Are one or two simple questions sufficient to detect depression in cancer and palliative care? A Bayesian meta-analysis

The purpose of this study is to examine the value of one or two simple verbal questions in the detection of depression in cancer settings. This study is a systematic literature search of abstract and full text databases to January 2008. Key authors were contacted for unpublished studies. Seventeen analyses were found. Of these, 13 were conducted in late stage palliative settings. (1) Single depression question: across nine studies, the prevalence of depression was 16%. A single ‘depression’ question enabled the detection of depression in 160 out of 223 true cases, a sensitivity of 72%, and correctly reassured 964 out of 1166 non-depressed cancer sufferers, a specificity of 83%. The positive predictive value (PPV) was 44% and the negative predictive value (NPV) 94%. (2) Single interest question: there were only three studies examining the ‘loss-of-interest’ question, with a combined prevalence of 14%. This question allowed the detection of 60 out of 72 cases (sensitivity 83%) and excluded 394 from 459 non-depressed cases (specificity of 86%). The PPV was 48% and the NPV 97%. (3) Two questions (low mood and low interest): five studies examined two questions with a combined prevalence of 17%. The two-question combination facilitated a diagnosis of depression in 138 of 151 true cases (sensitivity 91%) and gave correct reassurance to 645 of 749 non-cases (specificity 86%). The PPV was 57% and the NPV 98%. Simple verbal methods perform well at excluding depression in the non-depressed but perform poorly at confirming depression. The ‘two question’ method is significantly more accurate than either single question but clinicians should not rely on these simple questions alone and should be prepared to assess the patient more thoroughly.

British Journal of Cancer (2008) 98, 1934–1943. doi:10.1038/sj.bjc.6604396 www.bjcancer.com
Published online 27 May 2008
© 2008 Cancer Research UK

Keywords: depression; diagnostic validity; meta-analysis; sensitivity
and 8, 6 and 5 item versions of the Edinburgh Depression scale (EPDS) (Pallant et al, 2006; Eberhard-Gran et al, 2007; Lloyd-Williams et al, 2007). Occasionally, authors have developed entirely new short scales, such as the four item case-find for depression (Jefford et al, 2004), or attempted to develop a short scale specifically for palliative settings (Lloyd-Williams et al, 2007).

In the 1990s, several groups working in cancer settings suggested that two questions, or in some cases just a single question, might be sufficient to detect depression in palliative care. Usually, these ultra-short tests formed part of symptom checklists and were not validated against an accepted standard (Miller and Walsh, 1991; Donnelly et al, 1995; Conill et al, 1997; Brünnhagen et al, 1998; Edmonds et al, 1998; Ng and von Gunten, 1998; Pratheepawanit et al, 1999). At the same time, simple (non-verbal) visual-analogue methods of assessing depression, anxiety or distress were developed, exemplified by the NCCN Distress Thermometer and Edmonton Symptom Assessment (Hürny et al, 1996; Vignaroli et al, 2006). The accuracy of these methods was reviewed in mixed cancer settings with the finding that they had reasonable rule-out accuracy but limited case-finding ability (Mitchell, 2007). Yet, it is not clear how simple verbal questions perform alone and when used specifically for patients with advanced cancer.

The aim of this study is to examine the diagnostic accuracy of simple verbal questions to detect depression in cancer and palliative care and to ascertain whether clinicians should rely upon either one question or two questions to detect major depression compared with more established screening tools.

MATERIALS AND METHODS

Search
A systematic literature search, critical appraisal of the collected studies and pooled analysis were conducted. The following abstract databases were searched: Medline 1966-January 2008, PsycINFO 1887-January 2008, Embase 1980-January 2008 and CINAHL 1982-January 2008. In these databases, the keywords (MeSH terms) were ‘distress or anx$ or depress$ or mood’ and ‘screen$ or detect$ or case-finding or recogni$ or recogni$’ and ‘cancer or oncology or malignant or transplant or tumour or metastatic.’ Four full text collections including Science Direct, Ingenta Select, Ovid Full text and Wiley Interscience were searched. In these online databases, the same search terms were used but as a full text search and citation search. The abstract database Web of Knowledge (4.0, ISI) was searched, using the above terms as a text word search, and using key papers in a reverse citation search. Conference abstracts from IPOS 2006 and 2007 were examined. Non-English language papers and abstracts were included but, where necessary, authors were contacted directly for primary data and data in press.

Critical appraisal
The review guidelines for diagnostic tests recently outlined in Evidence Based Medicine were followed (Pai et al, 2004). Questions for each report included the setting, the data integrity, the choice of reference criterion, the method of application of the screening questionnaire and, importantly, the type of outcome measured. Quality appraisal standards are listed in Table 1.

Pooled analysis and meta-analysis
Two methods are possible in combining diagnostic validity studies (Midgette et al, 1993; Irwig et al, 1995). (a) Simple pooling of the raw data and re-calculation of the cumulative sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). This method assumes a consistent prevalence between studies and in future work. (b) Correction for the variance in prevalence by relying on the stability of sensitivity and specificity by calculating a pooled weighted rate of sensitivity and specificity and then calculating PPV and NPV according to local prevalence data (Glasziou and Irwig, 1998). In this case, a Bayesian curve can be constructed of all post-test probabilities if a given test is positive or negative. Overall accuracy was calculated using the identification index which is the fraction correct minus faction incorrect. The reciprocal of the identification index is the number needed to screen (Mitchell, 2009).

Standards of accuracy
Performance was taken as follows: <0.2 poor, >0.2≤0.4 fair, >0.4≤0.6 moderate, >0.6≤0.8 good and >0.8≤1 very good; adapted from that originally proposed by Landis and Koch (Landis and Koch, 1977).

Outcome measures
The majority of studies defined depression using a psychiatric interview (applied in a semistructured or clinical interview) but a minority utilised standardised rating scales (Murphy, 2002).

RESULTS

Systematic literature search
The search identified 98 analyses specifically examining ultra-short methods (Figure 1). Studies that examined one or two question methods in non-cancer medical patients or in primary care were excluded. A total of 39 studies had no gold standard and 29 examined visual-analogue methods and were therefore excluded. Thirteen studies were not sufficiently detailed for inclusion. Thus, 17 analyses of verbal/written questions to detect depression in cancer were included. No attempt was made to separate questions read by investigators (verbal) from questions read by the patients (written). The data extraction is illustrated in Figure 1 in accordance with Quality of Reporting of Meta-analyses guidelines (Moher et al, 1999).

Of 17 analyses, 9 examined one single depression question: Chochinov et al (1997); Lloyd-Williams et al (2003a); Meyer et al (2003a); Jefford et al (2004); Akechi et al (2006); Kawase et al (2006); Ohno et al (2006); Payne et al (2007); and Mitchell et al (2008). Three analyses examined one single interest item: Akechi et al (2006); Payne et al (2007); and Mitchell et al (2008). Five analyses examined a combination of two questions: Akechi et al (2006); 65 Payne et al (2007); Chochinov et al (1997); Gessler et al (2007); and Mitchell et al (2008).

Critical appraisal
The mean sample size was 165.8 (s.d. 55.7). However, several studies examined different verbal methods in the same sample and thus there were actually 1579 unique patients under study. Eleven studies took place in palliative settings and/or specifically in those with late-stage cancer. Three took place in mixed stages and three in predominantly early cancers. All but two studies used DSM criteria by clinical interview or by structured clinical interview. Four looked at major or minor depression combined and the remainder looked at major depression alone. The most common question for depression was simply ‘Are you depressed?’ but variations included ‘Describe your mood over the last week’ and the PHQ2 question two ‘Over the last 2 weeks, how often have you been bothered by feeling down, depressed or hopeless?’ The loss of interest question was asked in three different
ways, namely 'Have you lost interest?', 'Have you experienced loss of interest in things or activities that you would normally enjoy?' and the PHQ2 question one 'Over the last 2 weeks, how often have you been bothered by little interest or pleasure in doing things?' The two-question approach was always asked as question 1 (Q1) or question 2 (Q2), which favours sensitivity at the expense of specificity compared to the Q1 and Q2 approach (Coyne and Mitchell, 2007). Further details of the studies are shown in Tables 1 and 2.

Pooled analysis and meta-analysis

Single depression/mood question Across nine studies, the prevalence of depression was 16%. Using the simple pooled method, the single depression question enabled the detection of depression in 160 out of 223 true cases, a sensitivity of 72% (95% CI 66.3–76.8%) and correctly reassured 964 out of 1166 non-depressed cancer sufferers, a specificity of 83% (95% CI 81.6–83.6). Thus, the PPV was 44% and the NPV 94%. The Youden score was 0.544 (95% CI 0.479–0.603). Using the meta-analytic approach, the weighted sensitivity was 74.1% (95% CI 74.2–89.9) and the specificity was 86% (95% CI 84.4–89.9). The PPV was 48% and the NPV 97%. The Youden score was 0.692 (95% CI 0.586–0.768).

Analysing the results by natural frequencies, out of every 100 screening applications, the single depression question would correctly rule out 69 non-depressed, rule in 11 out of 18 cases, missing 5 and also giving 15 false-positive diagnoses (Figure 2). Thus, the identification index (net gain) would be 61.8% and the number needed to screen in order to yield one additional correct identification would be 1.62.

Single interest question There were only three studies examining the 'loss-of-interest' question, with a pooled prevalence of 14%. Using the simple pooled method, this question allowed the detection of 60 out of 72 cases (sensitivity 83%) (95% CI 74.2–89.9) and excluded 394 from 459 non-depressed cases (specificity of 86%) (95% CI 84.4–89.9). The PPV was 48% and the NPV 97%. The Youden score was 0.692 (95% CI 0.586–0.768). Using the meta-analytic approach, the weighted sensitivity was 82.4% (95% CI 73.0–90.0) and the specificity was 86.4 (95% CI 83.0–89.3).

Analysing the results by natural frequencies, out of every 100 screening applications, the single interest question would correctly rule out 74 non-depressed, rule in 11 out of 14 cases, missing 2 and also giving 12 false-positive diagnoses (Figure 2). Thus, the identification index (net gain) would be 71% and the number needed to screen in order to yield one additional correct identification would be 1.41.

Two questions (low mood and low interest) In five studies using a two-question combination (Q1 or Q2), the prevalence of depression was 17%. Using the simple pooled method, the two-question combination facilitated a diagnosis of depression in 138 of 151 true cases (sensitivity 91.4%) (95% CI 86.4–94.8) and gave correct reassurance to 645 out of 749 non-depressed cases (specificity 86%) (95% CI 85.1–86.8). The PPV was 57% and the NPV 98%. The Youden score was 0.775 (95% CI 0.771–0.816), significantly higher than either of the single questions used alone. Using the meta-analytic approach, the weighted sensitivity was 92.7% (95% CI 87.9–96.9) and specificity 87.4% (95% CI 84.9–89.7).
## Table 1: Methodological aspects of simple verbal questions for depression in cancer

| Year | Author | Screening question | Reference Standard | Sample used for diagnostic testing | Setting | Comment | Quality appraisal |
|------|--------|--------------------|-------------------|----------------------------------|---------|---------|------------------|
| 1997 | Chochinov et al, 1997 | ‘Are you depressed? ‘Are you depressed OR have you lost interest? | RDC Mj+Mn Dep | 197 | Palliative | Sample was 94 male and 103 female inpatients receiving palliative care for advanced terminal cancer. Enhanced reliability was attempted by having an observer (a psychiatrist or psychologist) attend a random sample of 27 interviews. Chochinov et al examined whether the first two questions of the diagnostic interview could be used on their own when used at the same time as the clinical interview. The lack of administration of a blind independent criterion standard may partly explain the high accuracy of this method alone. Sample–Adequate Blinding –Poor Gold Standard –Adequate Sample integrity–Adequate |
| 2003 | Lloyd-Williams et al, 2003a | ‘Are you depressed?’ | Clinical interview based on DSMIV | 74 | Palliative | Recruited those in palliative and supportive day care over 6 months. Gold standard was the semi-structured clinical psychiatric interview based on DSMIV, although the exact method of administration was not disclosed. Sample–Poor Blinding –Adequate Gold Standard –Adequate Sample integrity–Good |
| 2006 | Akechi et al, 2006 | ‘Are you depressed?’ ‘Have you lost interest?’ | 1. Structured clinical interview for DSM-III-R. Mj Dep + adjustment | 205 | Palliative | The reliability (kappa coefficient) of the interview ratings was investigated by having another trained psychiatrist attend the first 29 consecutive interviews as a second rater. Mean age was 61 years; 137 (66%) subjects were male and 51 (24%) were in full-time employment. The most frequent primary cancer site was the lung (38%). In head-to-head analysis the single-item interview ‘Have you lost interest or pleasure?’ (AUC 0.92) performed better than ‘Are you depressed?’ (AUC 0.85) or indeed the HADS-D arm (AUC 0.82) or the HADS-Total score (AUC = 0.79). Sample–Good Blinding –Adequate Gold Standard –Adequate Sample integrity–Adequate |
| 2006 | Kawase et al, 2006 | ‘Are you depressed?’ | DSM-IV-TR Mj + Mn Dep | 282 | Mixed Cancer | Sample aged 26–90 years (mean 62.2). Interview by clinical psychologists with 100% concordance. Major and minor depression not separated in the analysis. Sample–Good Blinding –Adequate Gold Standard –Adequate Sample integrity–Adequate |
| 2007 | Payne et al (2007) | ‘Are you depressed?’ ‘Have you experienced loss of interest in things or activities that you would normally enjoy?’ | Clinical interview based on DSMIV major depressive disorder as defined by DSM conducted by one of two mental health professionals PRIME-MD (DSMIV) but validation based on the 4Q Brief Case-Find for Depression (BCD) | 167 (74% were suffering from cancer) | Inpatient Palliative Unit | A subgroup analysis of individuals with a past experience of depressive illness, (n 95) revealed that a significant number screened positive for depression by the screening test, 55.2% (16/29) compared with those with no background history of depression, 33.3% (22/66) (P = 0.045). Also used Prime-MD, BDI and HADS. Authors suggest that the BCD can be administered in 1 min in oncology and palliative settings. Sample–Adequate Blinding–Adequate Gold Standard –Adequate Sample integrity–Adequate |
| 2004 | Jefford et al, 2004 | ‘Over the past couple of weeks, have you been feeling unhappy or depressed?’ | PRIME-MD (DSMIV) | 100 | Late (60% palliative) | Sample–Poor Blinding–Poor Gold Standard –Poor Sample integrity–Adequate |
| Year   | Author                          | Screening question                                                                 | Reference Standard                  | Sample used for diagnostic testing (n) | Setting                     | Comment                                                                 | Quality appraisal |
|--------|---------------------------------|-------------------------------------------------------------------------------------|-------------------------------------|---------------------------------------|-----------------------------|-------------------------------------------------------------------------|-------------------|
| 2006   | Ohno et al, 2006                | 'Are you depressed or not?'                                                        | HADS Combined > 14                 | 160                                   | Mixed Cancers              | Assessed diagnostic accuracy of patients who refused to answer Q 'are you depressed' | Sample – Adequate |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Blinding – Poor   |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Gold Standard – Poor |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Sample integrity – Poor |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Sample integrity – Adequate |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Adequate          |
| 2003   | Meyer et al, 2003               | Schedule of affective disorders and schizophrenia (SADS) item: please describe your mood over the past week(s) (or since last seen) in terms of low mood or depression (not at all, slight vs mild, moderate or severe). | Structured clinical interview for DSM-III-R (SCID), a semi-structured interview for diagnosing depression | 45                           | Inpatients with advanced cancer (prognosis B/6 months) consecutively referred to a hospital palliative care team | Authors compared mood evaluation questionnaire (MEQ) and the structured clinical interview for DSM-III-R (SCID) as well as the single-item interview screening question. The MEQ and SCID had moderate agreement (weighted kappa 0.52 over all interviews). At first interview, 26 (38%) patients were depressed using MEQ, seven (16%) of these severely. | Sample – Adequate |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Blinding – Adequate |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Gold Standard – Adequate |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Sample integrity – Good |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Adequate          |
| 2007   | Gessler et al, 2007 (online published) | 2Q 'Have you lost interest or pleasure? OR 'Are you depressed?'                     | HADS combined > 11                 | 160                                   | Mixed outpatients with cancer | 53% female and mean age 60.0 years. Also tested were the HADS, GHQ-12, BSI-18 and DT. | Sample – Adequate |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Blinding – Poor   |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Gold Standard – Poor |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Sample integrity – Poor |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Adequate          |
| 2008   | Baker-Glenn, Thiagarajan, Chaudhuri, Granger, Symonds, Mitchell, 2008 (unpublished) | PHQ Q1 – ‘Over the last 2 weeks, how often have you been bothered by little interest or pleasure in doing things?’ PHQ Q2 ‘Over the last 2 weeks, how often have you been bothered by feeling down, depressed or hopeless?’ | DSMIV major depression              | 215                                   | Cancer patients attending chemotherapy | Mean age was 57.7 years and their mean time from diagnosis 7.3 months. The prevalence of major depressions was 11.1%. Patients also received the HADS, PHQ9 and DT. | Sample – Adequate |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Blinding – Poor   |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Gold Standard – Poor |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Sample integrity – Adequate |
|        |                                 |                                                                                     |                                     |                                       |                             |                                                                         | Adequate          |

(u) = Uncorrected; 1Q = single question test; 2Q = two question test; RDC = research diagnostic criteria; HADS = Hospital Anxiety and Depression Scale; CESD = Center for Epidemiologic Studies Depression Scale (CES-D); BCD = Brief Case Find for Depression includes the items: (A) been having restless or disturbed nights? (B) been feeling unhappy or depressed? (C) felt unable to overcome your difficulties! (D) been dissatisfied with the way you’ve been doing things? Quality appraisal: Sample: < 100 = poor; < 250 = adequate; < 500 = good; > 499 = excellent; Blinding – both scale and gold standard by same rater = poor; different rater = adequate; different rate blind to results = good; Gold Standard – Depression severity scale = poor; DSM or ICD10 criteria = adequate; semi-structured or structured interview = good; sample integrity – unexplained missing data = poor; partial loss to follow-up = adequate; no missing data = good. Dep = Depression; PHQ = Patient Health Questionnaire; Maj = Major; Min = Minor.
Table 2  Statistical summary of simple verbal questions for depression in cancer

| Reference | Assessment of depression | Total cases with depression | Total cases without depression | Specificity | PPV | NPV | Youden Score | Number needed to screen |
|-----------|--------------------------|----------------------------|-------------------------------|-------------|-----|-----|--------------|-------------------------|
| **Depression question** | | | | | | | | |
| Akechi et al, 2006 | 1Q – ‘Are you depressed?’ | 14 | 0.79 | 195 | 0.92 | 0.41 | 0.98 | 0.70 | 1.22 |
| Payne et al, 2007 | 1Q – ‘Are you depressed?’ | 43 | 0.70 | 124 | 0.81 | 0.57 | 0.89 | 0.51 | 1.76 |
| Chochinov et al, 1997 | 1Q – ‘Are you depressed?’ | 24 | 1.00 | 173 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Lloyd-Williams et al, 2003a | 1Q – ‘Are you depressed?’ | 20 | 0.55 | 54 | 0.74 | 0.44 | 0.82 | 0.29 | 2.64 |
| Jefford et al, 2004 | 1Q – ‘Over the past couple of weeks, have you been feeling unhappy or depressed?’ | 12 | 0.67 | 88 | 0.70 | 0.24 | 0.94 | 0.37 | 2.50 |
| Ohno et al, 2006 | 1Q – ‘Are you depressed or not?’ | 54 | 0.93 | 106 | 0.31 | 0.41 | 0.89 | 0.24 | 2.67 |
| Meyer et al, 2003 | 1Q – ‘Describe your mood over the last week’ | 17 | 0.35 | 28 | 0.75 | 0.46 | 0.66 | 0.10 | 5.00 |
| Baker-Glenn, Thiagarajan; Chaudhuri, Granger, Symonds, Mitchell, 2008 (unpublished) | 1Q – ‘Over the last 2 weeks, how often have you been bothered by feeling down depressed or hopeless?’ | 15 | 0.67 | 140 | 0.95 | 0.59 | 0.96 | 0.62 | 1.18 |
| Kawise et al, 2006 | 1Q – ‘Are you depressed?’ | 24 | 0.42 | 258 | 0.86 | 0.22 | 0.94 | 0.28 | 1.55 |
| **Pooled summary** | **Subtotal – depressed** | **223** | **0.72** | **1166** | **0.83** | **0.44** | **0.94** | **0.54** | **1.62** |
| **Interest question** | | | | | | | | |
| Payne et al, 2007 | 1Q – ‘Have you experienced loss of interest in things or activities that you would normally enjoy?’ | 43 | 0.79 | 124 | 0.73 | 0.50 | 0.91 | 0.52 | 2.06 |
| Baker-Glenn, Thiagarajan; Chaudhuri, Granger, Symonds, Mitchell, 2008 (unpublished) | 1Q – ‘Over the last 2 weeks, how often have you been bothered by little interest or pleasure in doing things?’ | 15 | 0.87 | 140 | 0.89 | 0.46 | 0.98 | 0.76 | 1.28 |
| Akechi et al, 2006 | 1Q – ‘Have you lost interest?’ | 14 | 0.93 | 195 | 0.92 | 0.45 | 0.99 | 0.85 | 1.19 |
| **Pooled summary** | **Subtotal – interest** | **72** | **0.83** | **459** | **0.86** | **0.48** | **0.97** | **0.69** | **1.41** |
| **Combination question** | | | | | | | | |
| Payne et al, 2007 | 2Q – ‘Depressed or loss of interest’ | 43 | 0.91 | 124 | 0.68 | 0.49 | 0.95 | 0.58 | 2.11 |
| Akechi et al, 2006 | 2Q – ‘Have you lost interest or pleasure? OR ‘Are you depressed? | 14 | 1.00 | 195 | 0.86 | 0.34 | 1.00 | 0.86 | 1.35 |
| Baker-Glenn, Thiagarajan; Chaudhuri, Granger, Symonds, Mitchell, 2007 (online published) | 2Q – PHQ1 or 2 | 15 | 1.00 | 140 | 0.87 | 0.45 | 1.00 | 0.87 | 1.30 |
| Gessler et al. 2007 (online published) | 2Q – ‘Have you lost interest or pleasure? OR ‘Are you depressed? | 43 | 0.79 | 117 | 0.86 | 0.68 | 0.92 | 0.65 | 1.45 |
| Chochinov et al (1997) | 2Q – Depressed mood OR loss of interest | 36 | 1.00 | 173 | 0.98 | 0.92 | 1.00 | 0.98 | 1.03 |
| **Pooled summary** | **Subtotal – 2Q** | **151** | **0.91** | **749** | **0.86** | **0.57** | **0.98** | **0.78** | **1.35** |

*PPV = positive predictive value; NPV = negative predictive value. 1Q = Single question; 2Q = Two question combination. *Number needed to screen = reciprocal of fraction correct minus fraction incorrect.*
Analysing the results by natural frequencies, out of every 100 screening applications, these two questions would correctly rule out 72 non-depressed, rule in 15 out of 18 cases, overlooking 1 true case and also giving 12 false-positive diagnoses (Figure 2). Thus, the identification index (net gain) would be 74% and the number needed to screen in order to yield one additional correct identification would be 1.35.

**Figure 2** Natural screening accuracy of simple verbal questions in the detection of cancer-related depression.

Bayesian pre-test–post-test gain  Assuming sensitivity and specificity hold for different rates of depression, a Bayesian curve was constructed of all post-test probabilities where a test result was either positive or negative. This illustrates the pre-test–post-test gain for each method and the predictive value conditional upon different baseline rates of depression. Figure 3 demonstrated the superior difference in gain for the two-question approach with the depression question alone.

**Figure 3** Bayesian graph: test probability (PPV and NPV) as function of test result – depression question vs combination question.

CONCLUSIONS

A previous pooled analysis found eight diagnostic validity analyses of one or two single item questions in the detection of depression in cancer settings (Mitchell, 2007). This study updates the previous analysis to include 17 analyses, 13 involving late-stage cancer and/or palliative settings. It is important to note that the average prevalence of depression across these studies was 16% (range 7–38%), which means that any case-finding method is likely to have difficulty detecting true cases without generating false positives. Results show that the loss of interest question is somewhat better than the depression question when used alone. This corresponds to research showing that, of many symptoms of depression, loss of interest best discriminated between patients with and without diagnosis of comorbid affective disorder (Reuter et al, 2004). However, two questions are significantly better than any one
question for detecting depression (Youden score was 0.78 for two questions, 0.54 for the depression question and 0.69 for the interest question). In fact, two questions are better for both ruling in and ruling out a diagnosis than either question alone, although the loss of interest question is also an excellent method of excluding depression. No method achieved a case-finding accuracy of more than 60% according to the PPVs. This means that at best there would be 43 false positives out of each 100 positive screens.

There has been increasing interest in short verbal and non-verbal screening methods. Lloyd-Williams and colleagues (Lawrie et al., 2004) found that among consultants working in palliative medicine, 10% asked the patient ‘are you depressed?’ to detect depression. Mitchell et al. (2008) and colleagues found that simple verbal questions were the most preferred active method of detecting depression, used by 30% of cancer professionals. However, this meta-analysis raises an important caution for all those using one or two questions. Assuming use of the two-question combination, this would mean that out of every 100 screening applications answering yes to either of these two questions, only 1 true case would be missed but 12 false-positive diagnoses would be generated (Figure 2). Thus, a second method with better PPV would be required. This could be a thorough clinical assessment by someone confidently able to diagnose depression or it could be a longer validated depression severity scale (Robinson and Crawford, 2005). However, there is no agreement on which is the optimal case-finding method and rarely has any method shown a case-finding (PPV) accuracy that exceeds 0.80. (Lloyd-Williams et al., 2003b; Trask, 2004). There is also no agreement on how often a tool should be applied (Lloyd-Williams and Riddleston, 2002). For example, Love et al. (2004) found that the HADS depression subscale had a PPV of 0.79 when a cut-off of 7+ was used. Using the same cut-off, Le Fevre et al. (1999) found a PPV of 0.42 in a palliative settings. Recently, Lloyd-Williams et al. (2007) showed that a six-item adaptation of the EPDS had a PPV of 0.65, also in a palliative setting.

In conclusion, no method has been shown to be sufficiently accurate to be considered the definitive screening or case-finding tool for cancer-related depression. Simple questions should be considered as a method of exclusion or combined with more detailed tests. Future work should move beyond screening for psychopathology alone to also consider unmet needs. That is, those individuals with emotional disorders (distress, anxiety, depression and anger) who require and desire professional help (Graves et al., 2007).

ACKNOWLEDGEMENTS

Declaration: I have no declarations, no conflicts of interest and no financial ties. I have received no grant support for this article. Thanks to the following individuals who very kindly supplied original data: Elisabeth Andritsch, Elisabeth Ladinek & Silke Zloklikovits, Medical University Graz, Austria; Greg Carter & Catherine Adams, Newcastle Mater Hospital, Australia; Joe Low, Marie Curie Palliative Care Research Unit, London; Anja Mehnert, Hamburg Germany; Anna Szekely, Eötvös Loránd University, Budapest. Thanks also to Elena Baker-Glenn for helpful comments and suggestions.

REFERENCES

Akechi T, Okuyama T, Sugawara Y, Shima Y, Furukawa TA, Uchitomi Y (2006) Screening for depression in terminally ill cancer patients in Japan. J Pain Symptom Manage 31: 5–12

Andresen EM, Malmgren JA, Carter WB, Patrick DL (1994) Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). Am J Prev Med 10(2): 77–84

Brunelli C, Costantini M, Di Giuolo P, Gallucci M, Fusco F, Miccinesi G, Paci E, Perusselli M, Morino P, Piazza M, Tamburini M, Toscani F (1998) Quality-of-life evaluation: when do terminal cancer patients and health care providers agree? J Pain Symptom Manage 15: 151–158

Bui QT, Ostir GV, Kuo YF, Freeman J, Goodwin JS (2005) Relationship of psychological distress with cancer stage and its recognition by doctors in patients with cancer. Br J Cancer 89(10): 1187–1189

Burnam MA, Wells KB, Leake B, Landsverk J (1988) Development of a brief screening instrument for detecting depressive disorders. Med Care 26: 775–789

Carlson LE, Angen M, Cullum J, Goodey E, Koopmans J, Lamont L, MacRae S, Ford S, Fallowfield L, Bultz B (2004) High levels of untreated distress and fatigue in cancer patients. Br J Cancer 90: 2297 – 2304

Chochinov HM, Wilson KG, Enns M, Lander S (1997) ‘Are you depressed?’ Screening for depression in the terminally ill. Am J Psychiatry 154: 674–676

Cole JC, Rabin AS, Smith TL, Kaufman AS (2004) Development and validation of a Rasch-derived CES-D short form. Psychological Assessment 16(4): 360 – 372

Coneill C, Verger E, Henriquez I, Saiz N, Espier M, Lugo F, Garrigos A (1997) Symptom prevalence in the last week of life. J Pain Symptom Manage 14: 328 – 331

Covic T, Pallant JF, Conaghan PG, Tannant A (2007) A longitudinal evaluation of the Center for Epidemiologic Studies-Depression scale (CES-D) in a rheumatoid arthritis population using Rasch analysis. Health Qual Life Outcomes 5: 41

Coyle JC, Mitchell AJ (2007) NICE may be discouraging detection of postpartum depression. BMJ 334(7593): 550 Donnelly S, Walsh D, Rybicki L (1995) The symptoms of advanced cancer: identification of clinical and research priorities by assessment of prevalence and severity. J Palliat Care 11: 27 – 32

Durkin I, Kearney M, O’Siorain L (2003) Psychiatric disorder in palliative care. Palliat Med 17: 212 – 218

Eberhard-Gran M, Eskild A, Samuelsen SO, Tambs K (2007) A short matrix-version of the Edinburgh Depression Scale. Acta Psychiatr Scand 116: 195 – 200

Edmonds PM, Stuttford JM, Peny J, Lynch AM, Chamberlain J (1998) Hospital palliative care teams improve symptom control? Use of a modified STAS as an evaluation tool. Palliat Med 12: 345–351

Fallowfield L, Ratcliffe D, Jenkins V, Saul J (2001) Psychiatric morbidity and its recognition by doctors in patients with cancer. Br J Cancer 84(6): 1011 – 1015

Ford S, Fallowfield L, Lewis S (1994) Can oncologists detect distress in their out-patients and how satisfied are they with their performance during bad news consultations? Br J Cancer 70: 767 – 770

Furlanetto L, Mendlowicz M, Bueno JR (2005) The validity of the Beck Depression Inventory-Short Form as a screening and diagnostic instrument for moderate and severe depression in medical inpatients. J Affect Disord 86(1): 87 – 91

Gessller S, Low J, Daniella E, Williams R, Brough V, Tookman A, Jones L (2007) Screening for distress in cancer patients: is the distress thermometer a valid measure in the UK and does it measure change over time? A prospective validation study. Psychooncology October 31 [Epub ahead of print]

Glasziou P, Irwig L (1998) Meta-analysis of diagnostic tests. In: Armitage P, Colton T (eds) Encyclopedia of Biostatistics. Indianapolis: Wiley, pp 2579–2585

Graves KD, Arnold SM, Love CL, Kirsh KL, Moore PG, Passik SD (2007) Distress screening in a multidisciplinary lung cancer clinic: prevalence and predictors of clinically significant distress. Lung Cancer 55(2): 215–224

Guo Y, Musselman DL, Manatunga AK, Gilles N, Lawson KG, Porter MR, McDaniel JS, Nemeroff CB (2006) The diagnosis of major depression in patients with cancer: a comparative approach. Psychosomatics 47(5): 376–384
Holton N, Barraclough J (2000) Is the Hospital Anxiety and Depression Scale (HADS) useful in assessing depression in palliative care? Palliat Med 14(3): 219 – 220

Hotopf M, Chidgey J, Additiong-Hall J, Ly KL (2002) Depression in advanced disease: a systematic review – Part 1. Prevalence and case finding. Palliat Med 16(2): 81 – 97

Hoyl MT, Alessi CA, Harker JD, Josephson KR, Pietruszka FM, Koelfgen M, Steiner JR, Fitten LJ. Rubenstein LZ (1999) Development and testing of a five-item version of the geriatric depression scale. J Am Geriatr Soc 47: 873 – 878

Huffman JC, Smith FA, Blais MA, Beiser ME, Januzzi JL, Frichionne GL (2006) Rapid screening for major depression in post-myocardial infarction patients: an investigation using Beck Depression Inventory II items. Heart 92: 1603 – 1606

Hürny C, Bernhard J, Coates A, Peterson HF, Castiglione-Gertsch M, Gelber RD, Rudenstam CM, Collins J, Lindtner J, Goldhirsch A, Senn HJ (1996) Responsiveness of a single-item indicator vs a multi-item scale: assessment of emotional well-being in an international adjuvant breast cancer trial. Med Care 34(5): 234 – 248

Irwig L, Macaskill P, Glasziou P, Fahey M (1995) Meta-analytic methods for diagnostic test accuracy. J Clin Epidemiol 48: 119 – 130

Jefford M, Mileshkin L, Richards K, Thomson J, Matthews JP, Zalcberg J, Jennens R, McLaclan S-A, Wein S, Antill Y, Clarke DM (2004) Rapid screening for depression – validation of the Brief Case-Find for Depression (BCD) in medical oncology and palliative care patients. Br J Cancer 91(5): 960 – 966

Kansanen-Klock NS, Vanderwerker LC, Block SD, Zhang BH, Prigerson HG (2005) Psychiatric disorders and mental health service use in patients with advanced cancer—a report from the coping with cancer study. Cancer 104(12): 2872 – 2881

Kawase E, Karasaki K, Shimotos I, Sameo T, Ito K, Matsuki H, Sakano Y, Horikawa N (2006) Evaluation of a one-question interview for depression in a radiation oncology department in Japan. Gen Hosp Psychiatry 28(4): 321 – 322

Kennard BD, Stewart WM, Olivera R, Bawon DE, O’Hanilin A, Lewis CP, Winick NJ (2004) Nonadherence in adolescent oncology patients: primary data on psychological risk factors and relationships to outcome. J Clin Psychol Med Settings 11: 30 – 39

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33: 159 – 174

Lawrie I, Lloyd-Williams M, Taylor F (2004) How do palliative medicine physicians assess and manage depression. Palliat Med 18(3): 234 – 238

Le Ferve P, Devereux J, Smith S, Lawrie SM, Cornbleet M (1999) Screening for psychiatric illness in the palliative care inpatient setting: a comparison between the Hospital Anxiety and Depression Scale and the General Health Questionnaire-12. Palliat Med 13: 399 – 407

Lloyd-Williams M, Dennis M, Taylor F (2004) A prospective study to compare three depression screening tools in patients who are terminally ill. Gen Hosp Psychiatry 26(5): 384 – 389

Lloyd-Williams M, Dennis M, Taylor F, Baker I (2003a) Is asking patients in palliative care – are you depressed – appropriate? Prospective study. BMJ 327: 372 – 373

Lloyd-Williams M, Friedman T, Rudd N (2001) An analysis of the validity of the Hospital Anxiety and Depression scale as a screening tool in patients with advanced metastatic cancer. J Pain Symptom Manage 22(6): 990 – 996

Lloyd-Williams M, Riddleston H (2002) The stability of depression scores in patients who are receiving palliative care. J Pain Symptom Manage 24(6): 593 – 597

Lloyd-Williams M, Shiers C, Dowrick C (2007) The development of the Brief Edinburgh Depression Scale (BEDS) to screen for depression in patients with advanced cancer. J Affect Disord 99(1–3): 259 – 264

Lloyd-Williams M, Spiller J, Ward J (2003b) Which depression screening tools should be used in palliative care? Palliat Med 17(1): 40 – 43

Love AW, Grabsch B, Clarke DM, Bloch S, Kinsasse DW (2004) Screening for depression in women with metastatic breast cancer: a comparison of the Beck Depression Inventory Short Form and the Hospital Anxiety and Depression Scale. Aust N Z J Psychiatry 38(7): 526 – 531

McIntyre RS, Konarski ZJ, Mancini DA, Fulton KA, Parikh SV, Grigoridias S, Grupp LA, Bakish D, Filipeau MJ, Gorman C, Nemerofer CF, Kenneth SH (2005) Measuring the severity of depression and remission in primary care: validation of the HAMD-7 scale. CMAJ 173(11): 1327 – 1334

Meyer HA, Sinnott C, Seed PT (2003) Depressive symptoms in advanced cancer. Part 1. Assessing depression: the Mood Evaluation Questionnaire. Palliat Med 17: 596

Midgett AS, Stukel TA, Littenberg B (1993) A meta-analytic method for summarising diagnostic test performance: receiver operating characteristic- summary point estimates. Med Decis Making 13: 253 – 257

Miller RD, Walsh TD (1991) Psychosocial aspects of palliative care in advanced cancer. J Pain Symptom Manage 6: 24 – 29

Mitchell AJ (2006) Predictive summary index and other summary measures of diagnostic accuracy. In: The Encyclopedia of Medical Decision Making. Kattan WM, Sage (in press)

Mitchell AJ (2007) Pooled results from 38 analyses of the accuracy of distress thermometer and other ultra-short methods of detecting cancer-related mood disorder. J Clin Oncol 25: 4670 – 4681

Mitchell AJ, Coyne JC (2007) Do ultra-short screening instruments accurately detect depression in primary care? Br J Gen Pract 57(538): 151

Mitchell AJ, Kaar S, Coggan C, Herdman J (2008) Acceptability of common screening methods used to detect distress and related mood disorders-preferences of cancer specialists and non-specialists. Psychooncology 17(3): 226 – 236

Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Strupf D (1999) Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of reporting of meta-analyses. Lancet 354(9193): 1896 – 1900

Murphy JM (2002) Symptom scales and diagnostic schedules in adult psychiatry. In: Ming T Tsuang and Mauricio Tohen (eds) Textbook of Psychiatric Epidemiology, 2nd edn, Chapter 12, pp 273 – 332. John Wiley & Sons Inc.: New York, Wiley-Liss, ISBN 0-471-40974-X

Ng K, von Gunten CF (1996) Symptom and attitudes of 100 consecutive patients admitted to an acute hospice/palliative care unit. J Pain Symptom Manage 16: 307 – 315

Noorani NH, Montagnini M (2007) Recognizing depression in palliative care patients. J Palliat Med 10(2): 458 – 464

Ohno T, Noguchi W, Nakayama Y, Kato S, Tsujii H, Suzuki Y (2006) How do we interpret the answer neither when physicians ask patients with cancer ‘are you depressed or not’? J Pall Med 9(4): 861 – 865

Pai M, McCulloch M, Enanoria W, Colford JM (2004) Systematic reviews of diagnostic test evaluations: what’s behind the scenes? Evid Based Med 9: 101 – 103

Pallant JF, Miller RL, Tannant A (2006) Evaluation of the Edinburgh Postnatal Depression Scale using Rasch analysis. BMC Psychiatry 6: 2

Pascoe S, Edelman S, Kidman A (2000) Prevalence of psychological distress and use of support services by cancer patients at Sydney Hospitals. Aust N Z J Psychiatry 34: 785 – 791

Payne A, Barry S, Creedon B, Stone C, Sweeney C, O’Brien T, O’Sullivan K (2007) Sensitivity and specificity of a two-question screening tool for depression in a specialist palliative care unit. Palliat Med 21: 193 – 198

Pratheepawanit N, Salek MS, Finlay IG (1999) The applicability of quality-of-life assessment in palliative care: combining two quality-of-life measures. Palliat Med 13: 325 – 334

Reuter K, Raugust S, Bengel J, Harter M (2004) Depressive symptom patterns and their consequences for diagnosis of affective disorders in cancer patients. Support Care Cancer 12(12): 864 – 870

Robinson JA, Crawford GB (2005) Identifying palliative care patients with symptoms of depression: an algorithm. Palliat Med 19(4): 278 – 287

Serrano-Duen M, Soledad SM (2008) Concurrent validation of the 21-item Hamilton Depression Rating Scale vs the DSM-IV diagnostic criteria to assess depression in patients with Parkinson’s disease: an exploratory analysis. Parkinsonism Relat Disord 14(3): 233 – 238

Sharpe M, Strong V, Allen K, Rush R, Postma K, Tulloh A, Maguire P, House A, Ramirez A, Culi A (2004) Major depression in outpatients attending a regional cancer centre: screening and unmet treatment needs. Br J Cancer 90: 314 – 320

Skarstein J, Aas N, Fossad SA, Skolvand E, Dahl AA (2000) Anxiety and depression in cancer patients: relations between the Hospital Anxiety and Depression Scale and the European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire. J Psychosom Res 49: 27 – 34

Solner W, DeVries A, Steixner E, Lukas P, Sprinzl G, Rumpold G, Maislinger S (2001) How successful are oncologists in identifying patient distress, perceived social support, and need for psychosocial counselling? Br J Cancer 84(2): 179 – 185

Spitzer RL, Kurt Kroenke K, Williams JBW, the Patient Health Questionnaire Primary Care Study Group (1999) Validation and Utility of a Self-report Version of PRIME-MD. The PHQ Primary Care Study. JAMA 282: 1757 – 1744


Stark D, Kiely M, Smith A, Velikova G, House A, Selby P (2002) Anxiety disorders in cancer patients: their nature, associations, and relation to quality of life. *J Clin Oncol* 20(14): 3137 – 3148

Steer RA, Cavallieri TA, Leonard DM, Beck AT (1999) Use of the beck depression inventory for primary care to screen for major depression disorders. *Gen Hosp Psychiatry* 21: 106 – 111

Stiefel F, Trill MD, Berney A, Olarte JMN, Razavi D (2001) Depression in palliative care: a pragmatic report from the expert working group of the European Association for Palliative Care. *Support Care Cancer* 9(7): 477 – 488

Thekkumpurath P, Venkateswaran C, Kumar M, Bennett M (2007) Screening for psychological distress in palliative care: validation of screening questionnaires in a hospice population in Leeds. *Psychooncology* 16(9): S257 (Suppl. S) P2–P154

Trask P (2004) Assessment of depression in cancer patients. *J Natl Cancer Inst* 2004(32): 80 – 92

van Marwijk HW, Wallace P, de Bock GH, Hermans J, Kaptein AA, Mulder JD (1995) Evaluation of the feasibility, reliability and diagnostic value of shortened versions of the geriatric depression scale. *Br J Gen Pract* 45(393): 195 – 199

Vignaroli E, Pace EA, Willey J, Palmer JL, Zhang T, Bruera E (2006) The Edmonton Symptom Assessment System as a screening tool for depression and anxiety. *J Palliat Med* 9(2): 296 – 303

Zabora J, Brintzenhofeszoc K, Curbow B, Hooker C, Piantadosi S (2001) The prevalence of distress by cancer site. *Psychooncology* 10: 19 – 28

A Bayesian meta-analysis

Aj Mitchell