Research

Nicholas Meader, Alex J Mitchell, Carolyn Chew-Graham, David Goldberg, Maria Rizzo, Victoria Bird, David Kessler, Jon Packham, Mark Haddad and Stephen Pilling

Case identification of depression in patients with chronic physical health problems: a diagnostic accuracy meta-analysis of 113 studies

Abstract

Background
Depression is more likely in patients with chronic physical illness, and is associated with increased rates of disability and mortality. Effective treatment of depression may reduce morbidity and mortality. The use of two stem questions for case finding in diabetes and coronary heart disease is advocated in the Quality and Outcomes Framework, and has become normalised into primary care.

Aim
To define the most effective tool for use in consultations to detect depression in people with chronic physical illness.

Design
Meta-analysis.

Method
The following data sources were searched: CENTRAL, CINAHL, Embase, HMIC, MEDLINE, PsychINFO, Web of Knowledge, from inception to July 2009. Three authors selected studies that examined identification tools and used an interview-based ICD (International Classification of Diseases) or DSM (Diagnostic and statistical Manual of Mental Disorders) diagnosis of depression as reference standard. At least two authors independently extracted study characteristics and outcome data and assessed methodological quality.

Results
A total of 113 studies met the eligibility criteria, providing data on 20,826 participants. It was found that two stem questions, PHQ-9 (Patient Health Questionnaire), the Zung, and GHQ-28 (General Health Questionnaire) were the optimal measures for case identification, but no method was sufficiently accurate to recommend as a definitive case-finding tool. Limitations were the moderate-to-high heterogeneity for most scales and the facts that few studies used ICD diagnoses as the reference standard, and that a variety of methods were used to determine DSM diagnoses.

Conclusion
Assessing both validity and ease of use, the two stem questions are the preferred method. However, clinicians should not rely on the two questions approach alone, but should be confident to engage in a more detailed clinical assessment of patients who score positively.

Keywords
depression; diagnosis; meta-analysis; primary care

INTRODUCTION
Depression is one of the leading causes of disability and disease burden. It is associated with the most years lost to disability of all diseases worldwide. Identifying depression in patients with chronic physical health problems is important for several reasons. First, a number of studies suggest depression is approximately two to three times as prevalent in such populations, including patients with cancer, chronic heart disease, and chronic obstructive pulmonary disease (COPD). Secondly, there appears to be greater disease burden, in terms of healthcare use and functional disability, in people with comorbid depression compared with those with physical health problems alone. Thirdly, mortality is greater in several medical conditions when depression is present — heart disease, COPD, stroke, cancer — and in medically ill older adults. Furthermore, morbidity and mortality may diminish with effective treatment of depression.

There is convincing evidence that many cases of depression go unrecognised in the general population and in primary care.

N Meader, PhD, systematic reviewer; M Rizzo, MSc, research assistant; V Bird, BSc, research assistant, National Collaborating Centre for Mental Health, London. AJ Mitchell, MD, in liaison psychiatry, Leicestershire Partnership Trust, and honorary senior lecturer in Psycho- Oncology, Department of Cancer Studies and Molecular Medicine, University of Leicester, Leicester. C Chew-Graham, MD, professor of primary care, University of Manchester, Manchester. D Goldberg, DM, FRCP, professor emeritus, M Haddad, PhD, RGN, RMN, clinical research fellow, Health Service and Population Research Department, Institute of Psychiatry, King’s College, London. D Kessler, MD, Watport Clinical Lecturer — primary care, University of Bristol, Bristol. J Packham, DM, consultant rheumatologist, Haywood Hospital and Senior Lecturer, Arthritis Research UK Primary Care Centre, Keele University, Keele. S Pilling, PhD, joint director, National Collaborating Centre for Mental Health, director of Centre for Outcomes Research and Effectiveness, University College London, and professor of clinical psychology and clinical effectiveness, UCL Research Department of Clinical, Educational and Health Psychology, London.

Address for correspondence
Nicholas Meader, Centre for Reviews and Dissemination, University of York, York, YO10 5DD.

E-mail: rmeader@hotmail.com


©British Journal of General Practice

This is the full-length article (published online 28 Nov 2011) of an abridged version published in print. Cite this article as: Br J Gen Pract 2011; D0i: 10.3399/bjgp11X613151.
How this fits in

There is strong evidence that the prevalence of depression is raised among patients with long-term conditions and that this comorbidity is associated with adverse outcomes. Inadequate and inaccurate identification of depression has been documented in both primary care and general medical settings. This meta-analysis provides evidence that several brief and feasible depression case-finding approaches can be used as a first assessment for patients with chronic physical health problems, and that two stem questions referring to core depression features appear the most efficient initial approach.

guidelines. However, there is much debate in the literature concerning the effectiveness of screening and case identification. Gilbody and colleagues have shown untargeted screening was not effective in improving the recognition of depression in primary care and general hospital settings. There is also much debate concerning the terminology used in the field. The present study proposes to separate overall accuracy (case identification) into more clinically understandable rule-in and rule-out performance. Rule-in accuracy (positive predictive value) is the ability to correctly identify those with the disorder with minimal false positives, whereas rule-out accuracy (negative predictive value) is the ability to correctly identify those without the disorder with minimal false negatives (missed cases). In order to differentiate from untargeted screening approaches, which appear to be ineffective, this data synthesis will focus on case identification in a population at higher risk of depression (that is, people with chronic physical health problems). This is vital before further case finding is advocated by the QOF for patients with other physical problems.

There are a large number of scales used both in clinical practice and in research studies, few of which have been originally developed for the physically ill. In addition, there are no existing definitive meta-analyses across a comprehensive range of measures. Therefore, a diagnostic accuracy meta-analysis was conducted to assess the sensitivity and specificity of the most widely used case-identification instruments in people who are physically ill.

METHOD

Data sources and searches

The full review protocol can be found in the guideline on depression in people with chronic physical health problems, which was commissioned by NICE. Briefly, a search for studies assessing the validity of case-identification instruments was made using seven electronic bibliographic databases (CENTRAL, CINAHL, Embase, HMIC, MEDLINE, PsycINFO, Web of Knowledge). Each database was searched from inception to October 2009. Additional papers were found by searching the references of retrieved articles, tables of contents of relevant journals, previous systematic reviews and meta-analyses of case identification for depression, written requests to experts, and suggestions made by the members of the Guideline Development Group (comprising clinicians, academics, and service users with expertise in depression and chronic physical health problems).

Study selection

The study included validation studies of mood questionnaires agreed by the authors [see Appendix 1 for further details]. The reference standard was diagnoses according to the Diagnostic and Statistical Manual of Mental Disorders (DSM) of the American Psychiatric Association (for example DSM-IV) or International Classification of Diseases (ICD) (for example ICD-10) of the World Health Organization criteria. Studies that did not clearly state the comparator to be DSM or ICD diagnosis of depression, or that did not provide sufficient data to be extracted in the meta-analysis were excluded.

Data extraction and quality assessment

All published studies that met the eligibility criteria were assessed for methodological quality using the Scottish Intercollegiate Guidelines Network (SIGN) checklist for diagnostic studies. Data were extracted independently by at least two authors, and 2×2 tables were constructed, from which the primary outcomes were calculated: that is, sensitivity, specificity, and likelihood ratios. To maximise the available data, the most consistently reported and recommended cut-off points were extracted for each of the scales. There are limitations to this approach, as noted by Furukawa and colleagues, who found that the optimal cut-offs for the General Health Questionnaire (GHQ)-12 and GHQ-28 differed according to the prevalence of depression, and it is likely there are similar problems for most other scales. However, a Bayesian approach makes allowance for variations according to prevalence [see
Data synthesis and analysis
A bivariate diagnostic accuracy meta-analysis was conducted using Stata (version 10) with the metandi commands, to obtain pooled estimates of sensitivity, specificity, and likelihood ratios. This method was originally developed as a mixed effects regression model for meta-analysis of trials, and modified more recently for studies of diagnostic accuracy. Between-study heterogeneity was assessed using the I² statistic. In addition, publication bias was assessed by visual inspection of funnel plots, and formal use of Egger’s test.

A Bayesian curve analysis was also undertaken; this plots post-test conditional probabilities from all possible pre-test probabilities (prevalence). The area under the Bayesian curve (AUC) for positive results can be used as a statistical comparison of rule-in success and 1 — AUC for negative results used as an indicator of rule-out success. An area of more than 0.75 can be interpreted as ‘satisfactory’ and more than 0.80 interpreted as ‘good’. If a test achieved more than 0.90 in a rule-in capacity, this was considered sufficient for a recommendation that this tool could be used on its own for case finding.

Additional meta-regression analyses were planned to assess differences in diagnostic accuracy for disease groups. Such analyses were conducted on a scale when there were a minimum of four studies for at least two disease groups.

RESULTS
A total of 113 studies on 20,826 participants met the eligibility criteria of the review (see Figure 1 for full details on study flow information). These studies were both on populations specifically targeted for a chronic physical health problem (such as cancer, heart disease, and stroke), and in general medical settings where all were physically ill and a substantial proportion had a chronic physical health problem. In total, 83 studies specifically targeted people with chronic physical health problems in any setting [Appendix 2]. The mean prevalence of depression was 0.25 (95% confidence interval [CI] = 0.05 to 0.61). A further 30 studies were on people in general medical settings, with a mean prevalence of depression of 0.24 (95% CI = 0.04 to 0.52).

Studies recruiting for chronic physical health problem
Sensitivity and specificity: Table 1 provides an evidence summary for the various scales on people recruited for specific chronic physical health problems. There was moderate to high sensitivity for most scales. The tools with the highest sensitivity were the two stem questions (0.98; 95% CI = 0.85 to 0.99), followed by the GHQ-28, Patient Health Questionnaire (PHQ)-9, Beck Depression Inventory (BDI), and BD non-somatic (Table 1). Sensitivity was lowest for the one-item measure.

The Zung Self Rating Depression Scale had the highest specificity 0.92 (95% CI = 0.68 to 0.98). This was followed by the two stem questions, the Hamilton Depression Rating Scale (HDRS), PHQ-9 and the Centre for Epidemiologic Studies Depression Scale (CES-D); all had high specificity. The lowest specificity was found for the one-item measure and the GHQ-12.

Rule-in (positive predictive value) and rule-out accuracy (negative predictive value). Using Bayesian plots of conditional probabilities to examine rule-in and rule-out performance, only three tools had less than satisfactory rule-in performance, namely the single question: the Geriatric Depression Scale (GDS-30) and GHQ-12. The optimal single tool was the Zung, although it did not reach the a priori standard for recommendation when applied alone. For rule-out performance, four methods were not satisfactory. These were the single question, the Hospital Anxiety and Depression Scale (HADS), GDS-30, and GHQ-12. The optimal tools were the two stem questions and GHQ-28. Overall accuracy was best for the two stem questions, Zung, PHQ-9, and GHQ-28. However, it should be noted that data for the Zung scale were based on just four studies and a relatively small total sample size (n = 190).
Meta-regression comparing the diagnostic accuracy for different disease groups was only possible for the BDI and HADS-D. There was no evidence of difference in sensitivity (beta = 0.93, P = 0.34) and specificity (beta = 1.56, P = 0.35) of the HADS between stroke and cancer patients. There was no evidence of difference in sensitivity (beta = 1.49, P = 0.60), but some evidence for differences in specificity (beta = 1.20, P = 0.02) of the BDI between heart disease and cancer patients.

Studies in general medical settings

Table 2 summarises the results for general medical settings. There were only three scales that provided sufficient data for analyses. All these scales performed equally well in this setting as compared to populations specifically targeted for chronic physical health problems with a large overlap in confidence intervals.

**Sensitivity and specificity.** Sensitivity was relatively high in all measures but particularly high in the GDS-15 (0.89, 95% CI = 0.84 to 0.92). Specificity was very similar for the GDS-30, GDS-15, and HADS when used in general medical settings (Table 2).

**Rule-in and rule-out accuracy.** Using the same methodology for each measure in general medical settings and correcting for prevalence using a Bayesian analysis, the GDS-15 was most successful and the HADS least successful. No method came close to the a priori standard for rule-in performance when applied alone. For rule-out accuracy, the HADS was significantly less accurate than the GDS-15 (Area HADS = 0.71, 95% CI = 0.68 to 0.74 versus Area GDS-15 = 0.78, 95% CI = 0.75 to 0.82).

**DISCUSSION**

Most of the scales performed adequately as case-identification measures for depression, with modest differences in validity coefficients. Most studies targeted chronic physically ill populations rather than general medical settings such as primary care. In order to detect depression in those with chronic physically ill health, the most sensitive instruments appear to be two stem questions, PHQ-9, and GHQ-28. The most specific measure was the Zung. Overall, optimal accuracy was achieved by the two stem questions, Zung, PHQ-9, and GHQ-28. However, it should be noted that estimates on the Zung and GHQ-28 analysis were based on a relatively small sample size; therefore, it is possible that conclusions regarding these scales may change with further data. No method came close to the a priori standard for case-finding recommendation when applied alone.

Another important factor to consider when comparing the different measures is the ease of implementation. The Zung is a 20-item scale and therefore is more resource intensive and less likely to be implemented in primary care compared to shorter measures. Taking into account both the psychometric properties and ease of implementation, it would appear the two stem questions may be the preferred measure for case identification in patients with chronic physical health problems. From these data, the authors do not recommend relying upon a single question
In general medical settings, there were fewer studies, and analysable data were only available for the GDS, GDS-15, and HADS-D. Specificity was similar for all three scales but sensitivity was highest in the GDS-15. Further research is needed to confirm whether the optimal tools in the chronically ill (two stem questions, PHQ-9, and the Zung) perform equally well in general medical samples.

There are several limitations to the results of this systematic review. First, there was moderate to high heterogeneity for most measures. Secondly, there is a paucity of validity studies using the ICD-10 as the criterion standard compared with the DSM-IV, which may favour tools using DSM items, and therefore the authors recommend future examination using this outcome. Thirdly, there were widely used or potentially useful scales that had few or no studies in the physically ill; these include the Montgomery–Asberg Depression Scale (MADRS), and the Clinically Useful Outcome Depression Scale (CUDOS). Further research is needed on these scales for people with chronic physical health problems. Fourthly, there were a number of different semi-structured methods used to determine the interview-based diagnosis, including the Schedules for Clinical Assessment in Neuropsychiatry (SCAN), the Composite International Diagnostic Interview (CIDI), the Structured and Clinical Interview for DSM-III-R (SCID), and the Diagnostic Interview schedule (DIS), all of which may vary in diagnostic accuracy. A further limitation is the lack of cost-effectiveness analyses assessing the cost impact of false positives associated with the use of case-identification measures. However, it should be noted that the cost-effectiveness of case identification is very complex to model and requires a number of assumptions concerning probabilities assigned to events in the depression treatment care pathway, and explicit values of treatment outcomes. Therefore, such issues were considered beyond the scope of this paper.

It should also be acknowledged that the use of case-identification tools may not be translated into real benefit in clinical practice. Case identification may bring limited benefit if there are no effective assessment and treatment services in place, as professionals may be reluctant to make a diagnosis of depression if they have limited resources on which to call. The aim of the NICE guideline for which this review was conducted is to promote the commissioning of such services. The impact of case finding on the individual consultation may be important, since the use of the PHQ-9 severity questionnaire can cause a tension within the consultation, with GPs struggling to manage formal assessment versus personal enquiry.

From this data synthesis, it appears that there are a number of instruments for the case identification of depression in the medically ill that have similar accuracy. A consideration of both accuracy and acceptability suggests that the two stem questions may be the most efficient initial method, although further validation is needed. We do not recommend the use of a single question used alone. GPs and practice nurses should not rely on the case-finding questions alone; they should be confident to complete an assessment of the patient’s mental state and risk, and a pathway within the practice should be in place (particularly when it is the practice nurse who has done the case finding). Resources within the practice should be available to support patients who have depression and a chronic physical health problem, and primary care practitioners should have well-defined links with local primary care mental health services, which should offer appropriate interventions for such patients, including a collaborative care approach as recommended by NICE.
REFERENCES


Appendix 1. Full list of instruments considered

1. Beck Depression Inventory (BDI):
   - BDI-II
   - BDI Cognitive-Affective scale
   - BDI Fast Screen
2. Patient Health Questionnaire (PHQ):
   - PHQ-9
   - PHQ-2
3. Two stem questions:
   - These are similar to the PHQ-2 except that the scoring system is dichotomous (‘yes’ or ‘no’) rather than the Likert scale used for both PHQ-9 and PHQ-2, and the period of reported low mood or loss of interest is 1 month rather than 2 weeks.
4. General Health Questionnaire:
   - GHQ-12
   - GHQ-28
5. Centre of Epidemiological Studies – Depression (CES-D)
6. Geriatric Depression Scale (GDS)
   - GDS-30
   - GDS-15
7. Zung Self Rating Depression Scale
8. Hospital Anxiety and Depression Scale — Depression
9. Hamilton Depression Rating Scale (HDRS)
   - Both 17- and 21-item versions
10. Montgomery-Asberg Depression Rating Scale (MADRS)
11. Clinically Useful Depression Outcome Scale
12. One-item measures of depression
13. Edinburgh Postnatal Depression Scale

---

### References

## Appendix 2. Summary characteristics of included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Patients</th>
<th>Female</th>
<th>Mean age, years</th>
<th>Physical health condition</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aben et al., 2002&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Netherlands</td>
<td>171</td>
<td>91</td>
<td>68</td>
<td>Stroke</td>
<td>BDI, HADS-D, HDRS</td>
</tr>
<tr>
<td>Agreti and Dehlin, 1989&lt;sup&gt;31&lt;/sup&gt;</td>
<td>Sweden</td>
<td>40</td>
<td>22</td>
<td>80</td>
<td>Stroke</td>
<td>CES-D, GDS, HDRS</td>
</tr>
<tr>
<td>Akechi et al., 2006&lt;sup&gt;27&lt;/sup&gt;</td>
<td>Japan</td>
<td>205</td>
<td>68</td>
<td>61</td>
<td>Cancer</td>
<td>1-item, Two stem questions</td>
</tr>
<tr>
<td>Akizuki et al., 2003&lt;sup&gt;24&lt;/sup&gt;</td>
<td>Japan</td>
<td>275</td>
<td>164</td>
<td>52</td>
<td>Cancer</td>
<td>HADS-D, 1-item</td>
</tr>
<tr>
<td>Akyzukiet al., 2003&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Japan</td>
<td>275</td>
<td>164</td>
<td>52</td>
<td>Cancer</td>
<td>HADS-D, BDI</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Turkey</td>
<td>100</td>
<td>93</td>
<td>67</td>
<td>Cancer</td>
<td>BDI, HDRS, BDI</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>US</td>
<td>125</td>
<td>32</td>
<td>55</td>
<td>Various (general medical)</td>
<td>GDS, CES-D, Two stem questions</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>US</td>
<td>67</td>
<td>44</td>
<td>77</td>
<td>Various (general medical)</td>
<td>GDS</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>UK</td>
<td>41</td>
<td>16</td>
<td>53</td>
<td>Dialysis</td>
<td>BDI</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>US</td>
<td>197</td>
<td>103</td>
<td>Not reported</td>
<td>Cancer</td>
<td>Two stem questions</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Italy</td>
<td>132</td>
<td>98</td>
<td>50</td>
<td>Breast cancer</td>
<td>HADS-D, GDS</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Canada</td>
<td>99</td>
<td>36</td>
<td>51</td>
<td>Renal dialysis</td>
<td>BDI</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>UK</td>
<td>618</td>
<td>371</td>
<td>80</td>
<td>Medically ill (general medical)</td>
<td>GDS</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Spain</td>
<td>1000</td>
<td>451</td>
<td>43</td>
<td>Medically ill (general medical)</td>
<td>GHQ-28</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Turkey</td>
<td>109</td>
<td>36</td>
<td>67</td>
<td>Parkinson’s Disease</td>
<td>GDS</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Germany</td>
<td>126</td>
<td>37</td>
<td>51</td>
<td>Heart disease</td>
<td>BDI</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Germany</td>
<td>682</td>
<td>327</td>
<td>66</td>
<td>Congestive heart failure</td>
<td>BDI non-somatic</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>US</td>
<td>155</td>
<td>98</td>
<td>50</td>
<td>Medically ill (general medical)</td>
<td>GDS, CES-D, GDS non-somatic items</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>US</td>
<td>70</td>
<td>41</td>
<td>77</td>
<td>Visual impairment</td>
<td>GDS, GDS-15</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>US</td>
<td>93</td>
<td>42</td>
<td>70</td>
<td>Medically ill (general medical)</td>
<td>GDS</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>US</td>
<td>88</td>
<td>23</td>
<td>55</td>
<td>Hepatitis C</td>
<td>BDI non-somatic</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Italy</td>
<td>109</td>
<td>83</td>
<td>55</td>
<td>Cancer</td>
<td>HADS-D</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Germany</td>
<td>204</td>
<td>98</td>
<td>50</td>
<td>Medically ill (general medical)</td>
<td>GHQ-12, GDS, GDS-15</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>UK</td>
<td>266</td>
<td>264</td>
<td>Not reported</td>
<td>Breast cancer</td>
<td>HADS-D</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>UK</td>
<td>39</td>
<td>19</td>
<td>58</td>
<td>Amyotrophic lateral sclerosis</td>
<td>GDS</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Germany</td>
<td>206</td>
<td>103</td>
<td>48</td>
<td>Musculoskeletal diseases</td>
<td>GHQ-12, HADS-D</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Germany</td>
<td>569</td>
<td>285</td>
<td>54</td>
<td>Medically ill (general medical)</td>
<td>GHQ-12, HADS-D</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Germany</td>
<td>226</td>
<td>226</td>
<td>40</td>
<td>Medically ill (general medical)</td>
<td>Two stem questions</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Germany</td>
<td>88</td>
<td>15</td>
<td>70</td>
<td>Heart failure</td>
<td>GDS-15, HADS-D</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>UK</td>
<td>49</td>
<td>26</td>
<td>79</td>
<td>Stroke</td>
<td>BDI non-somatic</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>US</td>
<td>98</td>
<td>44</td>
<td>57</td>
<td>Haemodialysis</td>
<td>CES-D, BDI</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Germany</td>
<td>376</td>
<td>148</td>
<td>52</td>
<td>Diabetes</td>
<td>BDI, CES-D</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Spain</td>
<td>385</td>
<td>181</td>
<td>38</td>
<td>Medically ill (general medical)</td>
<td>HADS-D, CES-D</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>UK</td>
<td>39</td>
<td>19</td>
<td>Not reported</td>
<td>Not reported</td>
<td>GDS</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>UK</td>
<td>33</td>
<td>25</td>
<td>54</td>
<td>Cancer</td>
<td>HAM-D, GDS, CES-D</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>US</td>
<td>74</td>
<td>2</td>
<td>74</td>
<td>Medically ill (general medical)</td>
<td>GDS-15, GHQ-28</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>UK</td>
<td>75</td>
<td>75</td>
<td>51</td>
<td>Cancer</td>
<td>GHQ-28</td>
</tr>
<tr>
<td>Blank et al., 2004&lt;sup&gt;12&lt;/sup&gt;</td>
<td>UK</td>
<td>513</td>
<td>282</td>
<td>Not reported</td>
<td>Cancer</td>
<td>HADS-D, GHQ-28</td>
</tr>
</tbody>
</table>
### Appendix 2 continued. Summary characteristics of included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Patients</th>
<th>Female</th>
<th>Mean age, years</th>
<th>Physical health condition</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jackson and Baldwin, 1993†</td>
<td>US</td>
<td>59</td>
<td>Not reported</td>
<td>77</td>
<td>Medically ill (general medical)</td>
<td>GDS, GDS-15</td>
</tr>
<tr>
<td>Jelford et al., 2004‡</td>
<td>US</td>
<td>100</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Cancer</td>
<td>1-item</td>
</tr>
<tr>
<td>Johnson et al., 1995‡</td>
<td>Australia</td>
<td>204</td>
<td>Not reported</td>
<td>71</td>
<td>Stroke</td>
<td>GDS, HADS-D, GHQ-28</td>
</tr>
<tr>
<td>Katz et al., 2004*</td>
<td>Canada</td>
<td>60</td>
<td>13</td>
<td>61</td>
<td>Cancer</td>
<td>BDI, HADS-D, CES-D</td>
</tr>
<tr>
<td>Kawase et al., 2006*</td>
<td>Japan</td>
<td>305</td>
<td>Not reported</td>
<td>62</td>
<td>Cancer</td>
<td>1-item</td>
</tr>
<tr>
<td>Koenig et al., 1992*</td>
<td>US</td>
<td>109</td>
<td>0</td>
<td>74</td>
<td>Medically ill (general medical)</td>
<td>GDS, GDS-15</td>
</tr>
<tr>
<td>Kugaya et al., 1998*</td>
<td>Japan</td>
<td>128</td>
<td>48</td>
<td>61</td>
<td>Cancer</td>
<td>HADS-D</td>
</tr>
<tr>
<td>Lam et al., 2004†</td>
<td>Hong Kong</td>
<td>100</td>
<td>56</td>
<td>69</td>
<td>Medically ill (general medical)</td>
<td>HADS</td>
</tr>
<tr>
<td>Lamers et al., 2006*</td>
<td>Netherlands</td>
<td>713</td>
<td>350</td>
<td>71</td>
<td>Medically ill (general medical)</td>
<td>PHQ-9</td>
</tr>
<tr>
<td>Laska et al., 2007*</td>
<td>Sweden</td>
<td>89</td>
<td>40</td>
<td>74</td>
<td>Stroke</td>
<td>MADRS</td>
</tr>
<tr>
<td>LeFevre et al., 1999*</td>
<td>UK</td>
<td>79</td>
<td>35</td>
<td>70</td>
<td>Cancer</td>
<td>HADS-D</td>
</tr>
<tr>
<td>Lee et al., 2008*</td>
<td>Japan</td>
<td>253</td>
<td>94</td>
<td>Not reported</td>
<td>Parkinson’s disease</td>
<td>GDS-15</td>
</tr>
<tr>
<td>Leung et al., 1998*</td>
<td>Taiwan</td>
<td>50</td>
<td>Not reported</td>
<td>54</td>
<td>Chronic medical disorders</td>
<td>Zung</td>
</tr>
<tr>
<td>Lightbody et al., 2007*</td>
<td>UK</td>
<td>28</td>
<td>Not reported</td>
<td>72</td>
<td>Stroke</td>
<td>MADRS</td>
</tr>
<tr>
<td>Lincoln et al., 2003*</td>
<td>UK</td>
<td>143</td>
<td>70</td>
<td>66</td>
<td>Stroke</td>
<td>BDI, GHQ-28</td>
</tr>
<tr>
<td>Lloyd-Williams et al., 2000*</td>
<td>UK</td>
<td>100</td>
<td>56</td>
<td>57</td>
<td>Cancer</td>
<td>EPDS, 1-item</td>
</tr>
<tr>
<td>Lloyd-Williams et al., 2001*</td>
<td>UK</td>
<td>100</td>
<td>56</td>
<td>57</td>
<td>Cancer</td>
<td>HADS-D, 1-item</td>
</tr>
<tr>
<td>Lloyd-Williams et al., 2004*</td>
<td>UK</td>
<td>74</td>
<td>37</td>
<td>68</td>
<td>Cancer</td>
<td>EPDS, 1-item</td>
</tr>
<tr>
<td>Love et al., 2002†</td>
<td>Australia</td>
<td>303</td>
<td>303</td>
<td>Not reported</td>
<td>Breast cancer</td>
<td>HADS-D, BDI non-somatic</td>
</tr>
<tr>
<td>Love et al., 2004*</td>
<td>Australia</td>
<td>227</td>
<td>227</td>
<td>52</td>
<td>Breast cancer</td>
<td>HADS, BDI non-somatic</td>
</tr>
<tr>
<td>Lowe et al., 2007*</td>
<td>Canada</td>
<td>119</td>
<td>30</td>
<td>63</td>
<td>Myocardial infarction or angina</td>
<td>BDI, GDS</td>
</tr>
<tr>
<td>Lowe et al., 2004*</td>
<td>Germany</td>
<td>501</td>
<td>167</td>
<td>42</td>
<td>Medically ill (general medical)</td>
<td>HADS-D, PHQ-9, PHQ-2</td>
</tr>
<tr>
<td>Lustman et al., 1997*</td>
<td>US</td>
<td>172</td>
<td>83</td>
<td>48</td>
<td>Diabetes</td>
<td>BDI</td>
</tr>
<tr>
<td>Lykouras et al., 1996*</td>
<td>Greece</td>
<td>107</td>
<td>57</td>
<td>43</td>
<td>Neurological disorder</td>
<td>GHQ-28</td>
</tr>
<tr>
<td>Magis et al., 1996*</td>
<td>Italy</td>
<td>220</td>
<td>109</td>
<td>76</td>
<td>Medically ill (general medical)</td>
<td>GDS</td>
</tr>
<tr>
<td>McManus et al., 2005*</td>
<td>US</td>
<td>1024</td>
<td>184</td>
<td>67</td>
<td>Chronic heart disease</td>
<td>PHQ-9, PHQ-2, Two stem questions</td>
</tr>
<tr>
<td>McQuillan et al., 2003‡</td>
<td>US</td>
<td>415</td>
<td>344</td>
<td>58</td>
<td>Rheumatoid Arthritis</td>
<td>CES-D</td>
</tr>
<tr>
<td>Meyer et al., 2003‡</td>
<td>US</td>
<td>45</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Cancer</td>
<td>1-item</td>
</tr>
<tr>
<td>Mitchell et al., 2008*</td>
<td>UK</td>
<td>129</td>
<td>Not reported</td>
<td>58</td>
<td>Cancer</td>
<td>PHQ-2, 1-item</td>
</tr>
<tr>
<td>Mohr et al., 2007*</td>
<td>US</td>
<td>260</td>
<td>190</td>
<td>51</td>
<td>Multiple sclerosis</td>
<td>Two stem questions, 1-item</td>
</tr>
<tr>
<td>Narding et al., 2002*</td>
<td>Netherlands</td>
<td>44</td>
<td>16</td>
<td>70</td>
<td>Stroke</td>
<td>HDRS</td>
</tr>
<tr>
<td>Neal et al., 1994*</td>
<td>UK</td>
<td>45</td>
<td>28</td>
<td>77</td>
<td>Medically ill (general medical)</td>
<td>GDS, GDS-15</td>
</tr>
<tr>
<td>O’Rourke et al., 1998†</td>
<td>UK</td>
<td>105</td>
<td>Not reported</td>
<td>68</td>
<td>Stroke</td>
<td>HADS-D</td>
</tr>
<tr>
<td>Okimoto et al., 1982*</td>
<td>Japan</td>
<td>55</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Medically ill (general medical)</td>
<td>Zung</td>
</tr>
<tr>
<td>Olden et al., 2003*</td>
<td>US</td>
<td>439</td>
<td>239</td>
<td>66</td>
<td>Cancer</td>
<td>HADS, HDRS, CES-D</td>
</tr>
<tr>
<td>Ozalp et al., 2008*</td>
<td>Turkey</td>
<td>208</td>
<td>208</td>
<td>51</td>
<td>Cancer</td>
<td>HADS-D</td>
</tr>
<tr>
<td>Parkhe et al., 1988*</td>
<td>US</td>
<td>80</td>
<td>40</td>
<td>58</td>
<td>Stroke</td>
<td>CES-D</td>
</tr>
<tr>
<td>Parker et al., 2002*</td>
<td>Australia</td>
<td>302</td>
<td>175</td>
<td>47</td>
<td>Medically ill (general medical)</td>
<td>BDI non-somatic</td>
</tr>
<tr>
<td>Passik et al., 2001*</td>
<td>US</td>
<td>60</td>
<td>58</td>
<td>31</td>
<td>Cancer</td>
<td>Zung</td>
</tr>
<tr>
<td>Patterson et al., 2006‡</td>
<td>US</td>
<td>310</td>
<td>37</td>
<td>40</td>
<td>HIV</td>
<td>BDI non-somatic</td>
</tr>
</tbody>
</table>
## Appendix 2 continued. Summary characteristics of included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Patients</th>
<th>Female</th>
<th>Mean age, years</th>
<th>Physical health condition</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payne et al, 2007</td>
<td>US</td>
<td>145</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Cancer</td>
<td>Two stem questions 1-item</td>
</tr>
<tr>
<td>Persoons et al, 2003</td>
<td>Netherlands</td>
<td>97</td>
<td>64</td>
<td>48</td>
<td>Otolaryngology</td>
<td>PHQ-9</td>
</tr>
<tr>
<td>Picardi et al, 2005</td>
<td>Italy</td>
<td>141</td>
<td>79</td>
<td>38</td>
<td>Dermatology</td>
<td>GHS-12</td>
</tr>
<tr>
<td>Pomeroy et al, 2001</td>
<td>US</td>
<td>87</td>
<td>52</td>
<td>78</td>
<td>Medically ill (general medical)</td>
<td>GDS-15</td>
</tr>
<tr>
<td>Pogge et al, 2004</td>
<td>US</td>
<td>115</td>
<td>47</td>
<td>43</td>
<td>Heart disease</td>
<td>PHQ-9</td>
</tr>
<tr>
<td>Rapp et al, 1988</td>
<td>US</td>
<td>150</td>
<td>48</td>
<td>69</td>
<td>Medically ill (general medical)</td>
<td>GDS</td>
</tr>
<tr>
<td>Razavi et al, 1990</td>
<td>Belgium</td>
<td>210</td>
<td>140</td>
<td>55</td>
<td>Cancer</td>
<td>HADS-D</td>
</tr>
<tr>
<td>Reuter and Harter, 2000</td>
<td>Germany</td>
<td>188</td>
<td>51</td>
<td>54</td>
<td>Cancer</td>
<td>HADS-D</td>
</tr>
<tr>
<td>Rinaldi et al, 2003</td>
<td>US</td>
<td>181</td>
<td>Not reported</td>
<td>79</td>
<td>Medically ill (general medical)</td>
<td>GHS-15</td>
</tr>
<tr>
<td>Roger et al, 2009</td>
<td>US</td>
<td>87</td>
<td>35</td>
<td>71</td>
<td>Stroke</td>
<td>CES-D</td>
</tr>
<tr>
<td>Rovner et al, 1997</td>
<td>US</td>
<td>70</td>
<td>41</td>
<td>77</td>
<td>Visual impairment</td>
<td>GHS-15</td>
</tr>
<tr>
<td>Sagen et al, 2009</td>
<td>Norway</td>
<td>104</td>
<td>43</td>
<td>65</td>
<td>Stroke</td>
<td>HADS-D</td>
</tr>
<tr>
<td>Scheinthal et al, 2001</td>
<td>US</td>
<td>75</td>
<td>42</td>
<td>74</td>
<td>Medically ill (general medical)</td>
<td>GHS-15</td>
</tr>
<tr>
<td>Schein et al, 1997</td>
<td>US</td>
<td>76</td>
<td>35</td>
<td>70</td>
<td>Medically ill (general medical)</td>
<td>CES-D</td>
</tr>
<tr>
<td>Serrano-Buenas et al, 2008</td>
<td>Ecuador</td>
<td>115</td>
<td>33</td>
<td>70</td>
<td>Parkinson’s disease</td>
<td>HDRS</td>
</tr>
<tr>
<td>Shinar et al, 1986</td>
<td>Israel</td>
<td>27</td>
<td>16</td>
<td>56</td>
<td>Stroke</td>
<td>CES-D</td>
</tr>
<tr>
<td>Silberman et al, 2006</td>
<td>Brazil</td>
<td>44</td>
<td>19</td>
<td>68</td>
<td>Parkinson’s disease</td>
<td>MADRS-BDI</td>
</tr>
<tr>
<td>Singer et al, 2008</td>
<td>Germany</td>
<td>250</td>
<td>23</td>
<td>Not reported</td>
<td>Cancer</td>
<td>HADS-D</td>
</tr>
<tr>
<td>Sirignoli et al, 2009</td>
<td>Turkey</td>
<td>85</td>
<td>53</td>
<td>59</td>
<td>Stroke</td>
<td>GHS-15</td>
</tr>
<tr>
<td>Stafford et al, 2007</td>
<td>Australia</td>
<td>193</td>
<td>39</td>
<td>64</td>
<td>Heart disease</td>
<td>HADS-D</td>
</tr>
<tr>
<td>Strik et al, 2001</td>
<td>US</td>
<td>206</td>
<td>49</td>
<td>59</td>
<td>Myocardial infarction</td>
<td>HADS-D</td>
</tr>
<tr>
<td>Tang et al, 2004</td>
<td>China</td>
<td>100</td>
<td>45</td>
<td>74</td>
<td>Stroke</td>
<td>GHS-15</td>
</tr>
<tr>
<td>Tang et al, 2004</td>
<td>China</td>
<td>60</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Stroke</td>
<td>BDI</td>
</tr>
<tr>
<td>Thekkumprath et al, 2009</td>
<td>UK</td>
<td>150</td>
<td>86</td>
<td>70</td>
<td>Cancer</td>
<td>GHS-15</td>
</tr>
<tr>
<td>Turner et al, 1984</td>
<td>US</td>
<td>40</td>
<td>20</td>
<td>47</td>
<td>Chronic pain</td>
<td>Zung</td>
</tr>
<tr>
<td>Upadhya et al, 1997</td>
<td>UK</td>
<td>72</td>
<td>35</td>
<td>71</td>
<td>Medically ill (general medical)</td>
<td>HADS-D</td>
</tr>
<tr>
<td>Vaher et al, 2007</td>
<td>Estonia</td>
<td>134</td>
<td>Not reported</td>
<td>44</td>
<td>Multiple sclerosis</td>
<td>1-item</td>
</tr>
<tr>
<td>Vargas et al, 2007</td>
<td>Portugal</td>
<td>484</td>
<td>276</td>
<td>70</td>
<td>Medically ill (general medical)</td>
<td>GHS-15</td>
</tr>
<tr>
<td>Walker et al, 2007</td>
<td>UK</td>
<td>361</td>
<td>238</td>
<td>Not reported</td>
<td>Cancer</td>
<td>GHS-15</td>
</tr>
<tr>
<td>Watnick et al, 2005</td>
<td>US</td>
<td>62</td>
<td>20</td>
<td>63</td>
<td>Dialysis</td>
<td>PHQ-9</td>
</tr>
<tr>
<td>Weintrub et al, 2004</td>
<td>US</td>
<td>148</td>
<td>Not reported</td>
<td>71</td>
<td>Parkinson’s disease</td>
<td>GHS-15</td>
</tr>
<tr>
<td>Wilhelm et al, 2004</td>
<td>Australia</td>
<td>212</td>
<td>117</td>
<td>Not reported</td>
<td>Medically ill (general medical)</td>
<td>BDI non-somatic</td>
</tr>
<tr>
<td>Williams et al, 2005</td>
<td>US</td>
<td>296</td>
<td>219</td>
<td>59</td>
<td>Stroke</td>
<td>PHQ-9</td>
</tr>
</tbody>
</table>