SYSTEMATIC REVIEW

Depression in older people in the general hospital: a systematic review of screening instruments

MICHAEL DENNIS1, ADIL KADRI2, JOHN COFFEY2

1College of Medicine, Swansea University, Grove Building, Swansea University Singleton Park, Swansea, Wales SA2 8PP, UK
2Old Age Psychiatry, ABM University Health Board, Swansea, Wales, UK

Address correspondence to: M. Dennis. Tel: (+44) 01792 602145; Fax: (+44) 01792 516579. Email: m.s.dennis@swansea.ac.uk

Abstract

Background: depression is common in later life, particularly in people with poor physical health. In the acute hospital setting this is associated with poor outcomes, increased length of stay and compromised care. The recognition and diagnosis of depression is therefore a key first step in managing depression in the general hospital, and this may be facilitated by the use of an appropriate screening instrument.

Objective: the aim of this study is to review all relevant literature on rating scales used to detect depression in older people in general hospitals so as to identify the most appropriate tool and cut-off score with optimal performance.

Method: an electronic search was conducted applying key search terms. Selection of articles was conducted in a staged manner and by utilising predetermined quality criteria. When appropriate pooled analysis was undertaken.

Findings: only 14 studies satisfied the inclusion criteria and only one instrument—the Geriatric Depression Scale (GDS)—has been studied to an adequate extent in older people in the acute general hospital setting. Best performance for the GDS was for a cut-off of 5/6 for the GDS-15 and 10/11 for the GDS-30.

Conclusion: further research is required before recommending the use of brief depression screening instruments (single or two items) in the acute hospital setting. The GDS would appear the most validated instrument currently (in either 15 or 30 item versions), though other tools such as the BASDEC show promise.

Keywords: depression, older people, screening, general hospital, elderly

Introduction

Depressive disorders are common in older people with prevalence in the community of around 12% [1]. Depression is even more common in older physically ill patients in general hospitals with estimates of 5–58% and a mean prevalence of 29% [2]. In addition, depression in later life is more heterogeneous than in younger people, not only in symptoms, but also aetiology and prognosis.

Identification of depression in older people in the general hospital will not only help to alleviate the distress from the condition, but could help to reduce the risk of suicide, and even have economic benefits for services in terms of reduced length of stay. There is widely accepted evidence to suggest that the presence of depression will adversely affect the outcome of a variety of medical conditions and is associated with increased mortality, affect compliance with treatment, and impair response to rehabilitation programmes [2]. Despite the significant prevalence of depression in hospitalised elderly patients, it continues to remain an undetected and undertreated illness [3]. The detection and diagnosis of depression in hospitalised elderly patients can pose particular problems. This is primarily because symptoms of depression, particularly somatic (such as loss of appetite, weight loss, decreased energy and fatigue, disturbed sleep) are similar to symptoms of physical illness. This is further complicated by the fact that the elderly frequently deny low mood, health professionals on
medical wards often feel too busy to take time to enquire about patients feelings and many doctors lack experience and skills in diagnosing depression.

In order to address the problem of poor detection rates of depression among the physically ill, the use of screening tests may be beneficial. The National Institute for Clinical Excellence (NICE) guidelines for the management of depression in primary and secondary care have highlighted the importance to screening at risk populations which would include the physically ill elderly in general hospitals [4]. There are already various screening instruments in common use in the general hospitals such as the Geriatric Depression Scale (GDS; [5]). These scales have usually been developed in community and primary care settings. This in itself poses its own challenges, the least of which are: do screening test for depression work accurately in this setting; which is the most appropriate instrument to use and what is the best cut-off score for the scale in the clinical setting of hospital in-patients?

The aim of this systematic review was to review all relevant literature on depression rating scales used to detect depression in older people in general hospitals with the purpose of trying to establish which is the most appropriate screening tool and the cut-off score providing the optimal performance.

Methods

Search strategy and selection criteria

We conducted an electronic search for articles published in three databases: Embase (from 1980 to June 2009); Medline (1950 to June 2009) and Psych Info (1806 to June 2009). There were 14 categories of search covering four main domains: depression, old age, hospitals and rating scales/screening instruments.

This initial search strategy yielded a total of 661 abstracts. For the second stage, a further set of criteria were devised comprising articles having a likelihood of the following:

- Studies that include subjects of whom some or all were in-patients in a hospital facility for non-psychiatric patients.
- Studies that comprise older people (defined as an age category commencing at 60 years or higher) or a sub-group of older people.
- The subjects had to be screened using some form of depression rating scale.
- The studies had to be published in the English language.

The abstracts were then independently checked against the above criteria by two researchers (A.K., J.C.) as suggested by Irwig et al. [6]. There were 102 abstracts of articles that appeared to satisfy these basic criteria and the full texts were requested and subsequently scrutinised. Only 53 of the articles were deemed to actually meet the second-stage criteria. The bibliographies of the 53 selected articles were checked for any further articles that may satisfying the above criteria and a further 18 papers were identified. An on-line search of particular key journals was also conducted for any further articles (Age and Ageing, International Journal of Geriatric Psychiatry, Journal of the American Geriatric Society, American Journal of Geriatric Psychiatry and International Psychogeriatrics).

The papers were then examined by three reviewers (J.C., A. K., M.D.) independently applying the final inclusion criteria described below. Any disagreements between the reviewers’ judgments were resolved by comprehensive discussion.

The final inclusion criteria for the review were as follows:

1. Studies had to comprise older people (defined as an age category commencing at 60 years or higher) or a clearly identifiable sub-group of older people.
2. The subjects had to comprise medical in-patients (in a general hospital or rehabilitation hospital facility not a psychiatric unit) or have a clearly identifiable sub-group of medical in-patients.
3. Each study had to utilise a depression rating scale to screen for depression and this had to be compared with quality external case criteria, i.e. ‘Gold standard’. The Gold Standard had to be one of the following: utilise recognised diagnostic criteria such as ICD-10 or DSM, GMS-AGECAT, RDC or involve a clinical interview and diagnosis by a psychiatrist.
4. There were a minimum of 10 depression cases according to the ‘Gold standard’. Less than 10 cases were considered insufficient for calculating sensitivity.
5. When data from the same study was published more than once, only one paper was selected.

A total of 12 papers satisfied these criteria from the 71 articles reviewed in detail. The key journal on-line search also identified a further two articles, resulting in a final 14 papers that satisfied the final strict inclusion criteria and were selected for the systematic review.

From the finally selected articles the following information was then identified and entered on to a standard pro-forma: full reference, sample size, sample setting, sample procedure, mean age, proportion of females, cognitive status of participants, depression screening instrument (and version), external case validation (‘gold standard’), sensitivity and specificity. To facilitate further analysis, we constructed two-by-two tables for each primary study (and for different cut-offs for various screening instruments) that categorised numbers of screening positive and negative persons who did or did not meet the ‘gold standard’ for depression.

Statistical analysis

In view of heterogeneity only when data were available from more than one study for the same cut-off point for an instrument was pooled analysis undertaken. Pooled analysis involved calculation of sensitivity, specificity, positive likelihood ratio (LR+), negative likelihood ratio (LR−), diagnostic odds ratio (DOR), and area under the curve (AUC) for the ROC curve when data were available from three or more studies.
Sensitivity of an instrument is the proportion of those with the condition that test positive, specificity refers to the proportion of those without the condition that test negative. LR+ = sensitivity/specificity, LR− = sensitivity/(1-specificity), DOR = sensitivity × specificity /[(1-sensitivity) × (1-specificity)]. Essentially, specificity and sensitivity are considered good above 0.8, and a test accurate if LR+ > 5, LR− < 0.3, and an instrument useful for clinical practice if DOR is greater than 20. Heterogeneity was presented as an I² statistic. The analysis was undertaken utilising Meta-Disc [7, 8] with a random effects model.

### Results

The fourteen articles that fully satisfied the strict inclusion criteria for the review appear in Table 1 [9–22]. In addition, there was one article that satisfied all criteria apart from having a minimum of 10 subjects diagnosed as depressed according to the external case validation [23], and further study excluded as the only a subset of patients screening above a threshold on the GDS were fully assessed by the psychiatrist [24]. Other reasons for exclusion of articles included the inability to extract data for in-patient populations from a more global clinical setting [25] or an evaluation in specialised in-patient settings such as a stroke rehabilitation [26].

From the 14 selected articles, there were six studies from the UK, five from the USA, two from Italy and one from Australia. The prevalence of depression in the study populations ranged widely from 8 [10] to 45% [16]. The clinical setting for the studies was usually an acute medical admission facility, or one that specialised in the care of acute medical illness for those aged 65 years and older (acute geriatric ward).

In most studies, patients with cognitive impairment were excluded. The study of Jackson and Baldwin [15] included patients with organic brain syndrome though they did include data for a sub-group of cognitively intact patients. However, definitions of cognitive impairment differed widely, some studies excluded confused patients on the basis of clinical impression, whereas others utilised screening instruments or diagnostic criteria. The most commonly used tools for identifying cognitive impairment were the Mini-Mental State Examination (MMSE) or Abbreviated Mental Test (AMT). The AMT cut-off was invariably 5 or less, but the MMSE cut-off varied from 15 or less [17, 18] through to 24 or less [19]. In most studies, the researcher assessing for the presence of depression defined as the gold standard was blind to the results for the depression screening instrument, though this was not the case in one study [10] or evident in two others [13, 15].

### Depression screening instruments

There were 13 different depression screening instruments examined in the 14 articles; many studies involved the valuation of more than one tool. By far the most frequently examined was the GDS, the full 30-item instrument [5] in 7

---

### Table 1. Description of all primary studies

<table>
<thead>
<tr>
<th>Authors</th>
<th>n</th>
<th>% Female</th>
<th>Mean age</th>
<th>Prevalence of depression (%)</th>
<th>Instrument(s)</th>
<th>Depression diagnosis</th>
<th>Diagnostic system</th>
<th>Research interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adshead et al. [9]</td>
<td>72</td>
<td>71</td>
<td>79</td>
<td>33</td>
<td>BASDEC, GDS-30</td>
<td>Depr, Dys</td>
<td>Clinical</td>
<td></td>
</tr>
<tr>
<td>Blank et al. [10]</td>
<td>150</td>
<td>51</td>
<td>80</td>
<td>8</td>
<td>CES-D 10, 20; GDS-15, GDS-30, Yale 1, PRIME-MD 2</td>
<td>MajD</td>
<td>DSM-IV</td>
<td>DIS</td>
</tr>
<tr>
<td>Cullum et al. [11]</td>
<td>221</td>
<td>60</td>
<td>80</td>
<td>18</td>
<td>GDS-15</td>
<td>Depr</td>
<td>ICD-10</td>
<td>GMS</td>
</tr>
<tr>
<td>Davies et al. [12]</td>
<td>100</td>
<td>64</td>
<td>82</td>
<td>23</td>
<td>HADS</td>
<td>Depr</td>
<td>AGE-CAT</td>
<td></td>
</tr>
<tr>
<td>Hammond et al. [14]</td>
<td>46</td>
<td>74</td>
<td>81</td>
<td>22</td>
<td>Brief Obs RS</td>
<td>Depr</td>
<td>AGE-CAT</td>
<td></td>
</tr>
<tr>
<td>Jackson and Baldwin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GDS-30, 30</td>
<td>Depr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kitchell et al. [16]</td>
<td>42</td>
<td>5</td>
<td>68</td>
<td>45</td>
<td>ZSDS, PID</td>
<td>MajD</td>
<td>DSM-III</td>
<td>Clinical</td>
</tr>
<tr>
<td>Koenig et al. [17]</td>
<td>128</td>
<td>12</td>
<td></td>
<td></td>
<td>GDS-30</td>
<td>MajD</td>
<td>DSM-III</td>
<td>+ DIS</td>
</tr>
<tr>
<td>Koenig et al. [18]</td>
<td>109</td>
<td>0</td>
<td>74</td>
<td>10</td>
<td>GDS-30, BCDRS</td>
<td>MajD</td>
<td>DSM-III</td>
<td>DIS</td>
</tr>
<tr>
<td>Liske et al. [19]</td>
<td>94</td>
<td>-</td>
<td>22</td>
<td>-</td>
<td>BASDEC, SCL-5</td>
<td>Depr</td>
<td>AGE-CAT</td>
<td></td>
</tr>
<tr>
<td>Magri et al. [20]</td>
<td>220</td>
<td>50</td>
<td>76</td>
<td>30</td>
<td>GDS-30, DFS</td>
<td>MajD, Dy</td>
<td>DSM-III</td>
<td>Clinical</td>
</tr>
<tr>
<td>Rapp et al. [21]</td>
<td>150</td>
<td>0</td>
<td>69</td>
<td>15</td>
<td>GDS-30, SADS, BDI</td>
<td>Depr*</td>
<td>RDC, SADS</td>
<td></td>
</tr>
<tr>
<td>Rinaldi et al. [22]</td>
<td>60</td>
<td>80</td>
<td>29</td>
<td>-</td>
<td>GDS-15, GDS-5</td>
<td>Depr*</td>
<td>DSM-IV, Clinical</td>
<td></td>
</tr>
</tbody>
</table>

Empty cells: information not available or insufficient.

Instrument: BASDEC, Brief Assessment Schedule Depression Cards; GDS, Geriatric Depression Scale 5, 15, 30-item versions; HADS, Hospital Anxiety and Depression Scale; ELDRS, Evans Liverpool Depression Rating Scale; Brief Obs RS, Brief Observer Rating Scale; ZSDS, Zung Self-Rating depression scale; PID, Popoff Index of Depression; BCDRS, Brief Carroll Depression Rating Scale; SCL, Symptom checklist; DFS, depression factor score of SCL-90; SADS, self-report depression scale; CES-D, Center for Epidemiological Studies Depression Scale; BDI, Beck Depression Inventory.

Depression diagnosis: Depr, depression; MajD, major depression; MinD, minor depression; Dys, dysthymia; Depr*, clinically significant depression in accordance with diagnostic criteria.


Research interview: GMS, Geriatric Mental State, DIS, diagnostic interview schedule.
External case validation: ‘gold standard’

The main diagnostic systems were DSM (III, IIIR or IV), and AGECAT; see Table 1. Frequently the diagnosis were made following structured or semi-structured interviews: the Geriatric Mental State for an AGECAT diagnosis (commonly used in UK-based studies); and the Diagnostic Interview Schedule for DSM. A clinical psychiatric interview was also used in five studies. Most studies used the diagnosis of major depression (DSM) or AGECAT depression of level 3 or above (corresponding to case of depressive illness). In some studies results were presented for a combination of major depression with mild but clinically relevant depressive syndromes (such as depression not otherwise specified, minor depression, dysthymia).

Instrument performance

**GDS-30**

Table 2 presents the pooled analysis for the GDS-30. Best performance was achieved with either a cut-off of 10/11 (maximising sensitivity while still achieving adequate LR+, LR−, DOR) or 13/14 with high specificity and adequate performance on other indices.

**GDS-15**

The pooled results for the GDS-15 (at cut-off points 4/5 and 5/6) are presented in Table 2. Cullum et al. (2006) found in their study that the optimum cut-off point for diagnosing depression in older people in the general hospital was 6/7—this gave a sensitivity of 74 (61–88), specificity of 81 (76–87), LR+ 3.98 (2.79–5.68), LR− 0.32 (0.18–0.54), DOR 12.62 (2.40–22.84).

Shorter versions of the GDS have also been evaluated in a variety of clinical settings, including the GDS-5. The GDS-5 was evaluated by Rinaldi et al. [22] in 60 hospitalized elderly people with a good performance employing a cut-off of 1/2: sensitivity 97 (90–100); specificity 74 (59–90); LR+ 3.74 (2.05–6.82), LR− 0.05 (0.01–0.32) and DOR 80.5 (92.64–253.64).

**BASDEC**

Both studies [9, 19] used a cut-off of 6/7 for the BASDEC allowing a pooled analysis to be undertaken. The sensitivity was 0.80 (0.66, 0.91, I² 68%), specificity 0.86 (0.78–0.92, I² 0%), LR+ 5.86 (3.71–9.25, I² 0%), LR− 0.22 (0.07–0.71, I² 64%) and DOR 27.48 (8.75–86.25, I² 7%).

**Other instruments examined only in single studies**

The other screening instruments evaluated in only one study that satisfied inclusion criteria for the review were: the HADS, Hospital Anxiety and Depression Scale; ELDRS, Evans Liverpool Depression Rating Scale; ZSDS, Zung Self-Rating depression scale; PID, Popoff Index of Depression; SCL, Symptom checklist; DFS, Depression factor score of SCL-90; SDS, Self-report depression scale; CES-D, Center for Epidemiological Studies Depression Scale and PRIME-MD 2. Also included here was the Brief Carroll Depression Rating Scale as in the study by Koenig.

### Table 2. Pooled analysis for Geriatric Depression Scale (GDS)

<table>
<thead>
<tr>
<th>Version (cut-off)</th>
<th>Sens % (95% CI)</th>
<th>Spec % (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR− (95% CI)</th>
<th>DOR (95% CI)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDS-15 (5/6)</td>
<td>79 (70–86%)</td>
<td>30 (77–81%)</td>
<td>0.29 (0.18–0.46)</td>
<td>10.88</td>
<td>0.29 (0.18–0.46)</td>
<td>11.08</td>
</tr>
<tr>
<td>GDS-15 (4/5)</td>
<td>88 (77–95%)</td>
<td>0 (64–77%)</td>
<td>0.18 (0.09–0.37)</td>
<td>13.61</td>
<td>0.18 (0.09–0.37)</td>
<td>10.83</td>
</tr>
<tr>
<td>GDS-30 (9/10)</td>
<td>80 (68–90%)</td>
<td>37 (81–76%)</td>
<td>0.28 (0.17–0.47)</td>
<td>17.83</td>
<td>0.28 (0.17–0.47)</td>
<td>10.88</td>
</tr>
<tr>
<td>GDS-30 (10/11)</td>
<td>85 (78–91%)</td>
<td>28 (82–78%)</td>
<td>0.21 (0.13–0.33)</td>
<td>23.10</td>
<td>0.21 (0.13–0.33)</td>
<td>10.83</td>
</tr>
<tr>
<td>GDS-30 (11/12)</td>
<td>75 (60–87%)</td>
<td>90 (90–84%)</td>
<td>0.29 (0.18–0.49)</td>
<td>23.55</td>
<td>0.29 (0.18–0.49)</td>
<td>10.88</td>
</tr>
<tr>
<td>GDS-30 (12/13)</td>
<td>66 (50–80%)</td>
<td>91 (85–95%)</td>
<td>0.39 (0.26–0.59)</td>
<td>18.05</td>
<td>0.39 (0.26–0.59)</td>
<td>10.83</td>
</tr>
<tr>
<td>GDS-30 (13/14)</td>
<td>72 (63–80%)</td>
<td>92 (88–93%)</td>
<td>0.33 (0.24–0.46)</td>
<td>23.50</td>
<td>0.33 (0.24–0.46)</td>
<td>10.83</td>
</tr>
</tbody>
</table>

LR+, positive likelihood ratio.  
LR−, negative likelihood ratio.  
DOR, diagnostic odds ratio.  
AUC, area under ROC curve.  
I², test of heterogeneity.
GDS-30, the best screening performance appears to be of only 12.4. Sensitivity can be improved by lowering the ——adequate sensitivity and specificity but with a DOR of only 12.4. Sensitivity can be improved by lowering the cut-off to 4/5 though with the inevitable cost of lowered specificity but retaining an adequate DOR. With the GDS-30, the best screening performance appears to be with a cut-off of 10/11 (maximum sensitivity while maintaining otherwise adequate LR+, LR− and DOR), though diagnostically it performs better with a higher cut-off (i.e. 13/14). With the exception of the BASDEC (two studies), all other instruments have only been validated in single reports with corresponding wide confidence intervals. It is difficult, therefore, to draw conclusions about the performance of instruments other than the GDS in this clinical setting, though the BASDEC, CES-D, Brief Observer Rating Scale, Cognitive-affective subscale of the BDI and Evans Liverpool Depression Scale have all performed well in single study validation.

In the UK, the NICE has advocated screening populations at high risk of depression [4], and those with chronic physical ill health [28]. NICE [4] originally suggested two simple screening questions; the ‘Whooley questions’ (‘During the past month have you often been bothered by feeling down, depressed or hopeless?’ and During the last month have you been bothered by having little interest or pleasure in doing things?’; [29]). NICE have also suggested the two-item version of the patient health questionnaire (PHQ-2), as an initial screen for depression in people with chronic physical ill-health [28]. The Whooley questions and PHQ-2 share similar origins and are the same questions with either a dichotomised response (yes/no in the Whooley questions) or likert scale (PHQ-2). In our review; we found no evaluation of the PHQ-2 in older people in the acute general hospital setting, and although the PRIME MD-2, which is essentially the same as the Whooley questions showed good sensitivity the specificity was poor [10]. The Yale single question showed promise in one study [10]. Pomeroy et al. have, however, shown that very short scales (4-item GDS, and the 1-item of the mental health inventory) performed adequately in elderly people in rehabilitation wards and day hospital setting [25].

<table>
<thead>
<tr>
<th>Instrument (cut-off)</th>
<th>Sensitivity % (95% CI)</th>
<th>Specificity % (95% CI)</th>
<th>Positive likelihood ratio (95% CI)</th>
<th>Negative likelihood ratio (95% CI)</th>
<th>Diagnostic odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDI (15/16)</td>
<td>70 (51–88)</td>
<td>87 (81–93)</td>
<td>5.20 (3.10–8.73)</td>
<td>0.35 (0.19–0.65)</td>
<td>14.79 (8–29.94)</td>
</tr>
<tr>
<td>BDI cog-off subscale (4/5)</td>
<td>74 (56–92)</td>
<td>92 (87–97)</td>
<td>9.39 (4.94–17.85)</td>
<td>0.28 (0.14–0.56)</td>
<td>33.15 (4–70.70)</td>
</tr>
<tr>
<td>SDS (49/50)</td>
<td>83 (67–98)</td>
<td>65 (57–74)</td>
<td>2.38 (1.76–3.23)</td>
<td>0.27 (0.11–0.65)</td>
<td>8.96 (1–19.16)</td>
</tr>
<tr>
<td>ELDRS (5/6)</td>
<td>88 (76–99)</td>
<td>96 (90–100)</td>
<td>19.54 (6.42–59.52)</td>
<td>0.13 (0.05–0.33)</td>
<td>149.33 (83.84–382.51)</td>
</tr>
<tr>
<td>Brief Obs RS (2/3)</td>
<td>90 (71–100)</td>
<td>72 (58–87)</td>
<td>3.24 (1.84–5.71)</td>
<td>0.14 (0.02–0.90)</td>
<td>23.4 (2.78–74.67)</td>
</tr>
<tr>
<td>ZSDS (54/55)</td>
<td>74 (54–94)</td>
<td>74 (56–92)</td>
<td>2.85 (1.35–5.92)</td>
<td>0.36 (0.16–0.79)</td>
<td>7.93 (3.03–18.89)</td>
</tr>
<tr>
<td>PID (11/12)</td>
<td>83 (66–100)</td>
<td>57 (36–77)</td>
<td>1.92 (1.15–3.19)</td>
<td>0.30 (0.10–0.88)</td>
<td>6.6 (3.18–16.18)</td>
</tr>
<tr>
<td>BCDRS (5/6)</td>
<td>73 (43–90)</td>
<td>79 (70–86)</td>
<td>3.39 (2.01–5.73)</td>
<td>0.35 (0.13–0.92)</td>
<td>9.78 (4.03–23.58)</td>
</tr>
<tr>
<td>SCL-5 (9/10)</td>
<td>77 (60–95)</td>
<td>74 (63–84)</td>
<td>2.93 (1.87–4.60)</td>
<td>0.31 (0.14–0.68)</td>
<td>9.48 (1.20–20.17)</td>
</tr>
<tr>
<td>DFS (1)</td>
<td>87 (78–95)</td>
<td>75 (68–81)</td>
<td>3.40 (2.55–4.52)</td>
<td>0.18 (0.10–0.33)</td>
<td>18.83 (3.94–37.33)</td>
</tr>
<tr>
<td>HADS (10/11)</td>
<td>30 (15–44)</td>
<td>89 (81–97)</td>
<td>2.68 (1.14–6.3)</td>
<td>0.79 (0.63–0.99)</td>
<td>3.39 (0.19–6.96)</td>
</tr>
<tr>
<td>CES-D 10 (3/4)</td>
<td>92 (76–100)</td>
<td>77 (70–84)</td>
<td>3.95 (2.79–5.60)</td>
<td>0.10 (0.02–0.71)</td>
<td>36.44 (39.53–112.41)</td>
</tr>
<tr>
<td>PRIME-MD 2 (0/1)</td>
<td>92 (76–100)</td>
<td>54 (46–63)</td>
<td>2.00 (1.57–2.58)</td>
<td>0.15 (0.02–1.00)</td>
<td>13.10 (14.10–40.26)</td>
</tr>
<tr>
<td>Yale-1 (1)</td>
<td>83 (62–97)</td>
<td>83 (77–90)</td>
<td>5.00 (3.19–7.85)</td>
<td>0.2 (0.06–0.71)</td>
<td>25 (14.57–64.57)</td>
</tr>
</tbody>
</table>

Instrument: HADS, Hospital Anxiety and Depression Scale; ELDRS, Evans Liverpool Depression Rating Scale; Brief Obs RS, Brief Observer Rating Scale; ZSDS, Zung Self-Rating depression scale; PID, Popoff Index of Depression; BCDRS, Brief Carroll Depression Rating Scale; SCL, Symptom checklist; DFS, Depression factor score of SCL-90; SDS, Self-report depression scale; CES-D, Center for Epidemiological Studies Depression Scale; BDI, Beck Depression Inventory.
Screening for depression in the general hospital

Conflicts of interest
None declared.

References

Key points
• Depression is common in older people in the general hospital setting and associated with adverse health outcomes.
• Screening for depression in people at high risk of depression and with chronic physical health difficulties has been widely advocated.
• This systematic review highlights the paucity of quality research for brief screening instruments for depression in older people in acute medical care.
• Only the GDS has been widely evaluated in this setting.

There are limitations to our review, in particular some validity studies may not be published as consequence of publication bias. Despite careful steps to reduce heterogeneity this remains an issue in particular with varying exclusion criteria for cognitive impairment and differing diagnostic classifications for clinically important depression. Although moderate to severe cognitive impairment was invariably excluded, studies often included patients with milder impairment; this may have a confounding effect in particular in the presence of delirium when depressive symptoms are common [30]. It is also important to highlight that there remains a debate concerning the efficacy of screening for depression unless there is enhanced care for depression and improvement of other aspects of patient care—this is clearly a very important issue in both acute and rehabilitation medicine and a clear stimulus for the continuing development of adequately resourced liaison psychiatry. Additionally, the screening instruments vary in their balance of somatic/biological and cognitive/affective symptoms. Instruments with a strong cognitive/affective bias (such as the GDS and Evans Liverpool Depression Rating Scale) avoid the dilemma of dual causality for somatic symptoms in the physically ill but this may actually mean that a different category of depression is more likely to be identified than with instruments retaining a balance of depression symptoms (such as the BASDEC). These issues may ultimately have implications concerning treatment delivery and development.

The conclusions of this review are that only the GDS has been thoroughly evaluated as a screening instrument for depression in older people in the acute medical inpatient setting. Further research is needed before other screening instruments that have only been evaluated in single studies are recommended. In particular, shorter instruments advocated by recent guidance (such as the PHQ-2 and Whooley two questions) should be carefully evaluated before being widely adopted in clinical practice. In addition, the use of care pathways utilising brief screening instruments followed by a more detailed questionnaire as recommended by NICE [28] also requires examination.

Received 4 February 2011; accepted in revised form 21 July 2011