive Impairment

Results of this study demonstrate that patients referred for specialized memory evaluation had virtually no idea of the degree of their cognitive impairment. Patients, on average, rated their function in 5 critical areas of cognition and behavior as “good.” While 80% of these patients demonstrated dementia on formal evaluation, more than 95% rated themselves as having good or very good cognition. Their ratings did not correlate with any objective memory measures or expert clinician opinion.

- Patient and informant ratings are unreliable. Patients with better cognition, who might visit their doctor alone for the Annual Wellness Visit and would appear more intact, were no better at judging their cognition than the total patient sample. Both the patients with good cognition and those with dementia rated themselves equally unimpaired. This finding is not unique to the visual analog scale that we used in this study. When 148 self-nominated “cognitively healthy” community-dwelling elders took the MOST and a battery of neuropsychological tests as part of a norming study for the MOST,31 more than 20% would be classified as having dementia based on their memory and executive function test scores. These findings strongly suggest that patients cannot be relied on to inform their physician of cognitive impairment.

While informants possessed some knowledge about a patient’s cognitive status and were able to supply helpful anecdotal information, their ratings correlated only modestly with objectively measured cognition. This is not surprising given the volume of research demonstrating rater and observer bias.

- Rely instead on an objective cognitive screening test. Of greatest relevance, these results indicate that an objective cognitive screening test is more accurate in identifying and measuring cognitive impairment than is

### Table 3

<table>
<thead>
<tr>
<th></th>
<th>Correlations of MOST, MMSE, VACS-P, and VACS-I to criterion measures</th>
<th>Pairwise comparison of correlations of MOST, MMSE, and VACS-I to criterion measures (absolute values)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pearson’s correlation coefficient (P value*)</td>
<td>Z-ratio (P value*)</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>Z</td>
</tr>
<tr>
<td>Dementia severity</td>
<td>–0.86</td>
<td>2.835</td>
</tr>
<tr>
<td>LM-II</td>
<td>0.67</td>
<td>2.245</td>
</tr>
<tr>
<td>VR-II</td>
<td>0.65</td>
<td>2.481</td>
</tr>
<tr>
<td>SLT-R</td>
<td>0.70</td>
<td>2.223</td>
</tr>
<tr>
<td>GDS-15</td>
<td>–0.01</td>
<td>–0.22</td>
</tr>
</tbody>
</table>


*The minimum acceptable measure of statistical significance was .05.

Pearson’s correlation coefficient (at left) measures the strength of relationship between 2 variables. It can range from 0.0 (no correlation) to –1.0 or 1.0 (perfect correlation). The larger the number, the stronger the relationship. A negative coefficient indicates an inverse relationship.

Z-ratio (at right) reflects the size, or magnitude, of the difference between 2 correlations.
The MOST outperformed the MMSE in detecting dementia and determining disease severity on a sample of 700 patients, and demonstrated twice the sensitivity for disease detection in those who were mildly impaired. The current study confirms that the MOST has a significantly higher correlation with dementia severity than does the MMSE, and significantly higher correlations with longer standardized memory tests.

**MOST, MMSE test-taking time varies, too.** Time constraints are an important consideration in a medical office. The average time to administer the MOST on cognitively impaired patients (a group that is slower to perform than patients with normal cognition) is 4.5 minutes. The MMSE, by comparison, takes 10 minutes or more. Cognition is as measurable as body mass index, blood pressure, height, weight, and level of depression, also mandated in the Annual Wellness Visit. Numbers are easily recorded and compared, while impressions or even a positive (>2) AD8 score are less precise. Provider observation, even if informed by family report, is not as sound a basis for risk analysis, treatment planning, or future monitoring as is an objective measure. Because several current screening tests for dementia possess known reliabilities over time, the physician can periodically repeat such a test to assess treatment response and ongoing risk.

**Is there a place for a subjective rating scale?** Possibly. A waiting room tool such as the VACS, combined with an objective test, may alert the clinician to a patient with anosognosia. These patients require different management strategies if treatment is to be effective. The care team faces an even greater challenge if an informant shares the patient’s lack of awareness. Conversely, a favorable cognitive screening result and a high score from the informant would give all parties assurance that cognition was normal.

**Study limitations.** The primary limitation of this study is that it was conducted in a tertiary memory center, where most patients have either suspected or demonstrated cognitive deficits. The relative proportion of normal to impaired patients is, consequently, different from that found in the primary care office, in which about 15% would have mild cognitive impairment and a similar percentage would have dementia. A replication of this study in such an environment would be helpful. On the other hand, without a companion neuropsychological evaluation as a criterion, the accuracy of self- or informant-report is more difficult to measure. As noted above, 20% of elders volunteering for a study on “normal cognitive functioning” showed significant objective deficits.

Assessment of cognitive impairment in the primary care physician’s office is uniquely challenging. Physicians are taught to respond to the complaints of patients. But when a patient has dementia, that approach does not work. Family reports are helpful, but not sufficiently accurate. The recent Alzheimer’s Association report notes that “Medicare’s new Annual Wellness Visit includes assessment for possible cognitive impairment,” but also points out that “many existing barriers affect the ability or willingness of individuals and their caregivers to recognize cognitive impairment and to discuss it with their physician.” We agree, and we believe that a sound approach to this problem would be for primary care physicians to consistently use an objective tool to measure cognitive functioning in the Annual Wellness Visit and in follow-up visits. A score that reflects the current level of cognition, provides diagnostic information, and reflects change in cognitive status over time will optimize this unique opportunity for earlier detection and potentially earlier treatment of dementia.

---

**CORRESPONDENCE**
Mitchell Clionsky, PhD, ABPP (CN), Clionsky Neuro Systems, 155 Maple Street, Suite 203, Springfield, MA 01105; mitch@cns-neuro.com
References


