Lea C. Watson, MD

Master's paper, Spring 2002
Health Care and Prevention
Public Health Leadership Program
UNC School of Public Health

A Review of Screening Accuracy for Geriatric Depression in Primary Care

Advisor signature: Joanne M. Garrett, PhD.

Second reader: Michael Pignone, MD, MPH
Abstract:

OBJECTIVE: To determine the accuracy of depression screening instruments for older adults in the primary care setting.

METHODS: Systematic review: databases employed were MEDLINE (search dates 1966-2001), and the Cochrane database on depression, anxiety and neurosis. We also searched the second Guide to Clinical Preventive Services, the 1993 AHCPR Clinical Practice Guideline on Depression and recent systematic reviews. Hand-checking of bibliographies and extensive peer review were also used to identify potential articles. Our pre-defined search strategy targeted only studies of adults aged 65 or greater in primary care or community settings, including long-term care. Articles were included in this review if they reported original data and tested depression screening instruments against a criterion standard, yielding sensitivity and specificity.

MAIN RESULTS: Seventeen articles met criteria and are included in this review, representing nine different screening instruments. The most commonly evaluated were the Geriatric Depression Scale, 30 and 15-item versions, the Center for Epidemiologic Studies-Depression scale, and the Self-Care D. There were minimal differences in the favorable performance of these three instruments, which had sensitivities ranging from 74-100% and specificities ranging from 53-98%.

CONCLUSIONS: Accurate and feasible screening instruments are available for detecting late-life depression in primary care. More research is needed to determine the accuracy of depression screening instruments in demented individuals, and in those with subthreshold depressive disorders.
Introduction:

Late-life depression, defined as depressive symptoms severe enough to affect social or occupational functioning in persons aged 65 years or older, is common and debilitating.(1-3) Fifteen to 20% of older community-dwellers report depressive symptoms severe enough to cause functional disability. (4-6) and 1-4% of this age group meet criteria for major depressive disorder.(1, 4, 5, 7) As many as 40% of nursing home patients suffer from depressive symptoms. (8) Depression in late-life is associated with reduced social functioning, added medical morbidity, and increased mortality.(9-12) (13)

Depression has a substantial economic and public health burden. Untreated episodes of depression in late life are associated with higher utilization of health services (2, 14). Unutzer et al. have shown that older medical outpatients who display depressive symptoms have substantially greater health care costs than those lacking depressive symptoms.(6) Depressive symptoms are also closely linked to suicidal behavior, and white males over 65 years are more likely to commit suicide than any other age group.(15)

Depression in late-life can be difficult to detect. Symptoms may arise in atypical fashion, the stigma of mental illness may be deeply entrenched in patients and their families, and physicians often fail to look for or detect depression. Simon and colleagues showed that only 35-43% of mixed-age patients with depression were identified by their primary care provider(14). Patients, families and physicians may assume that depression is a normal reaction to aging and loss.
When detected and treated, older patients with depression have improvement not only in symptoms and behavior, but also in cognitive functioning and overall quality of life. (7)

Several instruments have been developed to aid in identifying depression in older adults. We conducted a systematic review to determine the accuracy of these instruments for detecting unrecognized depression in the primary care setting.

Methods:

As a part of a broader review for the US Preventive Services Task Force and the Research Triangle Institute-University of North Carolina at Chapel Hill Evidenced Based Practice Center, we specifically prepared a strategy to identify articles relevant to the accuracy of depression screening instruments for older adults in the primary care setting. Databases employed were MEDLINE (search dates 1966-2001), and the Cochrane database on depression, anxiety and neurosis. We also searched the second Guide to Clinical Preventive Services,(16) the 1993 AHCPR Clinical Practice Guideline on Depression and recent systematic reviews. (17, 18). We also used hand-checking of bibliographies and extensive peer review to identify potential articles.

We used the search terms depression, depressive disorder, mass screening, sensitivity and specificity, reproducibility of results, primary health care, ambulatory care, family practice and the names of common screening and diagnostic instruments used to detect depression. Our search was limited to English language texts and to age greater than 65 years.

Eligibility for inclusion (table 1) required that articles report on screening in a primary care population using a criterion standard as comparison, and that
they provide information on diagnostic accuracy (usually sensitivity and specificity). The population could include studies performed in the community and in long-term care settings, but not in psychiatric facilities or clinics. We excluded studies that extracted briefer instruments from the parent version retrospectively; for example, if an investigator evaluated a 5-item version of the Geriatric Depression Scale (GDS) she must have defined the specific questions prior to administering the instrument versus extracting the five items based on post-hoc analyses. The criterion standards must have been commonly accepted structured or semi-structured diagnostic interviews or independent evaluations performed by psychiatrists based on DSM-IIIR, DSM-IV, ICD-10 or Research Diagnostic Criteria.

Both authors independently reviewed the abstracts generated from our searches. Discrepancies about eligibility were resolved by consensus after review of the entire article. For each included study, we extracted information about the screening instrument, the criterion standard, sensitivity and specificity, average age of participants, their dementia status, and the study setting.

Results:

Our initial search strategy yielded 1323 potential articles, (figure 1) 1270 of which could be eliminated by title review. Of the 56 articles remaining, 36 were eliminated after identifying exclusion criteria in the abstract or the manuscript: 17 because there was no criterion standard(19-35), 7 because the setting was not appropriate(36-42), 7 because the population was not geriatric(43-49), and 7 others with varying methodologic exclusions(50-55). Seventeen articles met our inclusion criteria and specifically examined the
performance of depression screening instruments for older adults in primary care (table 3).

The included studies were carried out among a wide spectrum of patients mostly in general practice settings, with the exception of one from a nursing home and one in the context of home care. Two studies specifically included patients with dementia. Of the remaining studies, 8 required patients to have mild or no dementia, and 8 did not test for cognitive impairment (table 3). Nine different instruments were used; most had 20 or fewer questions and were relatively easy to administer. Overall test performance in detecting major depression was similarly favorable among the instruments, with sensitivities ranging from 67% to 100% and specificities ranging from 53% to 98%.

The Geriatric Depression Scale (GDS), the Center for Epidemiologic Studies-Depression scale (CES-D), and the SELFCARE D were the most commonly evaluated screening instruments. The GDS has both a 30 and 15 item version and was designed in a yes/no format for self or caregiver administration, making it easy to use. (56, 57) It minimizes questions about somatic and vegetative symptoms, which can overlap with symptoms of concurrent medical illness. The GDS has been validated repeatedly in psychiatric settings. (37, 38, 56-58) Eight studies evaluated its use in primary care elderly, most using the 15 item version and a cutpoint of 3-5. (59-64) Sensitivity and specificity ranged from 79-100% and 67-80%, respectively.

The CES-D can be self-administered and lists 20 statements addressing depressive symptoms over the last week, asking the participant to rank the frequency of these feelings from “rarely” to “most of the time.” Its psychometric properties have been consistently strong in younger adults in the
community.(17) In the five studies that evaluated this instrument, cutpoints varied considerably, from 9-21.(65-69) The resultant sensitivities were 75-93%, with specificities ranging from 73-87%. One study also specifically evaluated the performance of the CES-D in mildly demented subjects with an average MMSE of 19, and showed similar test characteristics to the patients without dementia.(69) This instrument was perceived as generally easy to administer and feasible, except in a nursing home population where the questions had to be repeated multiple times.

The SELFCARE D is a self-administered instrument that requests responses to 12 items on a Likert scale, reflecting depressive symptoms over the last month. It was derived from a larger, previously validated instrument used commonly in England. (70) Bird and colleagues reported the initial study of its use in 1987 using an independent psychiatric assessment as the criterion standard in an outpatient clinic, showing a sensitivity of 77% and specificity of 98% with a cutpoint of 5. (70) Since then it has been validated using standardized structured interviews, both in general practice and in home care. (71, 72) Both settings revealed good sensitivities in the 90% range, but the specificity in home care was 53% versus 86% in general practice. Relative to the GDS and CES-D, this instrument has not been as extensively evaluated.

Papassotiropoulos et al. used the CES-D and the General Health Questionnaire (GHQ) to identify subthreshold depression in a community sample in Greece.(68) Accuracy was poor, with sensitivities below 50% and specificities of 75% and 72%, respectively. This is the only study we could identify that attempted to delineate subthreshold disorders. In an effort to address the potential cultural limitations of common instruments, Rait and
colleagues tested a screen specific to the growing contingent of African-Caribbeans in the United Kingdom (Carribean Culture Specific Screen). (73) They found that it performed well, but not better than the Brief Assessment Schedule Depression Cards or the GDS-15. Each had a sensitivity of 92%, with specificities ranging from 71-84%.(73)

Dementia poses barriers to effective screening for depression given the obvious limitations in self-report due to cognitive impairment. The Cornell Scale for Depression in Dementia was specifically designed for this population and calls for the clinician to use both patient and caregiver information to complete the screen. Most data generated about the Cornell Scale for Depression in Dementia (CSDD) have come from hospitalized patients, in whom it has demonstrated acceptable validity and reliability in demented and non-demented patients.(74-76) We identified only one study evaluating the CSDD. Vida et al. screened outpatients from a family medicine clinic and from a memory disorders clinic, and found a sensitivity of 90% and specificity of 75% for detecting major depression.(77)

Several very brief instruments have been validated in psychiatric or hospital settings where the prevalence of depressive symptoms is often high,(36, 40) but few have been tested in older primary care patients. Howe et al. attempted to validate a one question screen (MHI-1) derived from the mental health component of the SF-36, asking elderly participants, “in the past month, how much of the time have you felt downhearted or sad?” (1=none, 6=all the time) (78) They showed that as a “stand alone” screen, the MHI-1 did not perform well in the primary care setting, with a sensitivity of 67% and a specificity of 60%.
Discussion:

Our systematic review shows that several instruments demonstrate good accuracy for detecting late-life major depression in primary care. The GDS, CES-D and SELFCARE-D have comparable sensitivities and specificities and report adequate feasibility of administration. The CES-D and CCSD have similarly favorable accuracy in demented patients with an average MMSE score of 19. A one-question screen shows poor results, as does the one study using the GHQ and CES-D to detect subthreshold depression. Finally, Rait and colleagues demonstrate that a culturally specific screen in African-Caribbeans performs well, but no better than, the GDS.(73)

The GDS has longstanding success in identifying major depression in psychiatric and hospital settings and now demonstrates accuracy in primary care, where the 15-item version in its yes/no self-administered format represents a realistic tool for use in the community or the clinic. With a record of successful use in general adult research, the CES-D also has the benefit of a known track record and relative ease of administration. Evidence from this review suggests that it can be extended to the older primary care population. Although with no appreciable difference in accuracy from other instruments in our in our review, the SELFCARE-D has fewer data characterizing its use.

Our review highlights the need to further investigate the accuracy of screening tools for depression in patients with dementia, specifically where cognitive impairment may be severe. Using the CSDD, an instrument specifically designed for patients with dementia, Vida et al. found good accuracy for detecting depression; however, they studied patients with a mean MMSE score of 19.(77) The prevalence of depression in dementia is 15-40%(79-81). Given
the increasing incidence of dementia in our aging population, the availability of accurate screening tools that specifically account for the co-existence of these two common disorders will be important.

This review also reveals a notable lack of data in the area of screening accuracy for subthreshold depressive disorders. As the nosology for non-major depressive disorders evolves and treatment is shown to be effective, screening instruments should be targeted at identifying these syndromes. Lyness and colleagues showed that there is considerable functional disability in subsyndromal depression,(82) which is more prevalent than major depression. Others show similar findings, supporting the significant morbidity caused by depressive symptoms not severe enough to cross threshold for a major disorder.(4, 6)

Two recent reviews offer evidence that accurate screening instruments exist for detecting depression in general adult primary care populations.(17, 18) In her 1995 review of the general adult primary care population, Mulrow and colleagues showed that nine different instruments demonstrated good screening accuracy, and that none was superior based on psychometric properties.(17) A recent updated review for the U.S. Preventive Services Task Force also found that several brief, accurate screening tools were available.(18) The USPSTF review included, but did not independently evaluate, studies of late-life depression.

Late-life depressive disorders have a convincing burden of suffering, often go undetected, and have known effective treatments.(83) Our systematic review reveals that accurate and feasible screening instruments are available to detect depression in older primary care patients. Future work should include
tests of depression screening accuracy for demented populations, and for subthreshold depressive disorders.Investigators should also evaluate the accuracy of very short instruments, such as the 5-item version of the GDS,(64) in the primary care setting. Acceptable administration times and ease of use will likely determine the realistic application of proven instruments.
References:


Table 1. Inclusion/Exclusion criteria

<table>
<thead>
<tr>
<th>Category</th>
<th>Inclusion</th>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Databases</td>
<td>MEDLINE + Cochrane</td>
<td>Other databases</td>
</tr>
<tr>
<td>Languages</td>
<td>English only</td>
<td>Other languages</td>
</tr>
<tr>
<td>Populations</td>
<td>Humans only, age greater than 65</td>
<td>Animal studies</td>
</tr>
<tr>
<td></td>
<td>Primary care or community settings</td>
<td>Hospital settings</td>
</tr>
<tr>
<td></td>
<td>(including long-term care)</td>
<td>Psychiatry clinics</td>
</tr>
<tr>
<td>Study Design</td>
<td>Original data</td>
<td>Letters, editorials, and</td>
</tr>
<tr>
<td></td>
<td>Must have criterion standard</td>
<td>non-systematic reviews</td>
</tr>
<tr>
<td></td>
<td></td>
<td>that have no original</td>
</tr>
<tr>
<td></td>
<td></td>
<td>data</td>
</tr>
<tr>
<td>Publication Date</td>
<td>January 1966-September 2001</td>
<td></td>
</tr>
<tr>
<td>Outcomes of Interest</td>
<td>Sensitivity and specificity</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Selected articles for review

FROM INITIAL SEARCH
1323 articles
1270 articles eliminated after title review
53 articles
27 articles eliminated after abstract review
26 articles
9 articles eliminated after full text review
17 articles
FINAL INCLUDED ARTICLES
Table 2. Seventeen Articles Meeting Inclusion Criteria

<table>
<thead>
<tr>
<th>Author</th>
<th>Test/ Cutpoint</th>
<th>Sensitivity (%) (95% Conf. Intervals)</th>
<th>Specificity (%) (95% Conf. Intervals)</th>
<th>Criterion</th>
<th>Average Age</th>
<th>Dementia Setting</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>D’Ath et al., 1994 (61)</td>
<td>GDS-15/ 5</td>
<td>91 (86,96)</td>
<td>72 (66,78)</td>
<td>GMS/AGECAT</td>
<td>74</td>
<td>not tested</td>
<td>general</td>
</tr>
<tr>
<td>Gerety et al., 1994 (62)</td>
<td>GDS/ 11</td>
<td>89 (72,96)</td>
<td>68 (58,77)</td>
<td>SCID</td>
<td>79</td>
<td>avg.MMSE</td>
<td>nursing</td>
</tr>
<tr>
<td></td>
<td>CES-D/ 16</td>
<td>74 (55,86)</td>
<td>70 (60,79)</td>
<td></td>
<td>23</td>
<td></td>
<td>home</td>
</tr>
<tr>
<td>Neal and Baldwin, 1994 (63)</td>
<td>GDS/11</td>
<td>83 (72,94)</td>
<td>80 (68,92)</td>
<td>GMS/AGECAT</td>
<td>77</td>
<td>not tested</td>
<td>general</td>
</tr>
<tr>
<td>Van Marjwick et al., 1995 (64)</td>
<td>GDS/7</td>
<td>79 (76,82)</td>
<td>67 (63,71)</td>
<td>DIS</td>
<td>74</td>
<td>mild/none</td>
<td>general</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>practice</td>
</tr>
<tr>
<td>Study</td>
<td>Measure 1</td>
<td>Measure 2</td>
<td>Measure 3</td>
<td>Measure 4</td>
<td>Measure 5</td>
<td>Measure 6</td>
<td>Measure 7</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Arthur et al., 1999 (65)</td>
<td>GDS-15/3</td>
<td>100 (98,102)</td>
<td>72 (67,77)</td>
<td>ICD-10</td>
<td>80</td>
<td>none</td>
<td>general</td>
</tr>
<tr>
<td>Hoyl et al., 1999 (66)</td>
<td>GDS-15/5</td>
<td>94 (89,99)</td>
<td>82 (73,91)</td>
<td>SCID</td>
<td>75</td>
<td>avg. MMSE</td>
<td>frail</td>
</tr>
<tr>
<td>Rait et al., 1999 (75)</td>
<td>GDS-15/4</td>
<td>92 (64,100)</td>
<td>71 (63,79)</td>
<td>GMS/AGECAT</td>
<td>&gt;60</td>
<td>not tested</td>
<td>community</td>
</tr>
<tr>
<td></td>
<td>BASEDEC/6</td>
<td>92 (64,100)</td>
<td>84 (78,91)</td>
<td></td>
<td></td>
<td></td>
<td>survey of</td>
</tr>
<tr>
<td></td>
<td>CCSS/6</td>
<td>92 (64,100)</td>
<td>79 (71,86)</td>
<td></td>
<td></td>
<td></td>
<td>African-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Caribbeans</td>
</tr>
<tr>
<td>Beekman et al., 1997 (67)</td>
<td>CES-D/20</td>
<td>93 (91,95)</td>
<td>73 (69,77)</td>
<td>DIS</td>
<td>55-82</td>
<td>none</td>
<td>community</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>practice</td>
</tr>
<tr>
<td>Study</td>
<td>Scale</td>
<td>CES-D Score</td>
<td>GDS Score</td>
<td>Methodology</td>
<td>Age Cut-off</td>
<td>Sample Source</td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>---------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>Lewisohn et al., 1997 (68)</td>
<td>CES-D/12</td>
<td>76 (73,79)</td>
<td>77 (74,80)</td>
<td>RDC, DSM IIIR</td>
<td>64</td>
<td>not reported survey</td>
<td></td>
</tr>
<tr>
<td>Lyness et al., 1997 (69)</td>
<td>CES-D/21</td>
<td>92 (87,97)</td>
<td>87 (81,93)</td>
<td>SCID</td>
<td>71</td>
<td>not tested general practice</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GDS/10</td>
<td>100 (98,102)</td>
<td>84 (78,90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papassotiropoulos et al. 1999 (71)</td>
<td>CES-D/8</td>
<td>75 (70,80)</td>
<td>74 (67,81)</td>
<td>CIDI</td>
<td>&gt;60</td>
<td>avg. MMSE community survey 27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(demented excluded)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CES-D/9</td>
<td>75 (70,80)</td>
<td>72 (67,77)</td>
<td>CIDI</td>
<td>&gt;60</td>
<td>avg. MMSE community survey 19 in demented sample</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Tool</td>
<td>GHQ-12/0</td>
<td>CES-D/9</td>
<td>SELFCARED/5</td>
<td>SELFCARED/8</td>
<td>CIDI, DSM IIIR</td>
<td>GMS/AGECAT</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------</td>
<td>----------</td>
<td>----------</td>
<td>-------------</td>
<td>-------------</td>
<td>----------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Papassotiropoulos et al. 1999</td>
<td>GHQ-12/0</td>
<td>46 (40.52)</td>
<td>39 (33.45)</td>
<td>77 (67.87)</td>
<td>70 (70.80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bird et al., 1987</td>
<td>SELFCARED/5</td>
<td>77 (67.87)</td>
<td></td>
<td>77 (69.71)</td>
<td>98 (95.101)</td>
<td>Independent</td>
<td></td>
</tr>
<tr>
<td>Upadhyaya and Stanley, 1997</td>
<td>SELFCARED/5</td>
<td>95 (90.100)</td>
<td>86 (78.94)</td>
<td></td>
<td></td>
<td>GMS/AGECAT</td>
<td></td>
</tr>
<tr>
<td>Banerjee et al. 1998</td>
<td>SELFCARED/8</td>
<td>90 (86.94)</td>
<td>53 (46.60)</td>
<td></td>
<td></td>
<td>GMS/AGECAT</td>
<td>&gt;65</td>
</tr>
<tr>
<td>Source</td>
<td>Test</td>
<td>Code/Score</td>
<td>Control</td>
<td>Technique</td>
<td># Excluded</td>
<td>Population</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------</td>
<td>------------</td>
<td>---------</td>
<td>------------</td>
<td>------------</td>
<td>---------------------------</td>
<td></td>
</tr>
<tr>
<td>Howe et al., 2000(80)</td>
<td>MHI-1/2</td>
<td>67 (58,76)</td>
<td>60 (50,70)</td>
<td>GMS/AGECAT</td>
<td>81</td>
<td>excluded</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>community &quot;organic impairment&quot;</td>
<td></td>
</tr>
<tr>
<td>Vida et al., 1994(79)</td>
<td>Cornell Screen/7</td>
<td>90 (80,100)</td>
<td>75 (60,90)</td>
<td>RDC</td>
<td>72</td>
<td>Avg. MMSE 19 specialty clinic-demented sample</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. footnotes

GDS = Geriatric Depression Scale, 30 item
GDS-15 = Geriatric Depression Scale, 15 item
GHQ= General Health Questionnaire
DIS=Diagnostic Interview Schedule
BASEDEC= Brief Assessment Schedule Depression Cards
CES-D= Center for Epidemiologic Study-Depression
MHI-1= single question from the Mental Health Inventory: " in the past month, how much have you felt downhearted or sad (1: none-6: all the time)
GMS=Geriatric Mental State/ AGECAT computer program
CIDI= Composite International Diagnostic Interview
SCID=Structured Clinical Interview for DSM IIIR
CCSS= Caribbean Culture Specific Screen
RDC=Research Diagnostic Criteria
DSM IIIR=Diagnostic and Statistical Manual of Mental Disorders, 3rd edition revised
MMSE=Mini Mental State Examination
ICD-10=International Classification of Diseases, 10th edition