Reliability and Validity of Prehospital Case Finding for Depression and Cognitive Impairment

Manish N. Shah, MD, MPH,* Jurgis Karuza, PhD,§ Erik Rueckmann, MD,* Peter Swanson,* Yeates Conwell, MD,§ and Paul Katz, MD

OBJECTIVES: To evaluate the test–retest reliability, the concurrent criterion validity, and the construct validity of prehospital, emergency medical service (EMS) case finding for depression and cognitive impairment in older adults.

DESIGN: Cross-sectional study.

SETTING: Prehospital EMS system and hospital emergency department.

PARTICIPANTS: EMS providers and community-dwelling older adult (aged ≥ 60) patients.

INTERVENTIONS: Case finding instruments for depression (Patient Health Questionnaire-2; PHQ-2) and cognitive impairment (Six-Item Screener).

MEASUREMENTS: The reliability and validity of these instruments.

RESULTS: Moderate test–retest reliability was found for prehospital application of the PHQ-2 (kappa = 0.50) and Six-Item Screener (kappa = 0.52), fair concurrent criterion validity for depression (kappa = 0.36), and slight to fair concurrent criterion validity for cognitive impairment (kappa = 0.11–0.23). Construct validity was demonstrated using the Multitrait-Multimethod Matrix.

CONCLUSION: Moderate test–retest reliability and construct validity were demonstrated for prehospital case finding by EMS providers for cognitive impairment and depression using these instruments. Slight to fair concurrent criterion validity was found, a result that methodological limitations could explain. These findings provide additional support for the concept of using EMS providers to detect older adults at risk for these conditions. Further work is needed to confirm the validity and effectiveness of prehospital screening before such programs are implemented.

Key words: emergency medical services; prehospital care; depression; cognitive impairment; geriatrics

Older adults are a large, vulnerable segment of the population, and many suffer from undetected health problems. Failure of healthcare providers to identify and treat these conditions can lead to unnecessary morbidity and mortality and a poorer quality of life.1,2 Depression and cognitive impairment are significant conditions that affect older adults and often go undetected. Up to 20% of community-dwelling older adults suffer from clinically significant depressive symptoms, which are associated with greater use of medical services, disability, and death,3,4 although estimates suggest that less than half of depressed older adults are identified and treated, despite the existence of effective therapies.5 This failure to identify and treat depressed older adults results in many at-risk individuals suffering unnecessarily.

Cognitive impairment is also common, affecting 10% of community-dwelling seniors and more than 20% of older adult emergency department (ED) patients.6,7 It has been associated with significant morbidity and mortality and high healthcare use and cost.2,8,9 Detection, particularly early detection, of cognitive impairment benefits older adults, because reversible causes can be treated, existing therapies to slow the progression of the disease can be provided, and families can prepare for the advancement of the disease, but cognitive impairment is poorly identified and treated in primary care practices.10,11 This failure to identify and treat cognitively impaired patients results in many at-risk individuals suffering unnecessarily.

Traditionally, screening and case finding have been the responsibility of primary care providers (PCPs). This role has become more difficult as demands on PCPs’ time increase and results in individuals who are at risk for preventable or treatable conditions not being identified.12,13 One potential solution to this problem involves developing innovative case finding systems in alternative settings and...
then communicating those findings to intervention programs. If successful, integrated case finding and intervention programs would improve the health and quality of life of these individuals. Given the prevalence and clinical importance of cognitive impairment and depression and the existence of interventions, older adults may particularly benefit from case finding to identify these conditions.

One alternative setting proposed for case finding is the prehospital, emergency medical services (EMS) setting, because older adults frequently use the EMS system, and EMS providers spend extended periods of time with patients. During emergency (911) responses, EMS providers may be able to use screening instruments to identify older adults at risk for various conditions, thus helping to ensure that at-risk individuals are identified. This information can then be provided to emergency physicians for acute evaluation and intervention, to PCPs for outpatient evaluation and interventions, to the patient and family members to pursue further evaluation and interventions, or to a single point-of-entry system that would coordinate evaluation and interventions. Recent work has documented EMS providers’ interest in participating in public health activities during emergency responses and the feasibility of prehospital case finding, although the reliability and validity of EMS case finding during emergency responses have not been evaluated. This is an essential next step in evaluating the effectiveness of EMS case finding, which is critical to informing policy changes that make greater public health use of EMS.

The objective of this study was to evaluate the test–retest reliability, the concurrent criterion validity, and the construct validity of EMS case finding for depression and cognitive impairment using previously validated screening tools for depression and cognitive impairment. It was hypothesized that EMS providers can reliably use the Patient Health Questionnaire-2 (PHQ-2) instrument to identify older adults with depression and the Six Item Screener (SIS) to identify older adults with cognitive impairment, that the PHQ-2 and SIS instruments have concurrent criterion validity in the prehospital setting, and that prehospital case finding for these conditions has construct validity when evaluated using the Multitrait-Multimethod Matrix.

METHODS

Design
A cross-sectional study of older adults (aged ≥60) cared for by participating EMS providers was performed between June and December 2007. The University of Rochester Medical Center’s Research Subjects Review Board approved this study with informed consent.

Setting
This study took place in Monroe County, New York, and involved two EMS agencies, Rural Metro Medical Services and Henrietta Volunteer Ambulance. Both agencies provide basic life support care, such as splinting and lifting, with emergency medical technicians (EMTs) and advanced life support care, such as intravenous medications and intubation, with paramedics.

Protocol
To develop the case finding instrument, tools validated in primary care to identify older adults with depression or cognitive impairment were identified. For depression, the PHQ-2 was chosen, and for cognitive impairment, the SIS was chose.

EMS providers from both agencies were invited to participate. Eighteen EMS providers, 12 paramedics, and six EMTs consented, completed the training, and participated. The training consisted of the basic life support version of the Geriatrics Education for EMS course and reviewing the study procedures and the case finding instruments.

Participating EMS providers were instructed to apply the case finding instrument to all community-dwelling patients aged 60 and older who requested emergency assistance and transport to either of two university EDs. Patients who could not speak English or who refused transport were excluded.

After arriving at the ED, the EMS providers called the study coordinator to provide the screening results. Separately, a trained research assistant, blinded to the EMS results, interviewed consenting patients. Subjects completed the SIS and the PHQ-2 to evaluate the test–retest reliability; completed the Mini-Cog and the CLOX to assess the concurrent criterion validity of the SIS and completed the PHQ-9 to assess the concurrent criterion validity of the PHQ-2. Finally, the research assistant used the Confusion Assessment Method (shortened version) to evaluate for delirium. The medical records were reviewed for patients’ medication list, medical history, and presenting complaint and vital signs.

Two physician investigators (MNS, ER) scored the Mini-Cog and CLOX independently. In the event of a discrepancy between the two scores, the two investigators discussed the results and reached a consensus score.

ED and EMS medical records were also reviewed for individuals who elected not to participate in the study using a short, structured data abstraction instrument.

Scales
Two scales, derived from the Primary Care Evaluation of Mental Disorders, were used to evaluate depression. A version of the PHQ-2 with dichotomous answers was used for EMS case finding because of the ease of administration and scoring and excellent test characteristics when used in a multistage process. A positive answer to either question indicated that the subject was depressed. The PHQ-9, used for criterion validation, has been validated in the primary care setting. This was chosen for the criterion standard, because it has excellent test characteristics, with a sensitivity and specificity of 88%. The diagnostic algorithm was used to determine whether patients screened positive for minor or major depression. From one perspective, the comparison of the PHQ-2 with the PHQ-9 could be seen as cross-validation of the PHQ-2, because the PHQ-2 items appear in the PHQ-9, but applying the term “concurrent criterion validity,” although not ideal, is reasonable given the dissimilarities in the administration and scoring of the version of the PHQ-2 used and the fact that the PHQ-9 has credibility as the established test in the literature.

The SIS was chosen for cognitive impairment, because it is easy to administer and score and has excellent test characteristics. It has been validated in the primary care setting.
Having more than two errors on the instrument was considered positive for cognitive impairment. For criterion validation, two instruments were chosen. The Mini-Cog has been validated in community samples and has excellent test characteristics. The CLOX evaluates executive dysfunction and has been shown to be easy to administer and to be well tolerated by older adults. It includes an unprompted clock-drawing task sensitive to executive control (CLOX1) and a prompted version that is not (CLOX2). Standard scoring schemes were used for both tests.

**Data Analysis**

All data were analyzed using Stata 8.0 (StataCorp., College Station, TX). Descriptive statistics were calculated to characterize the individuals participating in the study. Those who did not consent were compared with those participating to evaluate for enrollment bias. Continuous data were compared using a t-test, and categorical data were compared using the chi-square test.

Test–retest reliability was evaluated by calculating the percentage concordance in final diagnoses and the kappa statistic with 95% confidence intervals (CIs). Concurrent criterion validity was calculated in two ways. First, the sensitivity and specificity of the EMS-applied instrument were compared with those of the reference standard instrument. Second, the kappa statistic between the two instruments were calculated. Finally, the concurrent criterion validation was augmented by examining the validity of the scales using the construct validity approach. The study design permits the use of the multitrait-multimethod approach, which measures multiple traits (e.g., cognitive function and depression) using multiple screening methods and tools to determine the convergent and discriminant validity of the scales.

**RESULTS**

Participating EMS providers screened 269 eligible subjects; 187 (70%) consented to participate, 43 refused (16%) and 39 could not give consent (14%) because of the absence of decisional capacity and a surrogate. Table 1 demonstrates the characteristics of older adults who consented to participate in the study and the limited characteristics obtained on older adults who refused to participate. Table 1 also demonstrates the results from the application of the various instruments on the study population. Many subjects could not complete the Mini-Cog (14%) and CLOX (30%) because of the inability to draw clock faces. Higher levels of cognitive impairment were found in subjects completing the Mini-Cog and CLOX than in those completing the SIS.

Table 2 shows the reliability of performing the SIS and PHQ-2 in the prehospital setting. The kappas showed moderate test–retest reliability. The concordance between the two applications of the SIS was 87% (161/185; 95% CI = 81–92%) and the concordance between the two applications of the PHQ-2 was 75% (128/170; 95% CI = 68–82%). When stratified for presence of delirium, no statistically significant differences were found in the reliability of performing either instrument. Furthermore, stratifying according to cognitive impairment (SIS applied in the ED) did not indicate any statistically significant differences in the PHQ-2 reliability (results not shown).

### Table 1. Characteristics of Older Adults

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Consenting (n = 187)</th>
<th>Not Consenting (n = 82)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean ± standard deviation</td>
<td>75.6 ± 9.2</td>
<td>76.4 ± 9.9</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>120 (64)</td>
<td>43 (53)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>128 (68)</td>
<td>61 (74)</td>
</tr>
<tr>
<td>Black</td>
<td>55 (29)</td>
<td>18 (22)</td>
</tr>
<tr>
<td>Other or unknown</td>
<td>4 (2)</td>
<td>3 (4)</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Hispanic</td>
<td>181 (97)</td>
<td>NA</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1 (0.5)</td>
<td>NA</td>
</tr>
<tr>
<td>No answer</td>
<td>5 (2.7)</td>
<td>NA</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; High school</td>
<td>47 (25)</td>
<td>NA</td>
</tr>
<tr>
<td>High school</td>
<td>71 (38)</td>
<td>NA</td>
</tr>
<tr>
<td>College or more</td>
<td>58 (31)</td>
<td>NA</td>
</tr>
<tr>
<td>No answer</td>
<td>11 (5.9)</td>
<td></td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>22 (12)</td>
<td>NA</td>
</tr>
<tr>
<td>Married</td>
<td>59 (32)</td>
<td>NA</td>
</tr>
<tr>
<td>Widowed</td>
<td>61 (33)</td>
<td>NA</td>
</tr>
<tr>
<td>Separated or divorced</td>
<td>38 (20)</td>
<td>NA</td>
</tr>
<tr>
<td>No answer</td>
<td>7 (3.7)</td>
<td>NA</td>
</tr>
<tr>
<td>Accessed emergency department for care in previous 6 months, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospitalized in previous 6 months, n (%)</td>
<td>43 (23)</td>
<td>NA</td>
</tr>
<tr>
<td>Chief complaint, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trauma/hemorrhage</td>
<td>40 (21)</td>
<td>17 (21)</td>
</tr>
<tr>
<td>Cardiac</td>
<td>33 (18)</td>
<td>4 (5)</td>
</tr>
<tr>
<td>Gastrointestinal or genitourinary</td>
<td>22 (12)</td>
<td>9 (11)</td>
</tr>
<tr>
<td>Syncope or dizzy</td>
<td>23 (12)</td>
<td>6 (7)</td>
</tr>
<tr>
<td>Respiratory</td>
<td>25 (13)</td>
<td>4 (5)</td>
</tr>
<tr>
<td>Pain, not traumatic</td>
<td>12 (6)</td>
<td>9 (11)</td>
</tr>
<tr>
<td>Weakness or general illness</td>
<td>10 (5)</td>
<td>9 (11)</td>
</tr>
<tr>
<td>Neurological (cerebrovascular accident or seizure)</td>
<td>4 (2)</td>
<td>7 (9)</td>
</tr>
<tr>
<td>Endocrine</td>
<td>8 (4)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Alcohol of drug overdose</td>
<td>1 (1)</td>
<td>8 (10)</td>
</tr>
<tr>
<td>Confusion</td>
<td>2 (1)</td>
<td>3 (4)</td>
</tr>
<tr>
<td>Psychiatric</td>
<td>1 (1)</td>
<td>4 (5)</td>
</tr>
<tr>
<td>Other</td>
<td>6 (3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Length of illness symptoms, hours, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 4</td>
<td>53 (28)</td>
<td>NA</td>
</tr>
<tr>
<td>≥ 4</td>
<td>134 (72)</td>
<td>NA</td>
</tr>
<tr>
<td>Past medical history, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td>46 (25)</td>
<td>NA</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>25 (13)</td>
<td>NA</td>
</tr>
<tr>
<td>Dementia</td>
<td>16 (8.6)</td>
<td>NA</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>71 (38)</td>
<td>NA</td>
</tr>
<tr>
<td>Heart disease</td>
<td>68 (36)</td>
<td>NA</td>
</tr>
<tr>
<td>Hypertension</td>
<td>133 (71)</td>
<td>NA</td>
</tr>
<tr>
<td>Kidney disease</td>
<td>29 (16)</td>
<td>NA</td>
</tr>
<tr>
<td>Liver disease</td>
<td>8 (4.3)</td>
<td>NA</td>
</tr>
<tr>
<td>Lung disease</td>
<td>45 (24)</td>
<td>NA</td>
</tr>
</tbody>
</table>

(Continued)
which reflect the relationship between measures of the same trait, using different measures (range 0.12–0.58), followed by the heterotrait-heteromethod, which reflect the relationship between measures that share neither the same trait nor the same method (range -0.12–0.013). This pattern, in which the cognitive case finding results were more highly intercorrelated among themselves than with the depression screening results, and the depression screening results were more highly intercorrelated among themselves than with the cognitive screening results, provides evidence for the convergent and discriminant validity of the two screening tools. This is the order expected for a set of measures with construct validity.

DISCUSSION

This study found that the test–retest reliability of prehospital EMS provider application of the PHQ-2 was moderate (kappa = 0.50). The test–retest reliability of prehospital application of the SIS was also moderate (kappa = 0.52). This level of reliability is good and supports prehospital case finding for depression with the PHQ-2 and cognitive impairment with the SIS.

Three factors that may have altered the reliability must temper these results. First, the evaluation was not ideal, because the reference standard assessment was conducted in the ED, not the ambulance, and a short delay occurred between the two assessments. Practical and ethical considerations limited the ability to perform the ideal evaluation. This change threatens the external validity and potentially reduces the level of reliability identified. Second, a Hawthorne effect may have existed, potentially improving the results. This was a highly motivated group of EMS providers who were aware they were being evaluated. If universally implemented, this program may not be as successful. Third, before this study, EMS providers had not regularly evaluated their patients for depression or cognitive impairment as part of case finding or clinical care. As the EMS providers become more comfortable with these concepts and instruments, their skills and their results may improve, resulting in better reliability. Nonetheless, the reliability testing results support continued research evaluating prehospital case finding for depression and cognitive impairment with the PHQ-2 and SIS.

Table 1. (Contd.)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Consenting (n = 187)</th>
<th>Not Consenting (n = 82)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke</td>
<td>28 (15)</td>
<td>NA</td>
</tr>
<tr>
<td>Glasgow Coma Scale score = 15, n (%)</td>
<td>178 (95)</td>
<td>72 (88)</td>
</tr>
<tr>
<td>Cognitive impairment evaluation, n (%)</td>
<td>31/186 (17)</td>
<td>NA</td>
</tr>
<tr>
<td>Six-Item Screener by emergency medical services, impaired</td>
<td>29/186 (16)</td>
<td>NA</td>
</tr>
<tr>
<td>Mini-Cog by study staff, impaired</td>
<td>88/161 (55)</td>
<td>NA</td>
</tr>
<tr>
<td>CLOX1 by study staff, impaired</td>
<td>83/138 (60)</td>
<td>NA</td>
</tr>
<tr>
<td>CLOX2 by study staff, impaired</td>
<td>66/130 (51)</td>
<td>NA</td>
</tr>
<tr>
<td>Depression evaluation, n (%)</td>
<td>86/178 (48)</td>
<td>NA</td>
</tr>
<tr>
<td>PHQ-2 by emergency medical services, depressed</td>
<td>86/178 (48)</td>
<td>NA</td>
</tr>
<tr>
<td>PHQ-9, by study staff, minor or major depression</td>
<td>71/179 (40)</td>
<td>NA</td>
</tr>
<tr>
<td>Delirium evaluation, Confusion Assessment Method, by study staff, impaired, n (%)</td>
<td>16/175 (9.1)</td>
<td>NA</td>
</tr>
</tbody>
</table>

* P < .05.

NA = not applicable; PHQ = Patient Health Questionnaire.

The concurrent criterion validity of the prehospital case finding is also shown in Table 2. The kappas for depression show fair agreement, whereas the kappas for cognitive impairment show fair (SIS vs Mini-Cog) to slight (SIS vs CLOX1 or 2) agreement. When stratified for the presence of delirium, no statistically significant differences were found for cognitive impairment or depression screening (results not shown).

Construct validity is demonstrated in Table 3. It was found that the monotrait-monomethod correlations, which reflect the relationship between measures of the same trait, using the same method are among the greatest (0.48, 0.53), followed by the monotrait-heteromethod correlations, which reflect the relationship between measures of the same trait nor the same method (range -0.12–0.013). This pattern, in which the cognitive case finding results were more highly intercorrelated among themselves than with the depression screening results, and the depression screening results were more highly intercorrelated among themselves than with the cognitive screening results, provides evidence for the convergent and discriminant validity of the two screening tools. This is the order expected for a set of measures with construct validity.

Table 2. Reliability and Concurrent Criterion Validity of Depression and Cognitive Impairment Case Finding

<table>
<thead>
<tr>
<th>Evaluation</th>
<th>Domain</th>
<th>Instrument</th>
<th>Reference Standard</th>
<th>Sensitivity n/N; % (95% CI)</th>
<th>Specificity Kappa (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability testing</td>
<td>Cognitive impairment</td>
<td>SIS</td>
<td>NA</td>
<td>NA</td>
<td>0.52 (0.36–0.69)</td>
</tr>
<tr>
<td>Depression</td>
<td>PHQ-2*</td>
<td>NA</td>
<td>NA</td>
<td>0.50 (0.37–0.63)</td>
<td></td>
</tr>
<tr>
<td>Concurrent criterion validity testing</td>
<td>Cognitive impairment</td>
<td>EMS SIS</td>
<td>ED Mini-Cog</td>
<td>25/87; 29 (20–39)</td>
<td>70/73; 96 (88–99)</td>
</tr>
<tr>
<td></td>
<td>ED CLOX1</td>
<td>17/82; 21 (13–31)</td>
<td>51/55; 93 (82–98)</td>
<td>0.11 (0.017–0.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ED CLOX2</td>
<td>15/65; 23 (14–35)</td>
<td>59/64; 92 (83–97)</td>
<td>0.15 (0.029–0.27)</td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>EMS PHQ-2*</td>
<td>33/38; 87 (72–96)</td>
<td>85/133; 64 (55–72)</td>
<td>0.36 (0.24–0.48)</td>
<td></td>
</tr>
</tbody>
</table>

* Dichotomous scoring.

CI = confidence interval; NA = not applicable; SIS = Six-Item Screener; PHQ = Patient Health Questionnaire; EMS = emergency medical services; ED = emergency department.
The concurrent criterion validity testing for depression showed only fair performance of the prehospital PHQ-2. In particular, the prehospital PHQ-2 suffered from low specificity, which may have led to this unanticipated performance level. This study did not specifically explore the reasons for this result, but the low specificity may stem from the version of PHQ-2 used, which itself only has a specificity of 57%.22 The fact that trained study staff using the same instrument found similarly high levels of individuals screening positive (40% vs 48%) even though they found only 22% of individuals screening positive with the PHQ-9 supports this theory. It is possible that using the PHQ-2 with the interval scoring system rather than the dichotomous scoring system may result in better test performance, but the added difficulty of use may serve as a barrier to use.28 Future studies need to evaluate the performance of the PHQ-2 with the interval scoring system to better determine the reason for this fair performance and to identify ways to improve the validity of prehospital case finding for depression.

The construct validity testing for cognitive impairment showed only slight to fair performance of the SIS, with kappas ranging from 0.11 to 0.23, although evidence points to difficulties related to the chosen reference standards. In applying the Mini-Cog and CLOX, it became evident that these tests are inappropriate for the emergency setting. Almost one-quarter of subjects could not draw the clock faces because of limitations such as limited arm mobility or visual difficulties, thus potentially biasing the results. Additionally, for subjects who completed the test, the proportion with cognitive impairment was atypically high, lacking face validity. This study was not structured to evaluate this surprising finding, but distractions in the ED may have affected subjects’ ability to concentrate on the clock drawing, resulting in worse performance on the two instruments. Despite these challenges, the results are encouraging. Future studies need to evaluate the validity of the prehospital application of the SIS using instruments that are appropriate to the ED.

The construct validity, which found that the correlation coefficients of similar items were greater than those of different items, supports the hypotheses. One additional benefit of this approach is that it helps evaluate method variance. Especially problematic in the use of screening tools in such unorthodox environments is the confound of method variance—that the results reflect more of the variance due to the administration method of than the actual trait being measured. That the intercorrelations between the measures of cognitive function and depression within the ED and within the EMS setting were in the low range strongly suggests that the scores were not an artifact of the location and method of administration.

Despite the challenges encountered in this study, the results were encouraging. The moderate reliability between the prehospital- and study staff–applied PHQ-2 and SIS and the construct validity identified supports the continued development of the prehospital use of these instruments for case finding. The research agenda can move to validity testing, taking into account the lessons learned from the concurrent criterion validity testing that was performed. A diagnostic standard evaluation such as detailed neuropsychiatric testing for cognitive impairment and semistructured diagnostic interviews for depression to evaluate the validity of the EMS case finding activity would be appropriate. If those results prove to be accurate, then formal program development and effectiveness trials can be performed. In addition, case finding for other conditions can be considered, including for instance, vaccination status, fall risk, abuse, and medication management strategies.

A number of limitations were identified. First, the generalizability of these results is unknown, because this study took place with a small group of EMS providers who were motivated to participate, although the goal was to evaluate the reliability and criterion validity of prehospital implementation of these instruments, which relates primarily to the instruments. Second, the research staff in the ED performed the screening tests and the criterion standard evaluation such as detailed neuropsychiatric testing for cognitive impairment and semistructured diagnostic interviews for depression to evaluate the validity of the EMS case finding activity would be appropriate. If those results prove to be accurate, then formal program development and effectiveness trials can be performed. In addition, case finding for other conditions can be considered, including for instance, vaccination status, fall risk, abuse, and medication management strategies.

A number of limitations were identified. First, the generalizability of these results is unknown, because this study took place with a small group of EMS providers who were motivated to participate, although the goal was to evaluate the reliability and criterion validity of prehospital implementation of these instruments, which relates primarily to the instruments. Second, the research staff in the ED performed the screening tests and the criterion standard measurements. Thus, administration of one measure was done by individuals not blinded to another, potentially improving the concurrent criterion validity results. Third, a number of subjects did not consent or could not consent to participate in the study. Based on the limited data collected on these individuals, it is not expected that the bias would affect the reliability and validity results, although it may affect the proportion of older adults found to have needs. Finally, anxiety, which is interrelated with depression, was not evaluated. Future studies will also need to evaluate anxiety.

### Table 3. Multitrait-Multimethod Correlation Matrix Validation (N = 119)

<table>
<thead>
<tr>
<th>Cognitive Impairment</th>
<th>EMS SIS</th>
<th>ED SIS</th>
<th>Mini-Cog</th>
<th>CLOX1</th>
<th>CLOX2</th>
<th>Depression</th>
<th>EMS PHQ-2*</th>
<th>ED PHQ-2*</th>
<th>PHQ-9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive impairment</td>
<td>EMS SIS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED SIS</td>
<td>0.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mini-Cog</td>
<td>0.30</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLOX1</td>
<td>0.18</td>
<td>0.12</td>
<td>0.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLOX2</td>
<td>0.21</td>
<td>0.19</td>
<td>0.35</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>EMS PHQ-2*</td>
<td>-0.12</td>
<td>-0.060</td>
<td>-0.090</td>
<td>-0.12</td>
<td>-0.044</td>
<td>0.53</td>
<td>0.41</td>
<td>0.58</td>
</tr>
<tr>
<td>ED PHQ-2*</td>
<td>-0.067</td>
<td>-0.0091</td>
<td>0.081</td>
<td>0.069</td>
<td>0.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9 (minor or major depression)</td>
<td>-0.029</td>
<td>0.013</td>
<td>-0.083</td>
<td>-0.045</td>
<td>-0.053</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Dichotomous scoring.
EMS = emergency medical services; SIS = Six-Item Screener; PHQ = Patient Health Questionnaire; ED = emergency department.
CONCLUSION

Moderate test–retest reliability and construct validity was demonstrated for prehospital case finding by EMS providers for cognitive impairment using the SIS and for depression using the PHQ-2. Slight to fair concurrent criterion validity was found, a result that methodological limitations could explain. These findings provide additional support for the concept of using EMS providers to detect older adults at risk for these conditions. Further work is needed to confirm the validity and effectiveness of prehospital screening using these instruments before such programs are implemented.

ACKNOWLEDGMENTS

We would like to thank the EMS providers from Henrietta Volunteer Ambulance and Rural Metro Medical Services, Rochester, New York, for their assistance with this study.

Volunteer Ambulance and Rural Metro Medical Services, Rochester, New York, for their assistance with this study.

CONFLICT OF INTEREST: The editor in chief has reviewed the conflict of interest checklist provided by the authors and has determined that the authors have no financial or any other kind of personal conflicts with this paper.

Dr. Shah is supported by the Paul B. Beeson Career Development Award (NIA 1K23AG028942).

Author Contributions: Shah developed concept and design, acquired subjects, analyzed and interpreted, wrote manuscript. Karuza, Conwell, and Katz developed concept and design, interpreted data, edited manuscript. Rueckmann and Swanson acquired subjects, analyzed and interpreted data, edited manuscript.

Sponsor's Role: None.

REFERENCES


