Screening tools used for measuring depression among people with Type 1 and Type 2 diabetes: a systematic review

Received Date : 16-Feb-2011

Revised Date : 20-May-2011

Accepted Date : 03-Aug-2011

Article type : Review

Review

Screening tools used for measuring depression among people with Type 1 and Type 2 diabetes: a systematic review

T. Roy[1], C. E. Lloyd[2], F. Pouwer[3], R. I. G. Holt[4] and N Sartorius[5]

[1]Division of Social Research in Medicines and Health, University of Nottingham, Nottingham, [2]Faculty of Health and Social Care, The Open University, Milton Keynes, UK, [3]Centre of Research on Psychology in Somatic Diseases (C*o*RPS), Tilburg University, the Netherlands, [4]Developmental Origins of Health and Disease Division, University of Southampton Faculty of Medicine, Southampton, UK and [5]Association for the Improvement of Mental Health Programmes, Geneva, Switzerland

Correspondence to: Dr C. E. Lloyd, Faculty of Health and Social Care, The Open University, Milton Keynes MK7 6AA, UK. E-mail: C.E.Lloyd@open.ac.uk

Short title: Screening tools used for measuring depression in diabetes

Abstract

Background Depression is common in patients with Type 1 or Type 2 diabetes, has a strong negative impact on the quality of life of patients and is associated with poor outcomes and higher mortality rates. Several guidelines encourage screening of patients with diabetes for depression. It is unclear which depression screening tools are currently being used in people with diabetes and which are most appropriate.

Methods A systematic review was conducted to examine which depression screening instruments are currently being used in diabetes research, and the operating characteristics of these tools in diabetes populations. Literature searches for the period January 1970 to October 2010 were conducted using MEDLINE, PSYCH-INFO, ASSIA, SCOPUS, ACADEMIC SEARCH COMPLETE, CINAHL and SCIENCE DIRECT.

Results Data are presented for the 234 published studies that were examined. The Beck Depression Inventory and the Centre for Epidemiologic Studies Depression Scale were the most popular screening tools (used in 24% and 21% of studies). Information on the cultural applicability of screening tools was mostly unavailable and, where reported, included only details of the language translation process. A small number of studies reported reliability data, most of which showed moderate–good sensitivity and specificity but a high rate of false positives.

Conclusions Although a range of depression screening tools have been used in research, there remains few data on their reliability and validity. Information on the cultural applicability of these instruments is even scantier. Further research is required in order to determine the suitability of screening tools for use in clinical

practice and to address the increasing problem of co-morbid diabetes and depression.

Keywords depression, diabetes, screening tools

Introduction

Recently, there has been a heightened interest in the psychological well-being of people with diabetes. Current epidemiological evidence suggests that at least one third of people with diabetes suffer from sub-threshold or major depressive disorders [1–4]. Depression is associated with poor self-care, impaired glycaemic control, poor microvascular and macrovascular outcomes, higher healthcare costs, and compromises quality of life [5–10]. Effective psychological and pharmacological treatments are available, but an important barrier to treatment is the under-recognition of depression [11, 12]. For example, in one study, diabetes nurse specialists failed to recognize and document high levels of anxiety, depression or diabetes-specific emotional distress in approximately 75% of patients who had established high scores on the Hospital Anxiety and Depression Scale (HADS) or Problem Areas in Diabetes (PAID) Scale [12]. Although the stigma of mental illness has been much reduced in recent years, this remains a further barrier to the identification of depression as well as acceptance of diagnosis and treatment.

Regular screening for depression is recommended by several professional bodies, including the International Diabetes Federation, American Diabetes Association and the UK National Institute for Health and Clinical excellence. However, in spite of the huge impact of co-morbid depression and diabetes on the individual and its importance as a public health problem, questions still remain as to the most appropriate ways of identifying people suffering from depression. Brief screening instruments or questionnaires are quick and simple to administer and most have been shown to approximate with clinically significant levels of depressive disorder at certain cut-points or scores in the general population. It is unclear whether these tools are appropriate for people with diabetes, not least because there is overlap between the symptoms of depression and the symptoms of diabetes and its long-term complications. These symptoms include tiredness, lethargy, lack of energy, sleeping difficulties and appetite changes. Diabetes occurs more commonly in certain ethnic groups and there are emerging data suggesting concomitant high rates of depression [13, 14]. However, it remains unclear whether current screening instruments are suitable for use in other cultures of countries or in ethnic minority groups living in English-speaking countries.

It is important to recognize that the use of screening instruments is only one step towards the identification of individuals who may be depressed. In the UK and in some other European countries (for example, the Netherlands), a stepped care approach is recommended whereby treatment decisions are based on the severity of the symptoms. In order to assess the presence and severity of depressive symptoms, the general primary healthcare professional must first use two brief screening questions about possible depressive symptoms during the past month, followed by the use of one of three instruments [the Patient Health Questionnaire

(PHQ-9), the Beck Depression Inventory (BDI-II) and the Hospital Anxiety and Depression Scale (HADS)] to detect the extent of symptoms if the patient has answered in the affirmative to either of the two questions. All three of these scales measure symptoms that have been found to approximate clinical levels of depression [9, 15,16]. The third step, necessary in order to confirm any diagnosis of depression, is a full clinical interview.

A range of self-report screening tools have been devised for use in both research and clinical practice, many of which measure symptoms that approximate clinical levels of depression; however, the quality of these tools has not been assessed systematically for use in individuals with diabetes. This systematic review has addressed the following questions: (1) which depression screening tools are currently being used in diabetes research?; (2) which translations are available?; (3) do symptoms measured by the screening tools overlap with diabetes symptoms?; (4) what is the sensitivity and specificity of the tools in detecting depression in different studies? Given that healthcare practitioners encounter many different patient populations, it is not possible to make specific recommendations as to the criteria that should be used by all practitioners when identifying an appropriate screening tool. However, it is intended that this review will help inform practitioners in a range of settings and provide information about the suitability of different depression screening tools in people with diabetes

Methods

Literature searches for the period from January 1970 to October 2010 were conducted using MEDLINE, PSYCH-INFO, ASSIA, SCOPUS, ACADEMIC SEARCH COMPLETE, CHINAL and SCIENCE DIRECT. Three separate searches were performed for specific terms included in the keywords or title. The three search strategies and corresponding terms were: strategy I—'depression' AND 'diabetes' AND ('measurement tool' OR 'scale'); strategy II—'depression' AND 'diabetes' AND 'questionnaire'; ('depression' OR 'anxiety' OR 'mood disorder' OR 'psychological distress') AND ('measurement' OR 'assessment' OR 'evaluation' OR 'scale' OR 'tool' OR 'rating' OR 'inventory'); strategy III—'depression scale' AND ('measurement' OR 'assessment' OR 'evaluation' OR 'rating' OR 'inventory' OR 'monitoring'). Within each database, articles that appeared in all three searches were retained for further examination.

Only articles published in English were considered. No unpublished material was reviewed. Based on titles and abstracts, potentially relevant studies were identified and the full texts of these articles were examined for final determination of relevance. We included studies comparing depression-screening tools given the presence or absence of depressive symptoms and also those which assessed depression measurement scales for diabetes. As implicit in the search strategy adopted, for a study to be included, at least an adult sub-population within the study was required to have known diabetes and be specifically examined with respect to depressive symptomatology and a screening tool, questionnaire or measurement scale used to assess depression. Studies were excluded if they involved children or adolescents (an additional search limitation was used to restrict articles: 'search limited to Adult <

18 to 64 years > AND Aged < 65+ years >'); or if there were no explicit definitions of depressive symptoms. In addition to observational studies, we included randomized trials conducted in adults with Type 1 or Type 2 diabetes and depression, that attempted either (1) to reduce the degree of depressed mood through an intervention (e.g. intervention in depressed individuals with diabetes or other chronic illness) or (2) improve diabetes care outcome activity through a depression-specific intervention.

Using the databases described above (Fig. 1), we identified 1784 papers through the three search strategies. Of these, 1631 were available in English. Review of abstracts and titles identified 938 studies likely to be eligible for more complete review and the full texts of 234 studies were finally examined. These studies were categorized according to whether they reported (1) sensitivity/specificity and reliability data, (2) cultural applicability/validity of translations/culturally adapted scales, (3) investigated populations of Type 1 or Type 2 diabetes, and the type of setting in which the study was conducted.

For the purpose of this report, we classified study type according to the study design reported; i.e. (1) prevalence study—reported only prevalence/ incidence of depression; (2) experimental study—to test cause-and-effect relationships between variables. These were further categorized as either case–control study or randomized controlled trial (RCT); (3) observational study—designed to draw inferences about the possible effect of exposure on outcomes without the investigator's intervention, which was further categorized either as a case–control study or cohort study; and (4) validity/reliability study—a study designed specifically to test reliability and/or validity of an instrument.

Further, sensitivity was defined as how well a screening tool performs in terms of identifying all the people with symptoms of depression ('true positives'). Specificity was defined as the ability of a screening tool to identify people with symptoms of depression rather than other psychological morbidities.

We also considered both the positive and negative predictive values for each screening tool. These values are of considerable interest for clinical practice. A low positive predictive value (PPV) is associated with a high rate of false positives and a low negative predictive value (NPV) is associated with a high rate of false negatives. If the positive predictive value is low, the healthcare professional has to deal with numerous false positives, causing a lot of unnecessary additional diagnostic tests or referrals to mental health specialists and anxiety for the patient.

Results

Frequency of questionnaire use

Table 1 shows the full range of scales that have been used to screen for the symptoms of depression in people with diabetes. The Beck Depression Inventory (BDI) was the screening tool most often cited in 24% (55/234) of studies [15]; followed closely by the Centre for Epidemiologic Studies Depression Scale (CES-D) (21%; $n = 49$) [16]. The BDI was updated and renamed the BDI-II [17]. Both versions consist of 21 items; however, the later version asks the respondent to indicate how they have been feeling during the past 2 weeks (rather than 1 week) in line with the new edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) that contained a number of changes to the diagnostic criteria for major depression.

The CES-D has also been modified from a 20-item scale to a 10-item scale, although the longer version has been much more commonly used. The third most frequently cited screening tool was the Problem Areas in Diabetes (PAID) Scale (12%; 28/234 studies) [18–19]. Although not strictly a tool for measuring symptoms of depression, the PAID has been recognized in recent years for its utility in identifying symptoms of distress that frequently overlap with depressive symptomatology [20–22].

Two further depression screening tools frequently cited in the published literature were the Patient Health Questionnaire (PHQ) (11%; 25/234) [23] and the Hospital Anxiety and Depression Scale (HADS) (10%; 24/234) [24]. There are three versions of the PHQ (9-item, 8-item and 2-item); with the PHQ-9 being the instrument most commonly cited. The HADS is a brief self-complete scale measuring the presence of symptoms of both anxiety (seven items) and depression (seven items) during the past week [24]. It has been widely used in a range of populations, both with and without diabetes.

Other scales less frequently used included the Zung Self-Rating Depression Scale (SDS) (8%; 18/234) [25], which has been used more commonly in people with Type 2 rather than Type 1 diabetes. The Medical Outcomes Survey Short Form (SF) (either 36- or 12-item) (6.4%; 15/234) [26], a generic measure of quality of life with a mental health sub-scale, has also been utilized in a number of studies of diabetes and depression.

The Hamilton Depression Scale [27], the Geriatric Depression Scale [28] and the Well-Being Index [World Health Organization (WHO)-5] [29-30], among others, have only rarely been used (see Table 1) despite having been used extensively in populations without diabetes. A range of studies have been conducted to investigate

depression and diabetes co-morbidity, although these have predominantly been observational rather than experimental or intervention studies. Although studies have not always distinguished between type of diabetes, for those where this information was available the BDI has been more frequently used in people with Type 1 diabetes, whereas a greater number of studies using the CES-D, PHQ-9, and the Zung SDS have been conducted in people with Type 2 diabetes.

Depression was assessed using various techniques. As shown in Table 1, in most of the studies (80%; 187/234) depression was assessed using self-report questionnaires (e.g. BDI, CES-D, PAID, PHQ-9, Zung SDS, SF-36/12 etc) either by self-administered methods or standard assisted methods. The remainder (n = 47; 20%) of the studies assessed depression using diagnostic criteria and interviews conducted by a researcher/care provider (e.g. HADS, HRDS and MADRS etc).

Cultural adaptations/translated screening tools

Nearly half (102/234; 44%) of the publications identified in our systematic review used a translated version of a screening instrument; however, information on the translation process and data on the cultural applicability of the tools were lacking. Consequently, the available data are incomplete. Published studies of people with diabetes and depression were most commonly conducted using English versions of the relevant survey instrument; with a smaller number of studies conducted in a range of other languages (see Table 2). Of the 234 studies we examined, 62 (27%) reported using a translated version of a screening tool, but did not provide any further information; 31(13%) gave a reference for the origin of the translation, and only 10 (4%) studies provided more in-depth information on the translation procedure used and gave information on the validity of the tool. Of these 10 studies, four used

the PAID, two used the CES-D and the BDI, WHO-5, HADS and WBQ-12 were each used once. Only four studies [38, 45, 68, 80] mentioned the cultural applicability of their questionnaire, although the use of the terms ‘translation’ and ‘cultural applicability’ tended to be used interchangeably.

Symptom overlap

With the exception of the HADS, the General Health Questionnaire (GHQ) and the WHO-5, all instruments contained some items that could be confounded with the symptoms of poorly controlled diabetes; for example, the BDI includes items on tiredness and weight loss. Questions relating to symptoms such as loss of appetite and restless sleep (which may also be related to symptoms of diabetes) are included in the CES-D and the BDI, whilst the PHQ-9 contains items such as lack of energy, poor appetite and overeating.

Sensitivity, specificity and reliability data

Only 39 (17%) of the 234 studies reported reliability and validity data related to the depression questionnaire used. Of those, 23 (10%) reported only the reliability data and did not provide any further information on sensitivity and specificity. Studies in this category mostly performed internal consistency (usually measured with Cronbach's alpha) as a measure of reliability, and only a small number conducted test–retest studies to determine reliability (see section 2 of Table 3). Only 16 (7%) studies provided more in-depth information on validity and included data on reliability of the screening tool. To establish the criterion validity, studies of this category calculated sensitivity, specificity, the positive predictive value and negative predictive value of the tool (Table 3, section 1). Of these 16 studies, three used more than one screening tool and four used the PHQ-9 (mean sensitivity/specificity = 82/68%,

respectively). The WHO-5 (mean sensitivity/specificity = 76/48%), the HADS (mean sensitivity/specificity = 77/66%), and the Zung-SDS (mean sensitivity/specificity = 86/76%) were all used in three studies, and the CES-D (mean sensitivity/specificity = 90/78%) and the BDI (mean sensitivity/specificity = 85/79%) were used in two studies.

A good number of studies (108/234; 46%) relied on previous study reports of reliability and validity data to inform their choice of a particular instrument, justifying it as being a reliable and well-validated instrument for screening for depression. Given the limited data and lack of comparisons between instruments, it is not possible to conclude whether one instrument can be said to be more reliable than another; however, most (but not all) reports have demonstrated moderate–high sensitivity and specificity. As can be seen from Table 3, most studies reported a high rate of false positives (low positive predictive value) but a low rate of false negatives (high negative predictive value).

We further tabulated the prevalence of depression for the 16 studies that reported in-depth information on sensitivity and specificity data with positive and negative predictive values of the screening tool to investigate whether positive and negative predictive values were different in samples with varying prevalence rates (Table 3). Based on the limited data available, we observed a high rate of false positives for many of the screening tools (e.g. CES-D, PHQ-9, HADS, BDI and WHO-5), as indicated by low and varying positive predictive values (23–56% in all studies) and wide variations in depression prevalence using the CES-D (27–66%), PHQ-9 (19–56%) and HADS (22–41%) [37–43, 46–47]. Although, the studies reported wide variations in the prevalence of depression, most of them also reported high negative predictive values (> 90% in all studies). The studies using the BDI to screen for

depression reported an almost similar prevalence (37–38%) rate with varying positive predictive values (22–89%), but again they reported a similarly high negative predictive value (≥ 90%) [44–45]. Data available for the WHO-5 and Zung-SDS were incomplete and did not allow us to make such comparisons. Only one study reported sensitivity to change over time, and this was in relation to the PAID [81]. In this paper the authors note that, although the data available for analysis were from pilot studies, there was strong support for the responsiveness of the PAID.

Discussion

Our systematic review of the literature has shown that the CES-D and the BDI are the most frequently cited screening tools used in published research in diabetes and depression. However, the BDI, and also the HADS, despite being recommended for use in primary care in the UK, are rarely used in clinical practice, possibly because of prohibitive costs in using these instruments [78]. A recent study comparing several depression screening tools, including the HADS and the CES-D, found that the latter instrument was the best predictor of depression in people with Type 2 diabetes [79]. This study also examined the overlap of symptoms of diabetes and depression and reported that the CES-D had the best ability to discriminate between depression and other non-depressive symptoms. This has implications for clinical practice; not least because a number of studies have previously shown a high correlation between depression symptoms and symptoms of diabetes-related distress such as those measured by the PAID, as well as with symptoms of uncontrolled diabetes or the presence of diabetes complications (14, 80–83]. Identifying the underlying cause of symptoms is important to detect in order to inform treatment decisions. In the case of high depression scores, further diagnostic work should be performed by a mental

health specialist, such as a medical psychologist or liaison psychiatrist, who can distinguish between adjustment disorder, medically explained depression and major depressive disorder [84].

Also recommended for use in primary care in the UK, the PHQ-9 has been found to be acceptable in a number of different patient groups, including people with but also without diabetes, possibly because of its relative brevity, ease of interpretation and concurrence with the diagnostic criteria of major depression. It has begun to be used more extensively in research settings; for example, in the Pathways Study [42, 84]. However, despite being available in a number of different languages, we found only two published reports of the PHQ-9 being used in people with diabetes in languages other than English [40, 42]. Indeed, few of the studies we examined included details about the methods used to translate and validate the screening tool used in languages other than English, and issues remain as to the utility of particular instruments in different cultural or minority ethnic groups. [84]. Previous research has indicated that rates of depression may vary depending on the language in which the assessment instrument is applied [85]. Comparison of depression rates between different countries or cultures may therefore be problematic, unless comparable methods have been used to collect the data [85].

Given the limited data available and the lack of any published comparisons between instruments, it is not possible to conclude whether the performance of one scale as a depression screener can be said to be better than another, or if any variations in prevalence rates are dependent on the study setting. However, we observed wide variations in the prevalence of depression with low and varying positive predictive values but with high negative predictive values for many of the screening tools (e.g. CES-D, PHQ-9, HADS, BDI, Zung-SDS and WHO-5); this provides confidence in

these scales as valid screening tools for depression [37–-53]. It is important to investigate whether the screening performance of a specific scale differs with regard to symptom overlap, study setting and construct measures (depressive symptoms vs. well-being or diabetes-related distress). From the studies included in our review, however, it is not possible to elaborate on the screening performance of the identified screening tools. Further research should target these issues in order to provide evidence on the screening performance of the scales so as to inform clinical practice.

Only a small number of studies included in our systematic review have reported sensitivity and specificity and/or reliability data of the screening tools used, and most have relied on previous studies to inform their choice of a particular instrument. It is therefore difficult to evaluate the overall sensitivity and specificity of these screening tools. Although there were wide variations in sensitivity and specificity data across the studies, the published reports suggests that the PHQ-9 is a valid and reliable tool with high sensitivity and specificity. Indeed, where there were data reported, most of the scales displayed good levels of sensitivity and acceptable specificity.

At the same time, there may be some concerns with regard to the reported high rate of false positives of many of the screening tools, as indicated by a low positive predictive value. This has important implications for clinical practice, as it may lead to inappropriate referrals or follow-up for depression, as well as patient anxiety. Further, we also acknowledge the limitation of reviewing only diabetes studies that include depression measurement, which may not be sufficient to identify which depression screeners are actually used in clinical practice. To our knowledge, however, there are no reports describing this in primary care, although recently an observational study in secondary care has demonstrated the feasibility of using a computerized

assessment of psychological well-being (using the WHO-5 and the PAID) during patients' annual review for diabetes [86]. Further research is clearly required in order to inform choice of instrument and its interpretation in actual clinical practice.

It is important to note that screening tools are not intended for use as a diagnostic tool nor as instruments to assess severity of symptoms, but rather as a first step towards a definitive diagnosis, usually using a clinical interview. Screening should be embedded within a managed-care approach for depression. Indeed, approaches where screening has been followed by the monitoring of depression symptoms and more intensive treatment of depression appear to be successful [3, 20].

Most people with diabetes are cared for by their primary care physician and it is in this setting that there are key opportunities for screening and providing care for individuals with co-morbid diabetes and depression.

Competing interests

Nothing to declare.

Acknowledgements

The authors wish to acknowledge the support and encouragement of the members of the Dialogue on Diabetes and Depression (DDD)—an international collaborative effort addressing problems related to the co-morbidity of diabetes and depression aiming to review the evidence and experience relevant to this co-morbidity and to support research, training and education that might be necessary to deal with the problems resulting from it (http://www.diabetesanddepression.org/). This report has been carried out within the framework of the Dialogue on Diabetes and Depression (DDD), a project coordinated by Association for the Improvement of Mental Health

Programmes (AMH) The AMH has so far received support for the project from Eli Lilly and Company, the Lundbeck Neuroscience Institute, the Novo Nordisk Company, numerous experts and several universities. This report was partly funded by DDD.

References

[1] Barnard K, Skinner T, Peveler R. The prevalence of co-morbid depression in adults with Type 1 diabetes: systematic literature review. *Diabet Med* 2006; **23:** 445–448.

[2] Ali S, Stone M, Peters J, Davies M, Khunti K. The prevalence of co-morbid depression in adults with Type 2 diabetes: a systematic review and meta-analysis. *Diabet Med* 2006; **23:** 1165–1173.

[3] Pouwer F, Gellhoed-Duijvenstein P, Tack CJ, Bazelmans E, Beekman AJ, Heine RJ *et al*. Prevalence of co-morbid depression is high in outpatients with Type 1 or Type 2 diabetes mellitus: results from three outpatient clinics in the Netherlands. *Diabet Med* 2010; **27:** 217–224.

[4] Nouwen A, Winkley K, Twisk J, Lloyd CE, Peyrot M, Ismail K *et al*.; European Depression in Diabetes (EDID) Research Consortium. Type 2 diabetes mellitus as a risk factor for the onset of depression: a systematic review and meta-analysis. *Diabetologia* 2010; **53:** 2480–2486.

[5] Egede LE, Ellis C, Grubaugh AL. The effect of depression on self-care behaviors and quality of care in a national sample of adults with diabetes. *Gen Hosp Psychiatry* 2009; **31:** 422–427.

[6] Lustman PJ, Anderson RJ, Freedland KE, de Groot M, Carney RM, Clouse RE. Depression and poor glycemic control: a meta-analytic review of the literature. *Diabetes Care* 2000; **23:** 934–943.

[7] Egede LE, Nietert PJ, Zheng D. Depression and all-cause and coronary mortality among adults with and without diabetes. *Diabetes Care* 2005; **28:** 1339–1345.

[8] Lin EH, Heckbert SR, Rutter CM, Katon WJ, Ciechanowski P, Ludman EJ *et al*. Depression and increased mortality in diabetes: unexpected causes of death. *Ann Fam Med* 2009; **7:** 414–421.

[9] Katon W, Fan MY, Unützer J, Taylor J, Pincus H, Schoenbaum M. Depression and diabetes: a potentially lethal combination. *J Gen Intern Med* 2008; **23:** 1571–1575.

[10] Schram MT, Baan CA, Pouwer F. Depression and quality of life in patients with diabetes: a systematic review from the European Depression in Diabetes (EDID) research consortium. *Curr Diabetes Rev* 2009; **5:** 112–119.

[11] van der Feltz-Cornelis CM, Nuyen J, Stoop C, Chan J, Jacobson AM, Katon W *et al*. Effect of interventions for major depressive disorder and significant depressive symptoms in patients with diabetes mellitus: a systematic review and meta-analysis. *Gen Hosp Psychiatry* 2010; **32:** 380–395.

[12] Pouwer F, Beekman AT, Lubach C, Snoek FJ. Nurses' recognition and registration of depression, anxiety and diabetes-specific emotional problems in outpatients with diabetes mellitus. *Patient Educ Couns* 2006; **60:** 235–240.

[13] Pouwer F. Should we screen for emotional distress in patients with diabetes? *Nat Rev Endocrinol* 2009; **XX:** XX–XX.

[14] Asghar S, Hussain A, Ali SM, Khan AK, Magnusson A. Prevalence of depression and diabetes: a population-based study from rural Bangladesh. *Diabet Med* 2007; **24:** 872–877.

[15] Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. 'An inventory for measuring depression'. *Arch Gen Psychiatry* 1961; **4:** 561–571.

[16] Radloff L.S. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas* 1977; **1:** 385–401.

[17] Beck AT, Steer RA and Brown GK. *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation, 1996.

[18] Polonsky WH, Anderson BJ, Lohrer PA, Welch G, Jacobson AM, Aponte JE *et al*. Assessment of diabetes-related distress. *Diabetes Care* 1995; **18:** 754–760.

[19] Polonsky WH, Fisher L, Earles J, Dudl RJ, Lees J, Mullan J *et al*. Assessing psychosocial distress in diabetes: development of the diabetes distress scale. *Diabetes Care* 2005; **28:** 626–631.

[20] Pouwer F, Tack CJ, Geelhoed-Duijvestijn PH, Bazelmans E, Beekman AT, Heine RJ *et al*. Limited effect of screening for depression with written feedback in outpatients with diabetes mellitus: a randomised controlled trial. *Diabetologia* 2011; (In Press) DOI: 10.1007/s00125-010-2033-0

[21] Hermanns N, Mahr M, Kulzer B, Skovlund SE, Haak T. Barriers towards insulin therapy in type 2 diabetic patients: results of an observational longitudinal study. *Health Qual Life Outcomes* 2010; **8:** 113.

[22] Hermanns N, Kulzer B, Krichbaum M, Kubiak T, Haak T. How to screen for depression and emotional problems in patients with diabetes: comparison of screening characteristics of depression questionnaires, measurement of

diabetes-specific emotional problems and standard clinical assessment. *Diabetologia* 2006; **49:** 469–477.

[23] Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient Health Questionnaire. *J Am Med Assoc* 1999; **282:** 1737–1744.

[24] Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand* 1983; **67:** 361–370.

[25] Zung WW. 'A self-rating depression scale'. *Arch Gen Psychiatry* 1965; **12:** 63–70.

[26] Ware JE Jr. Snow KK, Kosinski M, Gandek B. *SF-36 Health Survey. Manual and Interpretation Guide*. Boston: The Health institute, New England Medical Center, 1993.

[27] Hamilton, M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960; **23**: 56–62.

[28] Yesavage JA. Geriatric Depression Scale. *Psychopharmacol Bull* 1988; **24:** 709–711.

[29] World Health Organization Regional Office for Europe and the International Diabetes Federation, Europe. Diabetes mellitus in Europe: a problem at all ages and in all countries. A model for prevention and self care. Meeting. Giorn Ital Diabetol 1990; 10 (suppl)

[30] Bech P, Gudex C, Staehr Johansen K. The WHO (Ten) Well-Being Index: Validation in Diabetes. *Psychother Psychosom* 1996; **65:** 183–190.

[31] Bradley C. The well-being questionnaire. In: Bradley C, ed. *Handbook of Psychology and Diabetes*. XXXXX: Harwood Academic, 1994.

[32] Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979; **134:** 382–389.

[33] Derogatis LR, Savitz KL. The SCL-90-R and the Brief Symptom Inventory (BSI) in primary care. In: Maruish ME, ed. *Handbook of Psychological Assessment in Primary Care Settings*. Volume 236. Mahwah, NJ: Lawrence Erlbaum Associates, 2000: 297–334.

[34] Goldberg DP, Blackwell B. Psychiatric illness in general practice. A detailed study using a new method of case identification. *Br Med J* 1970; **1:** 439–443.

[35] Spitzer RL. 'The diagnostic status of homosexuality in DSM-III: a reformulation of the issues'. *Am J Psychiatry* 1981; **138:** 210–215.

[36] The Euro-Qol Group. *Euro-Qol – A New Facility for the Measurement of Health-Related Quality of Life.* York: Centre for Health Economics, University of York, 1990.

[37] Wagner J, Tennen H, Mansoor G and Abbott G. Endothelial dysfunction and history of recurrent depression in postmenopausal women with Type 2 diabetes: a case–control study. J Diabetes Complications 2009; **23:** 18–24.

[38] Stahl D, Sum CF, Lum SS, Liow PH, Chan YH, Verma S *et al*. Screening for depressive symptoms: validation of the Center for Epidemiologic Studies depression scale (CES-D) in a multi-ethnic group of patients with diabetes in Singapore. *Diabetes Care* 2008; **31:** 1118–1119.

[39] Kahn LS, Fox CH, McIntyre RS, Tumiel-Berhalter L, Berdine DE, Lyle H. Assessing the prevalence of depression among individuals with diabetes in a Medicaid managed-care program. *Int J Psychiatr Med* 2008; **38:** 13–29.

[40] Lamers F, Jonkers CCM, Bosma H, Penninx BWJH Knottnerus JA, Van Eijk JTM. Summed score of the Patient Health Questionnaire-9 was a reliable and

valid method for depression screening in chronically ill elderly patients. J Clin Epidemiol 2008; **61:** 679–687.

[41] Ackermann RT, Rosenman MB, Downs SM, Holmes AM, Katz BP, LI J *et al*. Telephonic case-finding of major depression in a Medicaid chronic disease management program for diabetes and heart failure. *Gen Hosp Psychiatry* 2005; **27:** 338–343.

[42] van Steenbergen-Weijenburg KM, de Vroege L, Ploeger RR, Brals JW, Vloedbeld MG, Veneman TF *et al*. Validation of the PHQ-9 as a screening instrument for depression in diabetes patients in specialized outpatient clinics. *BMC Health Serv Res* 2010; **10:** 235.

[43] Sultan S, Luminet O, Hartemann A. Cognitive and anxiety symptoms in screening for clinical depression in diabetes. A systematic examination of diagnostic performances of the HADS and BDI-SF. *J Affect Disord* 2010; **123:** 332–336.

[44] Lustman PJ, Clouse RE, Griffith LS, Carney RM, Freedland KE. Screening for depression in diabetes using the Beck Depression Inventory. *Psychosom Med* 1997; **59:** 24–31.

[45] Nayani S. The evaluation of psychiatric illness in Asian patients by the Hospital Anxiety Depression Scale. *Br J Psychiatry* 1989; **155:** 545–547.

[46] Engum A. The role of depression and anxiety in onset of diabetes in a large population-based study. *J Psychosom Res* 2007; **62:** 31–38.

[47] Xu L, Ren J, Cheng M, Tang K, Dong M, Hou X *et al*. Depressive symptoms and risk factors in Chinese persons with type 2 diabetes. *Arch Med Res* 2004; **35:** 301–307.

[48] Yoshida S, Hirai M, Suzuki S, Awata S, Oka Y. Neuropathy is associated with depression independently of health-related quality of life in Japanese patients with diabetes. Psychiatry Clin Neurosci 2009; **63:** 65–72.

[49] Biggs JT, Wylie LT, Ziegler VE. Validity of the Zung Self-Rating Depression Scale. *Br J Psychiatry* 1978: **132:** 381–385.

[50] MCGuire BE, Morrison TG, Hermanns N, Skovlund S, Eldrup E, Gagliardino J *et al*. Short-form measures of diabetes-related emotional distress: the Problem Areas in Diabetes Scale (PAID)-5 and PAID-1. *Diabetologia* 2010; **53:** 66–69.

[51] Awata S, Bech P, Yoshida S, Hirai M, Suzuki S, Yamashita M *et al*. Reliability and validity of the Japanese version of the World Health Organization-Five Well-Being Index in the context of detecting depression in diabetic patients. *Psychiatry Clin Neurosci* 2007; **61:** 112–119.

[52] Newnham EA, Hooke GR, Page AC. Monitoring treatment response and outcomes using the World Health Organization's Wellbeing Index in psychiatric care. *J Affect Disord* 2010; **122:** 133–138.

[53] Furuya M, Hayashino Y, Tsujii S, Ishii H, Fukuhara S. Comparative validity of the WHO-5 Well-Being Index and two-question instrument for screening depressive symptoms in patients with type 2 diabetes. *Acta Diabetol* 2010; **XX:** XX–XX. DOI: 10.1007/s00592-010-0219-z

[54] Bai Y, Chiou C, Chang Y, Lam H. Correlates of depression in type 2 diabetic elderly patients: a correlational study. *Int J Nurs Stud* 2008; **45:** 571–579.

[55] Keawe'aimoku Kaholokula J, Haynes SN, Grandinetti A, Chang HK. Biological, psychosocial, and sociodemographic variables associated with

depressive symptoms in persons with type 2 diabetes. *J Behav Med* 2003; **26:** 435–458.

[56] Zauszniewski JA, Graham GC. Comparison of short scales to measure depressive symptoms in elders with diabetes. *West J Nurs Res* 2009; **31:** 219–234.

[57] Pibernik-Okanovic M, Begic D, Peros K, Szabo S, Metelko Z. Psychosocial factors contributing to persistent depressive symptoms in type 2 diabetic patients: a Croatian survey from the *European Depression in Diabetes Research Consortium. J* Diabetes Complications 2008; **22:** 246–253.

[58] Wagner J, Tsimikas J, Abbott G, De Groot M, Heapy A. Racial and ethnic differences in diabetic patient-reported depression symptoms, diagnosis, and treatment. *Diabetes Res Clin Pract* 2007; **75:** 119–122.

[59] Grandinetti A, Kaholokula JK, Crabbe KM, Kenui CK, Chen R, Chang HK. Relationship between depressive symptoms and diabetes among native Hawaiians. *Psychoneuroendocrinology* 2000; **25:** 239–246.

[60] Park H, Hong Y, Lee H, Ha E, Sung Y. Individuals with type 2 diabetes and depressive symptoms exhibited lower adherence with self-care. *J Clin Epidemiol* 2004; **57:** 978–984.

[61] Wagner JA, Abbott GL, Heapy A, Yong L. Depressive symptoms and diabetes control in African Americans. *J Immigr Minor Health* 2009; **11:** 66–70.

[62] Leonardson GR, Daniels MC, Nees FK, Kemper E, Mihura JL, Koplin BA *et al*. Validity and reliability of the general well-being schedule with Northern Plains American Indians diagnosed with type 2 diabetes mellitus. *Psychol Rep* 2003; **93:** 49–58.

[63] Kagee A. Symptoms of depression and anxiety among a aample of South African patients living with a chronic illness. *J Health Psychol* 2008; **13:** 547–555.

[64] Lee H, Chapa D, Kao C, Jones D, Kapustin J, Smith J *et al*. Depression, quality of life, and glycemic control in individuals with type 2 diabetes. *J Am Acad Nurse Pract* 2009; **21:** 214–224.

[65] Sousa VD, Zanetti ML, Zauszniewski JA, Mendes IA, Daguano MO. Psychometric properties of the Portuguese version of the Depressive Cognition Scale in Brazilian adults with diabetes mellitus. *J Nurs Meas* 2008; **16:** 125–135.

[66] Amsberg S, Anderbro T, Wredling R, Lisspers J, Lins P, Adamson U *et al*. A cognitive behavior therapy-based intervention among poorly controlled adult type 1 diabetes patients—a randomized controlled trial. *Patient Educ Couns* 2009; **77:** 72–80.

[67] Huang MF, Courtney M, Edwards H, Mcdowell J. Validation of the Chinese version of the Problem Areas in Diabetes (PAID-C) scale. *Diabetes Care* 2010; **33:** 38–40.

[68] Snoek FJ, Pouwer F, Welch GW, Polonsky WH. Diabetes-related emotional distress in Dutch and US diabetic patients: cross-cultural validity of the Problem Areas in Diabetes scale. *Diabetes Care* 2000; **23:** 1305–1309.

[69] Welch GW, Jacobson AM, Polonsky WH. The Problem Areas in Diabetes Scale. An evaluation of its clinical utility. *Diabetes Care* 1997; **20:** 760–766.

[70] Lange LJ, Piette JD. Perceived health status and perceived diabetes control: psychological indicators and accuracy. *J Psychosom Res* 2005; **58:** 129–137.

[71] Jacobson AM, de Groot M, Samson JA. The effects of psychiatric disorders and symptoms on quality of life in patients with type I and type II diabetes mellitus. *Qual Life Res* 1997; **6:** 11–20.

[72] Riazi A, Bradley C, Barendse S, Ishii H. Development of the well-being questionnaire short-form in Japanese: the W-BQ12. *Health Qual Life Outcomes* 2006; **4:** 40–10.

[73] Pouwer F, van der Ploeg HM, Adèr HJ, Heine RJ, Snoek FJ. The 12-item well-being questionnaire. An evaluation of its validity and reliability in Dutch people with diabetes. *Diabetes Care* 1999; **22:** 2004–2010.

[74] Cheng T, Boey KW. Coping, social support, and depressive symptoms of older adults with type II diabetes mellitus. *Clin Gerontol* 2000; **22:** 15–30.

[75] Chou KL, Chi I. Prevalence of depression among elderly Chinese with diabetes. *Int J Geriatr Psychiatry* 2005; **20:** 570–575.

[76] Yang FM, Jones RN. Measurement differences in depression: chronic health-related and socio-demographic effects in older Americans. *Psychosom Med* 2008; **70:** 993–1004.

[77] Bech P, Gudex C, Staehr Johansen K. The WHO (Ten) Well-Being Index: validation in diabetes. *Psychother Psychosom* 1996; **65:** 183–190.

[78] Lloyd CE. Measuring psychological well-being in South Asians with diabetes. *Divers Health Care* 2010; **7:** 288–290.

[79] McHale M, Hendrikz J, Dann F, Kenardy J. Screening for depression in patients with diabetes mellitus. *Psychosom Med* 2008; **70:** 869–874.

[80] Fisher L, Skaff MM, Chesla CA, Kanter RA, Mullan JT. Contributors to depression in Latino and European-American patients with type 2 diabetes. *Diabetes Care* 2001; **24:** 1751–1757.

[81] Welch G, Weinger K, Anderson B, Polonksy WH. Responsiveness of the Problem Areas in Diabetes (PAID) questionnaire. *Diabet Med* 2003; **20:** 69–72.

[82] Pouwer F, Skinner TC, Pibernik-Okanovic M, Beekman ATF, Cradock S, Szabo S *et al*. Serious diabetes-specific emotional problems and depression in a Croatian–Dutch–English Survey from the European Depression in Diabetes [EDID] Research Consortium. *Diabetes Res Clin Pract* 2005; **70:** 166–173.

[83] Adriaanse MC, Dekker JM, Spijkerman AM, Twisk JW, Nijpels G, van der Ploeg HM *et al*. Diabetes-related symptoms and negative mood in participants of a targeted population-screening program for type 2 diabetes: The Hoorn Screening Study. *Qual Life Res* 2005; **14:** 1501–1509.

[84] Bhui K, Bhugra D. Transcultural psychiatry: some social and epidemiological research issues. *Int J Soc Psychiatry* 2001; **47:** 1–9.

[85] Sartorius N. SCAN Translation. In: Wing JK, Sartorius N and Ustun TB, eds. *Diagnosis and Clinical Measurement in Psychiatry*. Cambridge: Cambridge University Press, 1998.

[86] Snoek FJ, Kersch NYA, Eldrup E, Harman-Boehm I, Hermanns N, Kokoszka A *et al*. Monitoring of individual needs in diabetes (MIND): baseline data from the cross-sectional Diabetes Attitudes Wishes and Needs (DAWN) MIND Study. *Diabetes Care* 2011; **34:** 601–603.

Figure 1 Flow diagram: database search (January 1970–October 2010).

Table 1 Depression screening tools: frequency of use, type of study and method of assessment

Screening tool*	Author(s), date first used [reference]	Total no. of studies (%)	Number of scale items (no. of studies)	Type of study (no. of studies)	Type of diabetes (no. of studies)	Methods of assessment (no. of studies)†
BDI-I BDI-II	Beck *et al.*, 1961 [15] Beck *et al.*, 1996 [17]	55 (23.5)	21 items BDI-I (44) 21 items BDI-II (11)	Incidence/ prevalence (11) Observational (34) Case–control (4) Experimental (6)	Type 1 (25) Type 2 (17) Not reported (19)	SRSA (53) SAM (2)
CES-D	Radloff, 1977 [16]	49 (20.09)	20 items (44) 10 items (short version) (4) 2 items (modified standard assisted) (1)	Incidence/ prevalence (17) Observational (43) Case–control (4) Validity (3)	Type 1 (10) Type 2 (28) Not reported (20)	SRSA (47) SAM (2)
PAID scale	Polonsky *et al.*, 1995 [18] Polonsky *et al.*, 2002 [19]	28 (12.0)	PAID-1 (27) 20 items PAID-2 (1) 17 items	Prevalence (2) Observational (13) RCT (2) Case–control (1) Validity (4)	Type 1 (4) Type 2 (9) Not reported (12)	SRSA (28)
PHQ-9 PHQ-8 PHQ-2	Spitzer *et al.*, 1999 [23]	25 (10.7)	PHQ-9 (20) PHQ-8 (3) PHQ-2 (2)	Incidence/ prevalence (8) Observational (16) RCT(1)	Type 1 (06) Type 2 (11) Not reported (16)	SR-Q (21) SAM (4)
HADS	Zigmond and Snaith, 1983 [24]	24 (10.3)	7 items measure depression/ 7 items measure anxiety (24)	Incidence/ prevalence (6) Observational (20) RCT (3)	Type 1 (9) Type 2 (11) Not reported (12)	SRSA (2) IDR (22)

Zung-SDS	Zung, 1965 [25]	18 (7.7)	20 items (18)	Prevalence (5) Observational (11) Case–control (3)	Type 1 (10) Type 2 (28) Not reported (20)	SRSA (18)
SF-36 SF-12	Ware *et al.*, 1993 [26]	15 (6.4)	SF-36 (11) SF-12 (4)	Prevalence (3) Observational (11) RCT (2)	Type 1 (1) Type 2 (6) Both (2) Not reported (6)	SRSA (15)
HRSD	Hamilton, 1960 [27]	10 (4.3)	17 items (10)	Observational (8) RCT (1) Case–control (1)	Type 1 (3) Type 2 (4) Not Reported (5)	IDR (10)
GDS	Yesavage *et al.*, 1982 [28]	10 (4.3)	GDS (8) GDS-15 (2)	Prevalence (4) Observational (6)	Type-1 (0) Type 2 (3) Not reported (7)	SRSA (10)
WHO-5 WHO-10	WHO, 1990 [29]; Bech *et al.*, 1996 [30]	09 (3.8)	WHO-5 (7) WHO-10 (2)	Prevalence (2) Observational (6) RCT (1)	Type 1 (04) Type 2 (07) Not reported (2)	SRSA (9)
WBQ-22 WBQ-12	Bradley [31]	10 (4.3)	G-WBQ (4) WBQ-22 (1) WBQ-12 (5)	Prevalence (1) Observational (6) RCT (1) Validity (2)	Type 1 (3) Type 2 (4) Not reported (6)	SRSA (10)
MADRS	Montgomery and Asberg, 1979 [32]	04 (1.7)	10 items (4)	Prevalence (3) Observational (1)	Type 1 (0) Type 2 (4) Not reported (0)	IDR (4)
SCL-20 SCL-90	Derogatis and Savitz, 2000 [33]	04 (1.7)	SCL-20 (2) SCL-90 (2)	Observational (4)	Type 1 (1) Type 2 (2) Not reported (2)	SRSA (4)
GHQ-12 GHQ-30	Goldberg and Blackwell, 1970 [34]	03 (1.3)	GHQ-12 (2) GHQ-30 (1)	Epidemiological (1) Observational (2)	Type 1 (0) Type 2 (3) Not reported (1)	SRSA (3)
DSM-III	Spitzer, 1981 [35]	03 (1.3)	DSM-III 265 items (2)	Incidence (1) Observational (3)	Type 1 (1) Type 2 (1)	SRSA (3)

			DSM-III R 292 items (1)		Not reported (2)	
EQ-5D	EuroQol Group, 1987 [36]	03 (1.3)	5 items	Incidence (1) Observational (2)	Type 1 (0) Type 2 (1) Not Reported (2)	SRSA (2) SAM (1)
Others§		33 (14.1)		Prevalence (6) Observational (23) RCT (2) Case–control (3) Validity (2) Experimental (1)	Type 1 (2) Type 2 (5) Not reported (36)	SRSA (16) IDR (17)

*BDI, Beck Depression Inventory; CES-D, Center for Epidemiologic Studies Depression Scale; PAID, Problem Areas in Diabetes scale; PHQ, Patient Health Questionnaire; HADS, Hospital Anxiety and Depression Scale; Zung-SDS, Zung Self-Rating Depression Scale; SF, Short Form Medical Outcomes Survey; HRSD, Hamilton Rating Scale for Depression; GDS, Geriatric Depression Scale; WHO-5/WHO-10, WHO Well-Being Index; WBQ, General Well-Being Questionnaire; MADRS, Montgomery–Åsberg Depression Rating Scale; SCL, Symptom Check List; GHQ, General Health Questionnaire; DSM, Diagnostic and Statistical Manual of Mental Disorders; EQ-5D, European Quality of Life-5 Dimensions.

†SRSA, self-reported self-administered; SAM, standard assisted method; IDR, interviewed by doctor/researcher.

‡Frequency adds up to more than total number of papers (n = 234) because many studies used more than one scale.
§Frequency of use is less than 3.
RCT, randomized controlled trial.

Table 2 Availability of sensitivity/specificity information, translations and the setting in which each screening tool has been used

Screening tool*	Sensitivity/ specificity	Translations	Settings
CES-D	References other studies (33) Reported for own study (5) Not reported (22)	English (40), Spanish (7), Dutch (4), Croatian (2), Chinese (3), Hawaiian (2), Italian (2), Japanese (2), Malay (2), Philipino (1), Turkish (3), Portuguese (1), Korean (1), Cantonese (2), Mandarin (2), German (2) Not reported (14)	Primary care (13) Secondary care (23) Community (27) Tertiary care (1)
BDI-I BDI-II	References other studies (14) Reported for own study (3) Not reported (38)	English (30), German (1), Chinese (1), Dutch (1), Finnish (2), French (4), Korean (1), Arabic (1), Spanish (2), Greek (2), Polish (1), South African (1) Not reported (17)	Primary care (13) Secondary care (21) Community (27) Tertiary care (1)
PHQ-9 PHQ-8 PHQ-2	References other studies (12) Reported for own study (3) Not reported (10)	English (22) Dutch (2) Not reported (1)	Primary care (10) Secondary care (2) Community (9) Tertiary care (1) Not reported (2)
HADS	References other studies (15) Reported for own study (3) Not reported (7)	English (13), German (1), Chinese (2), Dutch (2), Norwegian (2), French (2), Spanish (1), Greek (1), Swedish (1), Icelandic (1), Urdu (1) Not reported (5)	Primary care (7) Secondary care (16) Community (4) Tertiary care (2)
Zung-SDS	References other studies (5) Reported for own study (6) Not reported (7)	English (6), Chinese (3), Finnish (1), Japanese (2), Russian (1), Italian (1), Not reported (4)	Primary care (1) Secondary care (11) Community (4) Tertiary care (1) Not reported (1)
WHO-5 WHO-10	References other studies (7) Reported for own study (2) Not reported (2)	English (7), Dutch (1), Spanish (1), Spanish (Argentina) (1), Greek (1), Japanese (2) Not reported (5)	Primary care (3) Secondary care (6) Community (2)
PAID	References other	English (14), Croatian (2), Chinese (1),	Primary care (3)

	studies (14) Reported for own study (3) Not reported (12)	Japanese (2), Dutch (4),Turkish (2), Brazilian (1), German (1), Argentina (1), Swedish (2), Polish (1), Icelandic (1) Not reported (3)	Secondary care (22) Community (3) Tertiary care (1)
SF-36 SF-12	References other studies (5) Reported for own study (0) Not reported (10)	English (12), Hawaiian (1), Croatian (1), Not reported (1)	Primary care (7) Secondary care (5) Community (3)
GDS	References other studies (3) Reported for own study (0) Not reported (7)	English (4), Chinese (3), Italian (1), Turkish (1), Not reported (1)	Secondary care (5) Community (3) Tertiary care (2)
WBQ-22 WBQ-12	References other studies (7) Reported for own study (1) Not reported (2)	English (2), Japanese (1), Swedish (1), Greek (1), Yoruba (1), Dutch (4) Not reported (1)	Secondary care (5) Community (1)
HRSD	References other studies (2) Reported for own study (0) Not reported (8)	English (5), Turkish (1), Polish (1), Finnish (1), Italian (1), Yoruba (1) Not reported (3)	Primary care (1) Secondary care (9) Community (1) Tertiary (1)
MADRS	References other studies (1) Reported for own study (0) Not reported (3)	Urdu (1), Finnish (1), Bengali (1) Not reported (2)	Primary care (1) Community (3)
SCL-20 SCL-90	References other studies (2) Reported for own study (0) Not reported (2)	English (3), Dutch (1), Not reported (1)	Primary care (1) Secondary care (1) Community (1) Tertiary (1)
GHQ-12 GHQ-30	References other studies (1) Reported for own study (0)	English (2), Finnish (1), Swedish (1) Not reported (1)	Primary care (1) Community (3)

	Not reported (3)		
DSM-III	References other studies (1) Reported for own study (0) Not reported (2)	English (3)	Primary care (1) Secondary care (1) Community (1)
EQ-5D	References other studies (1) Reported for own study (0) Not reported (2)	English (1), Finnish (1), Swedish (1) Not reported (2)	Primary care (1) Community (2)
Others‡	References other studies (16) Reported for own study (2) Not reported (15)	English (22), Finnish (1), Swedish (2), French (1), Spanish (3), Greek (1), German (1), Russian (1), Italian (1) Not reported (2)	Primary care (7) Secondary care (18) Community (9) Tertiary (1)

* CES-D, Center for Epidemiologic Studies Depression Scale; BDI, Beck Depression Inventory; PHQ, Patient Health Questionnaire; HADS, Hospital Anxiety and Depression Scale; Zung-SDS, Zung Self-Rating Depression Scale; WHO-5/WHO-10, WHO Well-Being Index; PAID, Problem Areas in Diabetes scale; SF, Short Form Medical Outcomes Survey; GDS, Geriatric Depression Scale; WBQ, General Well-Being Questionnaire; HRSD, Hamilton Rating Scale for Depression; MADRS, Montgomery–Åsberg Depression Rating Scale; SCL, Symptom Check List; GHQ, General Health Questionnaire; DSM, Diagnostic and Statistical Manual of Mental Disorders; EQ-5D, European Quality of Life-5 Dimensions.

†Frequency adds up to more than total number of papers ($n = 234$) because many studies used more than one scale.
‡Frequency of use is less than 3.

Table 3 Studies reporting prevalence of depression with sensitivity/specificity and reliability/validity data

Section 1: sensitivity and specificity (*n* = 16)

Tool*	Study [reference]	Prevalence rate % (total *n*)	Sensitivity† (%)	Specificity‡ (%)	PPV (%)§	NPV (%)¶	Reliability/ validity (α)
CES-D	Wagner *et al*., 2009 [37]	66 (47)	60.0-100.0	86.7	28.6	97.0	0.80
CES-D		27 (374)					
—Chinese	Stahl *et al*., 2008 [38]		98.8	67.6	34.5	99.2	—
—Malay		36 (57)	66.7	60.6	23.5	90.9	—
—Indian		43 (73)	100.0	64.5	56.0	100.0	—
PHQ-9	Kahn *et al*., 2008 [39]	56 (249)	66.0	52.0	—	—	—
PHQ-9	Lamers *et al*., 2008 [40]	19.3 (5154)	89.0-100.0	75.0-85.0	54.9–64.4	96.2–100.0	0.80–0.84
PHQ-8	Ackermann *et al*., 2005 [41]	51 (890)	89.0-99.0	54.0-65.0	43.8–51.6	94.3–97.8	—
PHQ-9	Van Steenbergen *et al*., 2010 [42]	46 (197)	75.7	80.0	46.7	93.4	—
BDI/SF	Sultan *et al*., 2010 [43]	38 (256)	88.0	71.0	22.0	98.0	0.89
BDI	Lustman *et al*., 1997 [44]	36.6 (172)	82.0–90.0	84.0–89.0	59.0–89.0	82.0–97.0	
HADS	Sultan *et al*., 2010 [43]	41 (256)	83.0	65.0	28.0	96.0	0.78
HADS-A	Nayani, 2008 [45]	—	66.0	37.5	—	—	—
HADS-D		—	85.0	70.0	—	—	—
HADS	Engum, 2007 [46]	22.3 (37,291)	74.0	86.0	49.0	95.0	0.55
Zung-SDS	Xu *et al*., 2004 [47]	23 (222)	85.0–90.0	90.0–95.0	—	—	—
Zung-SDS	Yoshida *et al*., 2009 [48]	36.4 (129)	100.0	59.0	—	—	0.72–0.88
Zung-SDS	Biggs *et al*., 1978 [49]	—	69.0	—	—	—	—
HDRS	Biggs *et al*., 1978 [49]	—	80.0	—	—	—	—
PAID	McGuire *et al*., 2010 [50]	—	80.0–94.0	80.0–89.0	—	—	—
WHO-5	McGuire *et al*., 2010 [50]	—	100.0	78.2	45.5	100	0.89
WHO-5	Awata *et al*., 2007 [51]	15.4 (65)	33.0–71.0	5.0–36.0	8.0	24.0	0.89
WHO-5	Newnham *et al*., 2010 [52]	—	95.0	89.0	—	—	0.82–0.86
WHO-5	Furuya *et al*., 2010 [53]	—	53.6–57.1	67.7–82.5	—	—	—
GDS	Bai *et al*., 2008 [54]	15.4 (156)	93.3	92.3	—	—	0.87

Section 2: reliability/validity (*n* = 23)

							Reliability/

Instrument (*n*)	Study [reference no.]	validity (α)
CES-D (7)	Keawe'aimoku *et al*., 2003; Zauszniewski and Graham, 2009; Pibernik *et al*., 2008; Wagner *et al*., 2007; Grandinetti *et al*., 2000; Park *et al*., 2004; Wagner *et al*., 2009 [55–61]	0.60–0.93
BDI (4)	Leonardson *et al*., 2003; Kagee *et al*., 2008; Lee *et al*., 2009; Sousa *et al*., 2008 [62–65]	0.80–0.95
PAID (5)	Pibernik *et al*., 2008; Amsberg *et al*., 2009; Huang *et al*., 2010; Snoek *et al*., 2000; Welch *et al*., 1997 [57, 66–69]	0.86–0.93
SF-36/SF-12 (4)	Keawe'aimoku *et al*., 2003; Pibernik *et al*., 2008; Lange, 2005; Jacobson, 1997 [55, 57, 70–71]	0.74–0.84
WBQ-12 (4)	Leonardson, 2003; Amsberg *et al*., 2009; Riazi, 2006; Pouwer *et al*., 1999 [61, 65, 72–73]	0.85–0.92
GDS (2)	Cheng and Boey, 2000; Chou and Chi, 2005 [74–75]	0.77–0.89
HADS (1)	Amsberg *et al*., 2009 [66]	0.87
Zung-SDS (1)	Yang, 2008 [76]	0.90
WHO-10 (1)	Bech, 1996 [77]	0.87

*CES-D, Center for Epidemiologic Studies Depression Scale; PHQ, Patient Health Questionnaire; BDI, Beck Depression Inventory; SF, Short Form Medical Outcomes Survey; HADS, Hospital Anxiety and Depression Scale; Zung-SDS, Zung Self-Rating Depression Scale; HRSD, Hamilton Rating Scale for Depression; PAID, Problem Areas in Diabetes scale; WHO-5/WHO-10, WHO Well-Being Index; GDS, Geriatric Depression Scale; WBQ, General Well-Being Questionnaire.

†Sensitivity = number of true positives (cases of depression)/number of true positives + number of false negatives.

‡Specificity = number of true negatives/number of true negatives + false positives.

§Positive predictive value (PPV) = the proportion of cases with positive test results who are correctly diagnosed.

¶Negative predictive value (NPV) **=** the proportion of cases with negative test results who are correctly diagnosed.

Figure 1 Flow diagram: database search (January 1970–October 2010)