Focusing on Inattention: The Diagnostic Accuracy of Brief Measures of Inattention for Detecting Delirium

Annachiara Marra, MD, PhD1,2, James C. Jackson, PsyD1,3,4,5, E. Wesley Ely, MD, MPH1,3,4,6, Amy J. Graves, MS7, John F. Schnelle, PhD1,2,4,5, Robert S. Dittus, MD, MPH1,3,4,6, Amanda Wilson, MD9, Jin H. Han, MD, MSc3,6,10*

1Division of Allergy, Pulmonary, and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee; 2Department of Neurosciences, Reproductive and Odontostomatological Sciences, University of Naples, Federico II, Naples, Italy; 3Center for Health Services Research, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee; 4Research Service, Department of Veterans Affairs Medical Center, Tennessee Valley Healthcare System, Nashville, Tennessee; 5Geriatric Research, Education and Clinical Center (GRECC) Service, Department of Veterans Affairs Medical Center, Tennessee Valley Healthcare System, Nashville, Tennessee; 6Center for Quality Aging, Vanderbilt University Medical Center, Nashville, Tennessee; 7Division of General Internal Medicine and Public Health, Vanderbilt University Medical Center, Nashville, Tennessee; 8Department of Psychiatry, Vanderbilt University Medical Center, Nashville, Tennessee; 9Department of Emergency Medicine, Vanderbilt University Medical Center, Nashville, Tennessee; 10Department of Emergency Medicine, Vanderbilt University Medical Center, Nashville, Tennessee.

BACKGROUND: Delirium is frequently missed in most clinical settings. Brief delirium assessments are needed.

OBJECTIVE: To determine the diagnostic accuracy of reciting the months of year backwards (MOTYB) from December to July (MOTYB-6) and December to January (MOTYB-12) for delirium as diagnosed by a psychiatrist and to explore the diagnostic accuracies of the following other brief attention tasks: (1) spell the word “LUNCH” backwards, (2) recite the days of the week backwards, (3) 10-letter vigilance “A” task, and (4) 5 picture recognition task.

SETTING: Preplanned secondary analysis of a prospective observational study.

DESIGN: Emergency department located within an academic, tertiary care hospital.

PARTICIPANTS: 234 acutely ill patients who were ≥65 years old.

MEASUREMENTS: The inattention tasks were administered by a physician. The reference standard for delirium was a comprehensive psychiatrist assessment using Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision criteria.

RESULTS: Making any error on the MOTYB-6 task had a sensitivity of 80.0% (95% confidence interval [CI], 60.9%-91.1%) and specificity of 57.1% (95% CI, 50.4%-63.7%). Making any error on the MOTYB-12 task had a sensitivity of 84.0% (95% CI, 65.4%-93.6%) and specificity of 51.9% (95% CI, 45.2%-58.5%). The best combination of sensitivity and specificity was reciting the days of the week backwards task; if the patient made any error, this was 84.0% (95% CI, 65.4%-93.6%) sensitive and 81.9% (95% CI, 76.1%-86.5%) specific.

CONCLUSION: MOTYB-6 and MOTYB-12 had very good sensitivities but had modest specificities for delirium, limiting their use as a standalone assessment. Reciting the days of the week backwards appeared to have the best combination of sensitivity and specificity for delirium.

Journal of Hospital Medicine 2018;13:551-557. Published online first March 26, 2018. © 2018 Society of Hospital Medicine

Delirium is an acute neurocognitive disorder that affects up to 25% of older emergency department (ED) and hospitalized patients. The relationship between delirium and adverse outcomes is well documented. Delirium is a strong predictor of increased length of mechanical ventilation, longer intensive care unit and hospital stays, increased risk of falls, long-term cognitive impairment, and mortality. Delirium is frequently missed by healthcare professionals and goes undetected in up to three out of four patients by bedside nurses and medical practitioners in many hospital settings. A significant barrier to recognizing delirium is the absence of brief delirium assessments. In an effort to improve delirium recognition in the acute care setting, there has been a concerted effort to develop and validate brief delirium assessments. To address this unmet need, 4 ‘A’ Test (4AT), the Brief Confusion Assessment Method (bCAM), and the 3-minute diagnostic assessment for CAM-defined delirium (3D-CAM) are 1- to 3-minute delirium assessments that were validated in acutely ill older patients. However, 1 to 3 minutes may still be too long in busy clinical environments, and briefer (<30 seconds) delirium assessments may be needed.

One potential more-rapid method to screen for delirium is to specifically test for the presence of inattention, which is a
This study was conducted at a tertiary care, academic ED. The local institutional review board (IRB) reviewed and approved this study. Informed consent from the patient or an authorized surrogate was obtained whenever possible. Because this was an observational study and posed minimal risk to the patient, the IRB granted a waiver of consent for patients who were both unable to provide consent and were without an authorized surrogate available in the ED or by phone.

Selection of Participants
We enrolled a convenience sample of patients between June 2010 and February 2012 Monday through Friday from 8 AM to 4 PM. This enrollment window was based upon the psychiatrist's availability. Because of the extensiveness of the psychiatric evaluations, we limited enrollment to one patient per day. Patients who were 65 years or older, not in a hallway bed, and in the ED for less than 12 hours at the time of enrollment were included. We used a 12-hour cutoff so that patients who presented in the evening and early morning hours could be included. Patients were excluded if they were previously enrolled, non-English speaking, deaf or blind, comatose, suffered from end-stage dementia, or were unable to complete all the study assessments. The rationale for excluding patients with end-stage dementia was that diagnosing delirium in this patient population is challenging.

Research assistants approached patients who met inclusion criteria and determined if any exclusion criteria were present. If none of the exclusion criteria were present, then the research assistant reviewed the informed consent document with the patient or authorized surrogate if the patient was not capable of providing consent. If a patient was not capable of providing consent and no authorized surrogate was available, then the patient was enrolled (under the waiver of consent) as long as the patient assented to be a part of the study. Once the patient was enrolled, the research assistant contacted the physician rater and reference standard psychiatrists to approach the patient.

Measures of Inattention
An emergency physician (J.H.H.) who had no formal training in the mental status assessment of elders administered a cognitive battery to the patient, including tests of inattention. The following inattention tasks were administered:

- Spell the word “LUNCH” backwards. Patients were initially allowed to spell the word “LUNCH” forwards. Patients who were unable to perform the task were assigned 5 errors.
- Recite the months of the year backwards from December to July. Patients who were unable to perform the task were assigned 6 errors.
- Recite the days of the week backwards. Patients who were unable to perform the task were assigned 7 errors.
- Ten-letter vigilance “A” task. The patient was given a series of 10 letters (“S-A-V-E-A-H-A-A-R-T”) every 3 seconds and was asked to squeeze the rater’s hand every time the patient heard the letter “A.” Patients who were unable to perform the task were assigned 10 errors.
• Five-picture recognition task. Patients were shown 5 objects on picture cards. Afterwards, patients were shown 10 pictures with the previously shown objects intermingled. The patient had to identify which objects were seen previously in the first 5 pictures. Patients who were unable to perform the task were assigned 10 errors.

• Recite the months of the year backwards from December to January. Patients who were unable to perform the task were assigned 12 errors.

Reference Standard for Delirium
A comprehensive consultation- liaison psychiatrist assessment was the reference standard for delirium; the diagnosis of delirium was based on Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR) criteria. Three psychiatrists who each had an average of 11 years of clinical experience and regularly diagnosed delirium as part of their daily clinical practice were available to perform these assessments. To arrive at the diagnosis of delirium, they interviewed those who best understood the patient’s mental status (eg, the patient’s family members or caregivers, physician, and nurses). They also reviewed the patient’s medical record and radiology and laboratory test results. They performed bedside cognitive testing that included, but was not limited to, the Mini-Mental State Examination, Clock Drawing Test, Luria hand sequencing task, and tests for verbal fluency. A focused neurological examination was also performed (ie, screening for paraphasic errors, tremors, tone, asterixis, frontal release signs, etc.), and they also evaluated the patient for affective lability, hallucinations, and level of alertness. If the presence of delirium was still questionable, then confrontational naming, proverb interpretation or similarities, and assessments for apraxias were performed at the discretion of the psychiatrist. The psychiatrists were blinded to the physician’s assessments, and the assessments were conducted within three hours of each other.

Additional Variables Collected
Using medical record review, comorbidity burden, severity of illness, and premorbid cognition were ascertained. The Charlson Comorbidity Index, a weighted index that takes into account the number and seriousness of 19 preexisting comorbid conditions, was used to quantify comorbidity burden; higher scores indicate higher comorbid burden. The Acute Physiology Score of the Acute Physiology and Chronic Health Evaluation II was used to quantify severity of illness. This score is based upon the initial values of 12 routine physiologic measurements such as vital sign and laboratory abnormalities; higher scores represent higher severities of illness. The medical record was reviewed to ascertain the presence of premorbid cognitive impairment; any documentation of dementia in the patient’s clinical problem list or physician history and physical examination from the outpatient or inpatient settings was considered positive. The medical record review was performed by a research assistant and was double-checked for accuracy by one of the investigators (JHH).

Data Analyses
Measures of central tendency and dispersion for continuous variables were reported as medians and interquartile ranges. Categorical variables were reported as proportions. Receiver operating characteristic curves were constructed for each inattention task. Area under the receiver operating characteristic curves (AUC) was reported to provide a global measure of diagnostic accuracy. Sensitivities, specificities, positive likelihood ratios (PLRs), and negative likelihood ratios (NLRs) with their 95% CIs were calculated using the psychiatrist’s assessment as the reference standard. Cut-points with PLRs greater than 10 (strongly increased the likelihood of delirium) or NLRs less than 0.1 (strongly decreased the likelihood of delirium) were preferentially reported whenever possible.

All statistical analyses were performed with open source R statistical software version 3.0.1 (http://www.r-project.org/), SAS 9.4 (SAS Institute, Cary, North Carolina), and Microsoft Excel 2010 (Microsoft Inc., Redmond, Washington).

RESULTS
A total of 542 patients were screened; 214 patients refused to participate, and 93 were excluded, leaving 235 patients. The patient characteristics can be seen in Table 1. Compared with all patients (N = 15,359) who presented to the ED during the study period, enrolled patients were similar in age but more likely to be male, have cardiovascular chief complaints, and be admitted to the hospital. Of those enrolled, 25 (10.6%) were delirious. Delirious patients were older, more likely to be non-white, have a past history of dementia, have a graduate school degree, and have a chief complaint of altered mental status.

Making any error on the MOTYB-6 task had a sensitivity of 80.0% (95% CI, 60.9%-91.1%), specificity of 57.1% (95% CI, 50.4%-63.7%), PLR of 1.87 (95% CI, 1.45-2.40) and NLR of 0.35 (95% CI, 0.16-0.77) for delirium as diagnosed by a psychiatrist. Making any error on the MOTYB-12 task had a sensitivity of 84.0% (95% CI, 65.4%-93.6%), specificity of 51.9% (95% CI, 45.2%-58.5%), PLR of 1.75 (95% CI, 1.40-2.18), and NLR of 0.31 (95% CI, 0.12-0.76) for delirium. The AUCs for the MOTYB-6 and MOTYB-12 tasks were 0.77 and 0.78, respectively, indicating very good diagnostic performance.

The diagnostic performances of the other inattention tasks and additional cutoff values for the MOTYB-6 and MOTYB-12 tasks can be seen in Table 2. Increasing the MOTYB-6 cut-off to two or more errors and MOTYB-12 cut-off to three or more errors increased the specificity to 70.0% and 70.5%, respectively, without decreasing their sensitivity. The best combination of sensitivity and specificity was reciting the days of the week backwards; if the patient made any error, this was 84.0% (95% CI, 65.4%-93.6%) sensitive and 81.9% (95% CI, 76.1%-86.5%) specific for delirium. The inattention tasks that strongly increased the likelihood of delirium (PLR > 10) were the vigilance “A” and picture recognition tasks. If the patient made two or more errors on the vigilance task or three or more errors on the picture recognition task, then the likelihood of delirium strongly increased, as evidenced by a PLR of 16.80 (95% CI, 8.01-35.23) and 23.10 (95% CI, 7.95-67.12), respectively. No other inattention tasks...
were able to achieve a PLR of greater than 10, regardless of what cutoff was used. No inattention task was able to achieve a NLR of less than 0.10, which would have strongly decreased the likelihood of delirium. The best NLRs were if the patient made no errors spelling the word “LUNCH” backwards (NLR, 0.16; 95% CI, 0.04-0.60), no errors on the vigilance “A” task (NLR, 0.18; 95% CI, 0.07-0.43), and no errors on the days of the week backwards task (NLR, 0.20; 95% CI, 0.08-0.48).

**DISCUSSION**

Delirium is frequently missed by healthcare providers because it is not routinely screened for in the acute care setting. To help address this deficiency of care, we evaluated several brief measures of inattention that take less than 30 seconds to complete. We observed that any errors made on the MOTYB-6 and MOTYB-12 tasks had very good sensitivities (80% and 84%) but were limited by their modest specificities (approximately 50%) for delirium. As a result, these assessments have limited clinical utility as standalone delirium screens. We also explored other commonly used brief measures of inattention and at a variety of error cutoffs. Reciting the days of the week backwards appeared to best balance sensitivity and specificity. None of the inattention measures could convincingly rule out delirium (NLR < 0.10), but the vigilance “A” and picture recognition tasks may have clinical utility in ruling in delirium (PLR > 10). Overall, all the inattention tasks, including MOTYB-6 and MOTYB-12, had very good diagnostic performances based upon their AUC. However, achieving a high sensitivity often had to be sacrificed for specificity or, alternatively, achieving a high specificity had to be sacrificed for sensitivity.

Inattention has been shown to be the cardinal feature for delirium,
14 and its assessment using cognitive testing has been recommended to help identify the presence of delirium according to an expert consensus panel.26 The diagnostic performance of the MOTYB-12 observed in our study is similar to a study by Fick et al., who reported that MOTYB-6 and MOTYB-12, had very good diagnostic performances based upon their AUC. However, achieving a high sensitivity often had to be sacrificed for specificity or, alternatively, achieving a high specificity had to be sacrificed for sensitivity.

**TABLE 1.** Demographic and Clinical Characteristics of Patients

<table>
<thead>
<tr>
<th>Patient Characteristics</th>
<th>Enrolled Patients (n = 235)</th>
<th>All Potentially Eligible Patients (N = 15,359)</th>
<th>Nondelirious N = 210</th>
<th>Delirious N = 25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age (IQR)</td>
<td>74 (69, 79)</td>
<td>74 (69, 81)</td>
<td>74 (69, 79)</td>
<td>84 (67, 86)</td>
</tr>
<tr>
<td>Female gender</td>
<td>107 (45.5%)</td>
<td>8,198 (53.4%)</td>
<td>96 (45.7%)</td>
<td>11 (44.0%)</td>
</tr>
<tr>
<td>Nonwhite race</td>
<td>31 (13.9%)</td>
<td>-</td>
<td>24 (11.4%)</td>
<td>7 (28.0%)</td>
</tr>
<tr>
<td>Residence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home</td>
<td>215 (91.5%)</td>
<td>192 (91.4%)</td>
<td>23 (92.0%)</td>
<td></td>
</tr>
<tr>
<td>Assisted living</td>
<td>15 (6.4%)</td>
<td>14 (6.7%)</td>
<td>1 (4.0%)</td>
<td></td>
</tr>
<tr>
<td>Rehabilitation/postacute</td>
<td>2 (0.9%)</td>
<td>2 (1.0%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Nursing home</td>
<td>3 (1.3%)</td>
<td>2 (1.0%)</td>
<td>1 (4.0%)</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary or below</td>
<td>6 (2.6%)</td>
<td>5 (2.4%)</td>
<td>1 (4.0%)</td>
<td></td>
</tr>
<tr>
<td>Middle school</td>
<td>25 (10.6%)</td>
<td>21 (10.0%)</td>
<td>4 (16.0%)</td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>95 (40.4%)</td>
<td>86 (41.0%)</td>
<td>9 (36.0%)</td>
<td></td>
</tr>
<tr>
<td>College</td>
<td>71 (30.2%)</td>
<td>65 (31.0%)</td>
<td>6 (24.0%)</td>
<td></td>
</tr>
<tr>
<td>Graduate school</td>
<td>38 (16.2%)</td>
<td>33 (15.7%)</td>
<td>5 (20.0%)</td>
<td></td>
</tr>
<tr>
<td>Dementia in medical record</td>
<td>17 (7.2%)</td>
<td>10 (4.8%)</td>
<td>7 (28.0%)</td>
<td></td>
</tr>
<tr>
<td>Median Charlson (IQR)</td>
<td>2 (1, 4)</td>
<td>2 (1, 4)</td>
<td>3 (2, 6)</td>
<td></td>
</tr>
<tr>
<td>Median APS (IQR)</td>
<td>3 (1, 5)</td>
<td>3 (1, 5)</td>
<td>4 (1, 5)</td>
<td></td>
</tr>
<tr>
<td>ED chief complaint</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal pain</td>
<td>8 (3.4%)</td>
<td>854 (5.6%)</td>
<td>7 (3.3%)</td>
<td>1 (4.0%)</td>
</tr>
<tr>
<td>Altered mental status</td>
<td>14 (6.0%)</td>
<td>617 (4.0%)</td>
<td>2 (1.0%)</td>
<td>12 (48.0%)</td>
</tr>
<tr>
<td>Chest pain</td>
<td>43 (18.3%)</td>
<td>1575 (10.3%)</td>
<td>39 (18.6%)</td>
<td>4 (16.0%)</td>
</tr>
<tr>
<td>General weakness</td>
<td>21 (8.9%)</td>
<td>1,101 (7.2%)</td>
<td>18 (8.8%)</td>
<td>3 (12.0%)</td>
</tr>
<tr>
<td>Shortness of breath</td>
<td>31 (13.2%)</td>
<td>1,377 (9.0%)</td>
<td>29 (13.8%)</td>
<td>2 (8.0%)</td>
</tr>
<tr>
<td>Syncope</td>
<td>14 (6.0%)</td>
<td>422 (2.3%)</td>
<td>14 (6.7%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>Admitted to the hospital</td>
<td>168 (71.5%)</td>
<td>9491 (61.8%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: The APS is part of the APACHE II. Abbreviations: APACHE II, Acute Physiology and Chronic Health Evaluation II; APS, Acute Physiology Score; ED, emergency department; IQR, inter-quartile range.
Inattention measures can influence overall sensitivity and specificity. Objective measures of delirium, including the inattention screens studied, are particularly prone to spectrum bias. 31,34 However, the strength of this approach is that the assessment of inattention becomes less reliant upon clinical judgment and allows it to be used by raters from a wide range of clinical backgrounds. On the other hand, a subjective interpretation of these inattention tasks may allow the rater to capture the subtleties of inattention (ie, decreased speed of performance in a highly intelligent and well-educated patient without dementia). The disadvantage of this approach, however, is that it is more dependent on clinical judgment and may have decreased diagnostic accuracy in those with less clinical experience or with limited training.14,42,43

These factors must be carefully considered when determining which delirium assessment to use.
Additional research is required to determine the clinical utility of these brief inattention assessments. These findings need to be further validated in larger studies, and the optimal cutoff of each task for different subgroup of patients (eg, demented vs nondemented) needs to be further clarified. It is not completely clear whether these inattention tests can serve as standalone assessments. Depending on the cutoff used, some of these assessments may have unacceptable false negative or false positive rates that may lead to increased adverse patient outcomes or increased resource utilization, respectively. Additional components or assessments may be needed to improve the diagnostic accuracy of these assessments. In addition to understanding these inattention assessments’ diagnostic accuracies, their ability to predict adverse outcomes also needs to be investigated. While a previous study observed that making any error on the MOTYB-12 task was associated with increased physical restraint use and prolonged hospital length of stay,44 these assessments’ ability to prognosticate long-term outcomes such as mortality or long-term cognition or function need to be studied. Lastly, studies should also evaluate how easily implementable these assessments are and whether improved delirium recognition leads to improved patient outcomes.

This study has several notable limitations. Though planned a priori, this was a secondary analysis of a larger investigation designed to validate three delirium assessments. Our sample size was also relatively small, causing our 95% CIs to overlap in most cases and limiting the statistical power to truly determine whether one measure is better than the other. We also asked the patient to recite the months backwards from December to July as well as recite the months backwards from December to January. It is possible that the patient may have performed better at going from December to January because of learning effect. Our reference standard for delirium was based upon DSM-IV-TR criteria. The new DSM-V criteria may be more restrictive and may slightly change the sensitivities and specificities of the inattention tasks. We enrolled a convenience sample and enrolled patients who were more likely to be male, have cardiovascular chief complaints, and be admitted to the hospital; as a result, selection bias may have been introduced. Lastly, this study was conducted in a single center and enrolled patients who were 65 years and older. Our findings may not be generalizable to other settings and in those who are less than 65 years of age.

CONCLUSIONS

The MOTYB-6 and MOTYB-12 tasks had very good sensitivities but modest specificities (approximately 50%) using any error made as a cutoff; increasing cutoff to 2 errors and 3 errors, respectively, improved their specificities (approximately 70%) with minimal impact to their sensitivities. Reciting the days of the week backwards, spelling the word “LUNCH” backwards, and the 10-letter vigilance “A” task appeared to perform the best in ruling out delirium but only moderately decreased the likelihood of delirium. The 10-letter Vigilance “A” and picture recognition task appeared to perform the best in ruling in delirium. Days of the week backwards appeared to have the best combination of sensitivity and specificity.

Disclosure: The authors report no financial conflicts of interest.
Funding: This study was funded by the Emergency Medicine Foundation Career Development Award, National Institutes of Health K23AG032335, and National Center for Research Resources, Grant UL1 RR024975-01.

References


