Performance Characteristics of Depression Screening Instruments in Survivors of Acute Myocardial Infarction: Review of the Evidence

Brett D. Thombs, PhD[1,2]; Gina Magyar-Russell, PhD[1,2]; Eric B. Bass, MD, MPH[1,3,4,5]; Kerry J. Stewart, ED[1,3]; Konstantinos K. Tsilidis, MPH[1,4]; David E. Bush, MD[1,3]; James A. Fauerbach, PhD[1,2]; Una D. McCann, MD[1,2]; Roy C. Ziegelstein, MD[1,3]

[1]Johns Hopkins University Evidence-based Practice Center, Johns Hopkins University School of Medicine, Baltimore, Md; [2]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Md; [3]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Md; [4]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Md; [5]Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, Md

(Dr. Thombs is now a member of the Department of Psychiatry of the Sir Mortimer B. Davis-Jewish General Hospital and McGill University, Montreal, QC)

Running Title: Screening for Depression Post-AMI

Address for Correspondence:

Brett D. Thombs, Ph.D.

Institute of Community and Family Psychiatry

SMBD-Jewish General Hospital

4333 Cote Ste Catherine Road

Montreal, Quebec H3T 1E4

Tel (514) 340-8222 ext. 5112

Fax (514) 340-8124

E-mail: brett.thombs@mcgill.ca

This article is based on research conducted by the Johns Hopkins University Evidence-based Practice Center under contract to the Agency for Healthcare Research and Quality (Contract No. 290-02-0018), Rockville, MD. The authors of this article are responsible for its contents, including any clinical or treatment recommendations. No statement in this article should be construed as an official position of the Agency for Healthcare Research and Quality or of the U.S. Department of Health and Human Services. We thank Captain Ernestine Murray for her support as the Task Order Officer for this project.

ABSTRACT

A systematic review was conducted to assess performance characteristics of depression screening instruments post-acute myocardial infarction. Among the 7 studies identified, the Beck Depression Inventory (BDI) and the depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) were used most frequently. The HADS-D exhibited adequate specificity (89.9%), but poor sensitivity (65.0%), for combined major and minor depression. A BDI $\geq$ 10 was more sensitive (81.1%), but less specific (67.6%), for combined major and minor depression. Although based on limited evidence, the BDI appears to be a more sensitive measure than the HADS-D and provides more comprehensive symptom coverage.

INTRODUCTION

Major depression is present in approximately 20% of patients hospitalized with acute myocardial infarction (AMI).{{3701 Thombs,B.D. 2006; }} Symptoms of depression following AMI predict lower physical function,{{4025 Ades,P.A. 2002; }} poorer quality of life,{{4026 Ruo,B. 2003; }} substantially higher health-care costs,{{3663 Frasure-Smith,N. 2000; }} and increased cardiac morbidity.{{357 Lauzon,C. 2003;4027 Shiotani,I. 2002; }} Some studies, however, have failed to find this association.{{330 Lane,D. 2001;359 Mayou,R.A. 2000; }} A recent meta-analysis{{542 van Melle,J.P. 2004; }} and systematic review{{1160 Bush,D.E. 2005; }} each concluded that depression is associated with cardiac and all-cause mortality in post-AMI patients after controlling for other predictors. The authors of another systematic review, however, concluded that quality issues, such as the use of measurement instruments not specifically validated in AMI patients and insufficient power in many studies, make it difficult to draw unambiguous conclusions about the association of depression with post-AMI mortality.{{1158 Sorensenf,C. 2005; }}

American College of Cardiology / American Heart Association practice guidelines for AMI recommend that the psychosocial status of patients be evaluated, “including inquiries regarding symptoms of depression,”{{494 Antman,E.M. 2004; }} but do not recommend specific procedures for assessing depression. Canadian Cardiovascular Society guidelines for AMI{{3626 Fallen,E.L. 1995; }} recommend that all patients with AMI be screened for depression with the Beck Depression Inventory (BDI),{{3640 Beck, A.T. 1987; }} and the National Service Framework for Coronary Health Disease of the United Kingdom{{4028 Department of Health 2000; }} recommends that the Hospital Anxiety and Depression Scale

(HADS){{497 Zigmond,A.S. 1983; }} be used to screen for symptoms of depression. We recently reported in a systematic review that the BDI and the HADS depression subscale (HADS-D) are the most commonly used methods to screen for symptoms of depression in studies of post-AMI patients.{{3701 Thombs,B.D. 2006; }} The weighted prevalence of significant symptoms of depression across studies, however, differs substantially when screened using a BDI score $\geq$ 10 (31.1%) as compared to a score on the HADS-D $\geq$ 8, (15.5%) or $\geq$ 11, (7.3%).{{3701 Thombs,B.D. 2006; }}

Accurately assessing symptoms of depression can be difficult in patients hospitalized with AMI due to the overlap between somatic manifestations of depression and the physical symptoms of AMI and its treatment. Symptoms characteristically associated with depression, such as fatigue or loss of energy, anhedonia, changes in sleep patterns, changes in appetite, or poor concentration, may occur as a normal reaction to the AMI or to the hospitalization itself.{{512 Koenig,H.G. 1993;514 Cavanaugh,S. 1983; 4029 Morrison,M.F. 1997; }} The difference in rates obtained when using the BDI and the HADS-D could reflect whether or not somatic symptoms are assessed. Among medically ill patients, diagnostic paradigms that do not include somatic symptoms tend to produce lower prevalence rates.{{1162 Koenig,H.G. 1997;3487 Turner-Stokes,L. 2002; }} The BDI includes 7 items related to somatic symptoms among its 21 items. Conversely, the HADS-D is comprised of 7 items, none of which relates to somatic symptoms.

No studies have systematically reviewed the psychometric properties of screening tools for depression following AMI. Only one review has evaluated depression screening instruments in a defined medical population, and it reported very limited psychometric data for patients with

stroke.{{3487 Turner-Stokes,L. 2002; }} The high prevalence of depression in post-AMI patients and its relationship to important outcomes suggest that greater attention should be paid to the measurement characteristics of assessment tools in this population. Further, measurement characteristics of depression screening instruments may differ in post-AMI patients compared to other patient groups due to the acute nature of an AMI. A study by Strik et al., for instance,{{345 Strik,J.J. 2001; }} that compared three standard depression measures in a post-AMI sample found that cutoff points validated in other patient populations did not work well in this patient group. The present systematic review of the literature was done to 1) evaluate the performance characteristics (i.e., validity, reliability, sensitivity, specificity) of screening instruments for depressive symptoms in post-AMI patients, both in the hospital and within 3 months of hospitalization, and 2) make recommendations for the assessment of depressive symptoms in this population.

METHODS

The review was part of a comprehensive evidence report that addressed several different questions related to post-AMI depression (e.g., prevalence, association with morbidity and mortality, treatment).{{1160 Bush,D.E. 2005; }} Key aspects of the methods are summarized below.

Search Strategy

For the comprehensive evidence report,{{1160 Bush,D.E. 2005; }} the MEDLINE®, Cochrane, CINAHL®, PsycINFO®, and EMBASE® databases were searched for articles published between 1980 and March 2004. Search strategies and terms are found in Appendix A. Hand searching was done on 16 selected journals (see Appendix B) from October 2003 through

April 2004, and on references of reviews and eligible articles. ProCite® reference management software was used to create a database of reference material identified through an electronic search for relevant guidelines and reviews, through discussions with experts, and through the article review process. The article search for the comprehensive evidence report was updated by an additional search of the same databases for articles published between March 2004 and November 2005 that reported on psychometric performance.

Study Selection, Data Extraction, and Quality Assessment

Two investigators independently evaluated studies for inclusion with discrepancies resolved by consensus. Studies published since 1980 that reported psychometric data on a validated questionnaire used to assess symptoms of depression were included. When multiple articles were published on the same study cohort, the most relevant article was included. Articles were excluded if they consisted of case series or case reports, were not in English, if only a meeting abstract was provided, if depression was not measured by a validated method, or if the timing of the depression assessment relative to the MI was not reported. Studies with mixed populations were eligible for inclusion if data on MI patients were reported separately. Studies that reported on the diagnostic accuracy of a screening instrument were included if they used standard or empirically-derived cutoff scores, but not if they used arbitrary cutoff scores without justification. Although studies have reported psychometric data on depression instruments with other coronary artery disease patient groups, we only included articles on MI patients due to the focus of the comprehensive evidence report and because characteristics of hospitalized AMI patients may result in different measurement challenges, particularly due to the acute nature of

the event. Author and journal names were not masked since masking does not appear to influence inclusion and exclusion decisions.{{560 Berlin,J.A. 1997; }}

Two investigators independently extracted data, reconciling differences by face-to-face meeting. Data extraction forms were developed from consensus among the investigative team regarding the items that were most important for describing the characteristics of each study, and summarizing study results. Evidence from studies was classified using a 4-tiered system based on American Academy of Neurology guidelines{{790 Edlund,W. 2004; }} and intended to reflect the degree of potential bias in the findings (Level I, low; Level II, moderate; Level III, moderate to high; Level IV, very high). Some studies that were included in this review were not designed specifically to assess performance characteristics of measurement instruments. Thus, quality grades were assigned based on evidence regarding the utility of specific instruments as depression screening tools rather than on the evidence quality of the overall study.

Instruments Used to Screen for Depression Post-AMI

For this review, we defined depression as "symptoms meeting established clinical threshold criteria for depression as measured by validated questionnaires or standardized psychiatric interviews."{{1160 Bush,D.E. 2005; }} The present review was limited to self-report instruments, which are easy to administer and score, and can provide a quick estimate of symptom number and severity. Although the inclusion criteria for the comprehensive evidence report refer to the use of “validated questionnaires,” it is recognized that most instruments used with post-AMI patients have been validated in a patient group other than patients with AMI or cardiovascular disease.

Definitions of Reliability and Validity

Screening instruments for depression should display adequate coverage of the full range of symptoms of depression, including the core symptoms of depressed mood and anhedonia; should assess suicidality; and should display adequate population-specific reliability and validity. Reliability in this context refers to the extent to which each item of the scale reflects the same construct (internal consistency reliability) and the degree to which it produces the same score across administrations (test-retest reliability). Adequate construct validity is achieved if the instrument adequately covers the range of depressive symptoms (content validity), predicts depression status at a later time point (predictive validity), correlates with established measures of depression (convergent validity), and discriminates from constructs other than depression, such as anxiety (discriminative validity).{{3627 Joiner,T.E.,Jr 2005; }}

RESULTS

Search Results

The search process for the entire post-AMI depression review{{1160 Bush,D.E. 2005; }} identified 3,770 unique titles. During the title and abstract reviews, 2,597 and 825 citations were excluded, respectively. Of the remaining articles that were deemed eligible for at least one of the 6 comprehensive review questions, 5 articles met inclusion criteria for this review of performance characteristics of depression screening instruments.{{ 345 Strik,J.J. 2001; 366 Martin,C.R. 2003; 380 Martin,C.R. 2000; 379 Freedland,K.E. 2002; 4022 Wojciechowski,F.L. 2000; }} A subsequent supplementary search, which was done to identify articles on psychometric properties that were published between March 2004 and November 2005, uncovered 2 additional articles.{{543 Dickens,C.M. 2004;3688 Boersma,S.N. 2005; }}

Characteristics of Studies

A summary of key aspects of the 7 eligible studies is presented in Table 1. Included studies were published between 2000 and 2005. All of the studies were composed entirely of post-MI patients. Two studies reported data on factor/construct validity,{{380 Martin,C.R. 2000; 366 Martin,C.R. 2003; }} 2 on convergent validity,{{379 Freedland,K.E. 2002; 4022 Wojciechowski,F.L. 2000; }} 3 on internal consistency reliability,{{380 Martin,C.R. 2000;366 Martin,C.R. 2003; 3688 Boersma,S.N. 2005; }} and 2 on diagnostic utility.{{ 345 Strik,J.J. 2001; 543 Dickens,C.M. 2004; }} Eligible studies reported psychometric data on the BDI,{{345 Strik,J.J. 2001; 379 Freedland,K.E. 2002; }} the HADS-D{{366 Martin,C.R. 2003;380 Martin,C.R. 2000; 345 Strik,J.J. 2001; 3688 Boersma,S.N. 2005; }} or overall HADS,{{543 Dickens,C.M. 2004; }} the Symptom Checklist-90 Depression subscale (SCL-90-Dep),{{345 Strik,J.J. 2001;4022 Wojciechowski,F.L. 2000; }} and the Zung Depression Scale (ZDS).{{4022 Wojciechowski,F.L. 2000; }}

Six studies were conducted in Europe,{{345 Strik,J.J. 2001;366 Martin,C.R. 2003; 380 Martin,C.R. 2000; 4022 Wojciechowski,F.L. 2000; 3688 Boersma,S.N. 2005; 543 Dickens,C.M. 2004; }} and 1 was conducted in the United States.{{379 Freedland,K.E. 2002; }} Four studies were multi-center studies.{{379 Freedland,K.E. 2002;543 Dickens,C.M. 2004; 366 Martin,C.R. 2003; 3688 Boersma,S.N. 2005; }} The mean age of participants ranged from 55 to 67 years. The data reported here from the ENRICHD study,{{508 Berkman,L.F. 2003; }} which was designed to maximize diversity of participants, included 56% male and 66% white participants.{{379 Freedland,K.E. 2002; }} The other studies enrolled between 63% and 81% men. Only 2 studies provided data on race; they enrolled 94%{{543 Dickens,C.M. 2004; }} and

100%{{345 Strik,J.J. 2001; }} white participants. Only 1 study reported cardiac risk factors,{{379 Freedland,K.E. 2002; }} and none of the studies reported MI characteristics such as Killip class or left ventricular ejection fraction.

Of the 7 studies reviewed, 5 were assigned a Level IV quality of evidence rating{{3688 Boersma,S.N. 2005;379 Freedland,K.E. 2002; 366 Martin,C.R. 2003; 380 Martin,C.R. 2000; 4022 Wojciechowski,F.L. 2000; }} and 2 were assigned a Level III rating.{{345 Strik,J.J. 2001; 543 Dickens,C.M. 2004; }} Reasons for low evidence grades included the lack of evidence directly addressing screening utility, relatively small sample sizes, the lack of representative samples, or the failure to disclose the percent of eligible patients recruited. One study, which overall was a Level I study, received a Level IV rating for the purposes of this review since it only included patients with significant symptoms of depression or with low social support and only reported a correlation between two measures.{{379 Freedland,K.E. 2002; }}

Validity of Depression Screening Instruments Post-AMI

Two studies{{380 Martin,C.R. 2000; 366 Martin,C.R. 2003; }} examined the factor/construct validity of depression screening instruments in post-AMI patients. One of the studies examining the factor structure of the HADS used exploratory factor analysis (EFA),{{380 Martin,C.R. 2000; }} and the other used confirmatory factor analysis (CFA).{{366 Martin,C.R. 2003; }} In the EFA study, the HADS was administered within 24 hours of admission to the cardiac care unit. In the CFA study, the HADS was administered at three separate time points: at 1 week, 6 weeks, and 6 months post-AMI. Both studies found that a 3-factor solution represented the data better than the 2-factor depression and anxiety factors proposed by the authors of the HADS.{{497 Zigmond,A.S. 1983; }} In each case, however, a

single depression factor represented by the items of the HADS-D was found to be distinct from two anxiety-related factors from the anxiety subscale.

Two studies reported data on the convergent validity of the BDI, the ZDS, and the SCL-90-Dep.{{379 Freedland,K.E. 2002; 4022 Wojciechowski,F.L. 2000; }} The ENRICHD investigators reported a Pearson correlation of .64 between the BDI and the Hamilton Rating Scale for Depression (HRSD) at least 2 weeks post-AMI, close to the low end of findings reported in other clinical settings (.61 to .87).{{379 Freedland,K.E. 2002; }} Since all patients included in the analysis had significant symptoms of depression, however, the relatively low correlation may have been a result of a restricted range of scores on both the BDI and HRSD. The second study reported Pearson correlation coefficients from .70 to .77 between the ZDS and the SCL-90-Dep across 4 time periods post-AMI (1, 3, 6, and 12 months).{{ 4022 Wojciechowski,F.L. 2000; }}

Reliability of Depression Screening Instruments Post-AMI

Internal consistency reliability for the HADS-D as measured by Cronbach's alpha was .72 in a sample of 194 male and female patients assessed within 24 hours of admission for AMI.{{380 Martin,C.R. 2000; }} A second study of 335 male and female patients reported Cronbach's alpha for the HADS-D as .76, .80, and .81 at 1 week, 6 weeks, and 6 months post-AMI.{{366 Martin,C.R. 2003; }} A third study reported a Cronbach’s alpha of .82 for the HADS-D measured 2-7 weeks post-AMI.{{3688 Boersma,S.N. 2005; }}

Diagnostic Utility of Depression Screening Instruments Post-AMI

Two studies{{345 Strik,J.J. 2001; 543 Dickens,C.M. 2004; }} assessed the diagnostic utility of depression screening instruments. One study reported on the diagnostic utility of the

BDI, HADS-D, and SCL-90-Dep as compared with a diagnosis of major depression or combined major or minor depression based on a physician-administered Structured Clinical Interview for DSM-IV (SCID-IV){{1072 Spitzer,R. 1988;818 First,M. B. 1995; }} one month post-AMI.{{345 Strik,J.J. 2001; }} The number of patients in each analysis ranged from 179 to 199, depending on the measure. SCID-IV prevalence was 11.2% for major depression and 7.8% for minor depression for a combined prevalence of 18.9%. Using standard cutoff scores for combined major or minor depression, a BDI score of 10 or greater was 81.1% sensitive and 67.6% specific; a HADS-D ≥ 8 was 65.0% sensitive and 89.9% specific; and an SCL-90-Dep score ≥ 23 for men and ≥ 28 for women was 75.3% sensitive and 81.1% specific (Figure 1). Corrected specificity and sensitivity figures for the HADS-D were provided by the authors of the original study due to inaccuracies that were identified in the published manuscript (personal communication, December 16, 2005).

The authors of this study also reported results from empirically-derived cutoff points that were determined through visual assessment of receiver operating characteristic (ROC) curves. Empirically-derived cutoffs for combined major or minor depression were 8 or greater for the BDI (sensitivity 83.8%, specificity 71.7%), 4 or greater for the HADS-D (sensitivity 75.0%, specificity 77.6%), and 27 or greater for the SCL-90-Dep (sensitivity 81.1%, specificity 83.5%) (Figure 2). For major depression alone, empirically-derived cutoffs were 10 or greater on the BDI (81.8% sensitivity, 78.7% specificity), 4 or greater on the HADS-D (85.0% sensitivity, 74.8% specificity), and 25 or greater on the SCL-90-Dep (95.5% sensitivity, 74.0% specificity) (Figure 3).

The interpretability of these results, however, is limited by non-uniform timing of assessments with the screening questionnaires. Patients in the study were evaluated with the SCID-IV 1 month post-MI. At the time of the evaluation, they were asked to fill out the BDI, the HADS-D, and the SCL-90 at home and then return them to the investigators. Patients who did not return the questionnaires within 2 weeks received a reminder phone call. The study did not report data on the delay between assessment with the SCID-IV and return of the questionnaires. Thus, for an unknown proportion of patients, these results do not appear to represent concurrent sensitivity and specificity data. In addition, the number of patients assessed varied across instruments. Given the relatively small number of cases identified by the SCID-IV in this study, it is possible that this could have impacted comparative sensitivity and specificity results.

A second study, which included 314 patients hospitalized for an AMI,{{543 Dickens,C.M. 2004; }} reported good sensitivity (87.7%) and specificity (84.7%) for the overall HADS (≥ 18) compared to a diagnosis of ICD-10 depressive disorder in the month before the AMI based on the Schedule for Assessment of Neuropsychiatric Disorders{{4021 WHO Division of Mental Health 1996; }} (Figure 3). Assessments were done as soon as medically possible after the MI (mean 3.6 days post-AMI), and the prevalence of depressive disorder was 20.7%.

DISCUSSION

This is the first systematic review of the literature examining the psychometric properties of screening instruments for symptoms of depression in post-AMI patients. Only 7 studies were identified that provided data on validity, reliability, or diagnostic utility characteristics of depression assessment tools in this population. Of these studies, however, most were carried out

in relatively small samples, were generally of low quality, and provided limited information. Inconsistencies in results related to diagnostic utility from these studies highlight the shortcomings of the existing evidence base. Strik et al.,{{345 Strik,J.J. 2001; }} for instance, reported empirically derived cutoff scores for the SCL-90-Dep of 27 or greater to screen for major or minor depression, but a *lower* threshold of 25 or greater to screen for major depression alone. In addition, Strik et al. reported that a score of 4 or greater on the HADS-D maximized combined sensitivity and specificity regardless of whether major depression or combined major and minor depression was being assessed. Across studies, the relatively low empirically-derived cutoff score of 4 or greater on the 7-item HADS-D for major depression contrasts with the relatively high cutoff score of 18 or greater on the overall 14-item HADS that Dickens et al. {{543 Dickens,C.M. 2004; }} found to maximize diagnostic utility. Thus, it is difficult to conclude from the existing evidence that any one of the three instruments for which diagnostic utility data were reported (BDI, HADS, SCL-90-Dep) performed better than the others. Notably, none of the cutoff points derived in studies included in this review were replicated through cross-validation procedures. Cutoffs generated in single samples tend to capitalize on chance and maximize sensitivity and specificity, and cross-validation is necessary before cutoffs can be accepted as useful for clinical practice.{{3586 Charlson,M.E. 1987;334 Dawes,Robyn M. 1989; }}

A recent review of evidence-based assessment methods for depression recommended that, among other characteristics, screening instruments should display adequate coverage of the full range of symptoms of depression, including the core symptoms of depressed mood and anhedonia; should assess suicidality; and should display adequate psychometric standards,

including: 1) internal consistency and test-retest reliabilities of at least .70, 2) evidence for construct validity, and 3) demonstration of psychometric properties in at least 2 samples.{{3627 Joiner,T.E.,Jr 2005; }}

Consistent with values reported in studies of other medical populations,{{3690 Bjelland,I. 2002; }} the HADS-D was found to have adequate internal consistency reliability in three different samples of post-AMI patients.{{366 Martin,C.R. 2003;380 Martin,C.R. 2000; 3688 Boersma,S.N. 2005; }} In addition, two studies reported evidence for its construct validity as a measure of depression.{{366 Martin,C.R. 2003;380 Martin,C.R. 2000; }} The HADS-D, however, does not assess somatic symptoms of depression, does not screen for suicidality, and consistently identifies a relatively low percentage of post-AMI patients with a diagnosis of major depression.{{3701 Thombs,B.D. 2006; }} No studies have reported reliability data for the BDI with post-AMI patients, and only very limited convergent validity data is available.{{379 Freedland,K.E. 2002;4022 Wojciechowski,F.L. 2000; }} The BDI, however, identifies approximately 30% of post-AMI patients as having at least mild to moderate symptoms of depression, somewhat higher than the approximately 20% with major depression,{{3701 Thombs,B.D. 2006; }} and screens for suicidality, as well as for a full range of mood, cognitive, and somatic symptoms. No reliability data and only very limited validity data are available for the SCL-90-Dep with post-AMI patients.{{4022 Wojciechowski,F.L. 2000; }} The SCL-90-Dep does, however, assess a range of symptoms, including dysphoric mood and affect, anhedonia, lack of motivation and energy, hopelessness, suicidality, and other cognitive and somatic aspects of depression.

The BDI and the HADS-D are the most commonly used methods to screen for symptoms of depression among post-AMI patients.{{3701 Thombs,B.D. 2006; }} Different guidelines have recommended each for this purpose,{{3626 Fallen,E.L. 1995; 4028 Department of Health 2000; }} although comparative evidence regarding the usefulness of each screening measure among post-AMI patients is limited. A great deal of work, however, has been done on screening for depression in primary care settings. A recent systematic review of case-finding instruments for identifying depression in primary care found 38 different studies of 16 different case-finding instruments that ranged in length from 1 to 30 questions, including the BDI, the HADS, and the SCL-90.{{4023 Williams,J.W.,Jr 2002; }} The authors of the review reported that the median sensitivity and specificity for major depression were 85% and 74%, respectively, similar to values reported here for post-AMI patients. Notably, they found that there were no significant differences in accuracy across instruments. Within individual instruments, however, there was significant variation across studies, which is consistent with results reported here from post-AMI samples.

Limitations of this review include variations in the designs of the studies from which data were synthesized, including differences in study size, inconsistencies in in-hospital and follow-up assessment timing and in the relationship between the time of screening and diagnostic assessment, the overall low quality of the studies, and the high percentage of White males included each study, which may limit generalizability. The small number of studies reviewed did not allow for systematic review of potentially important factors, such as the exclusion of older patients{{3688 Boersma,S.N. 2005;543 Dickens,C.M. 2004; 4022 Wojciechowski,F.L. 2000; }} or of patients receiving psychiatric treatment.{{3688 Boersma,S.N. 2005; }} In addition, the

review did not include abstracts, non-published studies, or studies published in non-English language journals. Furthermore, a stronger design for article review might have included measures of interrater reliability and the utilization of a third investigator to resolve any differences rather than consensus by the initial two reviewers.

Practice guidelines recommend that physicians assess for depression among patients with AMI.{{494 Antman,E.M. 2004; }} Evidence from this review and from the larger literature on assessment of depression in primary care{{4023 Williams,J.W.,Jr 2002; }} suggests that many different depression screening tools will likely perform adequately, and that no single instrument has been shown to clearly outperforms the others. Physicians in practice have limited time with each patient and are responsible for screening for a variety of different disorders and conditions. In this context, considerations such as instrument brevity, readability and comprehensibility are of great importance. The BDI, HADS-D, and SCL-90 are each comprised of items with several different response options related to symptom severity, and item response options vary from item to item, increasing complexity for patients. Instruments that require only yes-no responses or estimates of symptom frequency may be easier to administer or for patients to complete independently.{{4023 Williams,J.W.,Jr 2002; }}. The Patient Health Questionnaire – 9 (PHQ-9) is a 9-item measure of depression severity that includes a user-friendly frequency rating response format, and that has been shown to be a highly sensitive (88%) and specific (88%) case-finding tool for depression among primary care patients.{{3764 Kroenke,K. 2001; }} The Patient Health Questionnaire – 2 (PHQ-2) is an even briefer 2-item measure that is also sensitive (83%) and specific (92%) for major depression in primary care. Limitations in existing research on screening tools for depression among post-AMI patients, along with the user-friendly attributes

and strong performance characteristics of instruments like the PHQ-9 and the PHQ-2 among primary care patients, suggest the need for further validation of case-finding instruments for post-AMI patients. Future research should compare the BDI and the HADS-D with the PHQ-9 and PHQ-2, and should attend to important elements of the screening process, including when, where, and how often to screen patients; whether patient characteristics, such as age, gender, and race, influence the accuracy of depression screening in the post-AMI setting; and whether serial screening with more than one instrument improves efficiency and accuracy.

In summary, identification of patients with elevated symptoms of depression may provide an opportunity to improve overall medical care, as well as an opportunity for the diagnosis and treatment of a burdensome health condition. Large studies, such as SADHEART{{3867 Glassman,A.H. 2002; }} and ENRICHD,{{3737 Taylor,C.B. 2005; }} indicate that SSRIs are a safe and potentially efficacious treatment for depression among patients with AMI. Thus, given the limited availability of psychiatry consultation, a reasonable strategy, although one that needs further research, would be to screen initially with one of several short screening tools (1-3 items) that have been validated in primary care settings{{3703 Henkel,V. 2004;3702 Kroenke,K. 2003; }} followed by a more thorough screening tool, such as the BDI, HADS-D, or PHQ-9, and referral for psychiatric consultation or to an affiliated mental health professional for patients with significant symptoms of depression.

REFERENCES

Table 1. Characteristics of Studies Reviewed

Study Author, Year	Study Design	Exclusion Criteria	Study Site	Recruitment Period	Study Objective	Number of Subjects	Mean Age (Years)	Males (%)	Instrument(s) Evaluated	Psychometric Analysis
Martin, 2000 {{380 Martin,C.R. 2000; }}	Cross-sectional	Non-English speaking	Europe	Not Reported	To determine the utility of Hospital Anxiety and Depression Scale in acute AMI patients by examination of the instrument's underlying factor structure.	194	63	73	HADS-D	Validity Reliability
Wojciechowski, 2000 {{4022 Wojciechowski,F .L. 2000; }}	Prospective cohort	Age > 75; previous AMI	Europe	May 1994 - Jan 1996	To investigate whether or not depression and vital exhaustion are separate entities.	143	58	81	SCL-90-Dep ZDS	Validity
Strik, 2001 {{345 Strik,J.J. 2001; }}	Cross-sectional	Recurrent AMI	Europe	May 1997 - Sep 1999	To assess sensitivity and specificity of 3 self-report questionnaires and one observer rating scale as screening instruments for major and minor depression following first AMI.	206	60	76	BDI HADS-D SCL-90-Dep	Diagnostic Utility
Freedland, 2002 {{379 Freedland,K.E. 2002; }}	Prospective cohort	AMI post CABG/invasive procedure; significant other medical; major psychiatric comorbidity; psychiatric medications or psychotherapy	United States	Oct 1996 - Oct 1999	To evaluate the effects of a psychosocial intervention on cardiovascular morbidity and mortality in post-AMI patients exhibiting depression and/or social isolation.	2404	61	56	BDI	Validity
Martin, 2003 {{366 Martin,C.R. 2003; }}	Prospective cohort	Not Reported	Europe	Not Reported	To determine the factor structure and assess change-sensitivity and reliability of the Hospital Anxiety and Depression Scale following AMI.	335	67	67	HADS-D	Validity Reliability

Dickens, 2004 {{543 Dickens,C.M. 2004; }}	Prospective cohort	Previous AMI; age ≥ 80, cognitive impairment; insufficient English; significant psychiatric illness other than depression	Europe	Oct 1997-Nov 1999	To understand the risk factors for depression preceding AMI and that develops in the 12 months post-AMI.	314	58	63	HADS	Diagnostic Utility
Boersma, 2005 {{3688 Boersma,S.N. 2005; }}	Cross-sectional	Age > 70; currently seeing psychiatrist or psychotherapist; congestive heart failure; insufficient Dutch; recent CABG	Europe	Sep 1999-Mar 2002	To investigate whether the experience of an AMI had an impact on important life goals.	160	55	81	HADS-D	Reliability

AMI = acute myocardial infarction; BDI = Beck Depression Inventory; CABG = coronary artery bypass graft; HADS = Hospital Anxiety and Depression Scale, HADS-D = Depression Subscale of the Hospital Anxiety and Depression Scale; SCL-90-Dep = Depression Subscale of the Symptom Checklist 90; ZDS = Zung Depression Scale.

Figure 1.

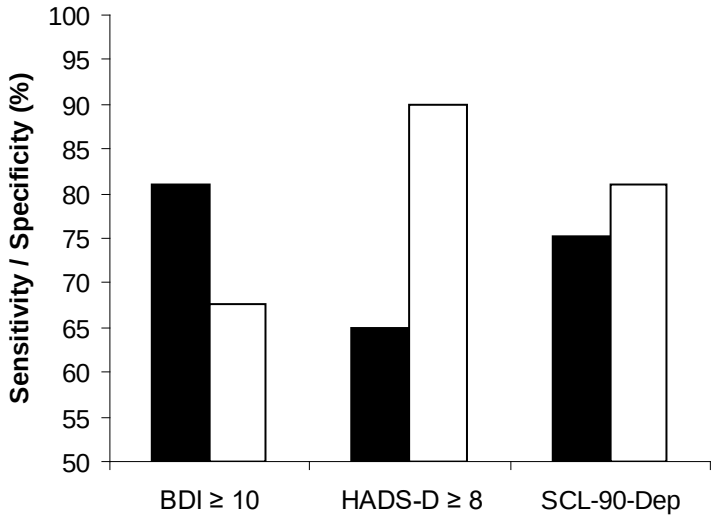
Sensitivity / Specificity (%)
100
95
90
85
80
75
70
65
60
55
50
BDI ≥ 10
HADS-D ≥ 8
SCL-90-Dep

Figure 2.

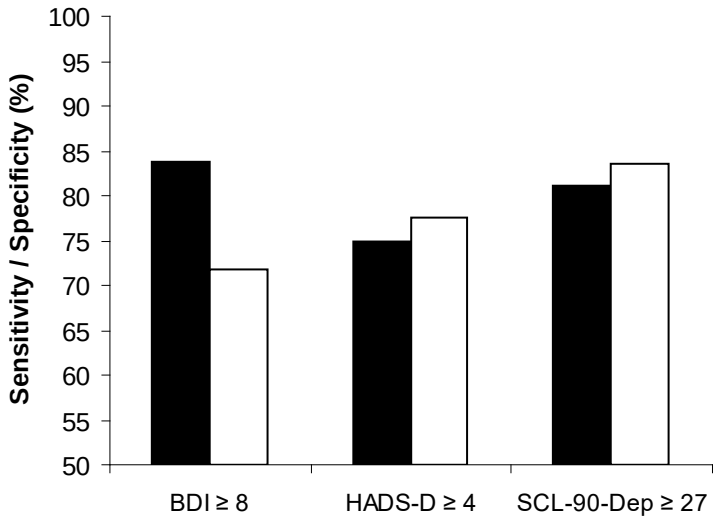
Sensitivity / Specificity (%)
100
95
90
85
80
75
70
65
60
55
50
BDI ≥ 8
HADS-D ≥ 4
SCL-90-Dep ≥ 27

Figure 3.

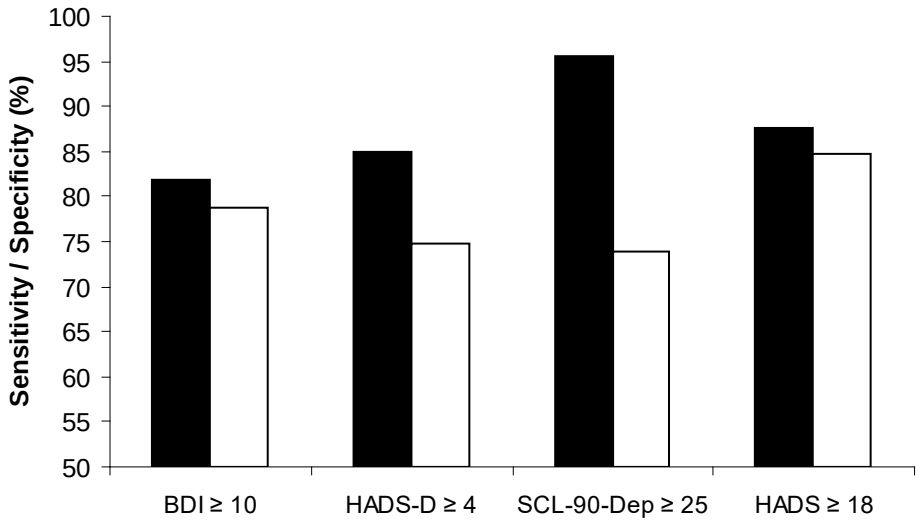
Sensitivity / Specificity (%)
100
95
90
85
80
75
70
65
60
55
50
BDI ≥ 10
HADS-D ≥ 4
SCL-90-Dep ≥ 25
HADS ≥ 18

FIGURE LEGENDS

Figure 1. Sensitivity and Specificity Data for Combined Major and Minor Depression Based on Standard Cutoff Scores

Sensitivity and specificity data for major and minor depression are depicted using standard cutoff scores for the Beck Depression Inventory (BDI ≥ 10), the Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D ≥ 8), and the Depression subscale of the Symptom Checklist-90 (SCL-90-Dep ≥ 23 for men, ≥ 28 for women).{{345 Strik,J.J. 2001; }} Sensitivity data are shown in solid black bars, and specificity data are presented in open bars.

Figure 2. Sensitivity and Specificity Data for Combined Major and Minor Depression Based on Empirically-Derived Cutoff Scores

Sensitivity and specificity data for major and minor depression are depicted using empirically-derived cutoff scores for the Beck Depression Inventory (BDI ≥ 8), the Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D ≥ 4), and the Depression subscale of the Symptom Checklist-90 (SCL-90-Dep ≥ 27).{{345 Strik,J.J. 2001; }} Sensitivity data are shown in solid black bars, and specificity data are presented in open bars.

Figure 3. Sensitivity and Specificity Data for Major Depression Based on Empirically-Derived Cutoff Scores

Sensitivity and specificity data for major depression are depicted using empirically-derived cutoff scores for the Beck Depression Inventory (BDI ≥ 10), the Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D ≥ 4), the Depression subscale of the Symptom Checklist-90 (SCL-90-Dep ≥ 25),{{345 Strik,J.J. 2001; }} and for the overall Hospital Anxiety Depression Scale (HADS ≥ 18).{{543 Dickens,C.M. 2004; }} Sensitivity data are shown in solid black bars, and specificity data are presented in open bars.

APPENDIX A: Literature Search Strategies

Medline

(myocardial infarction[mh] OR myocardial infarct*[tiab]) AND (depression[mh] OR mental disorder[mh] OR mood disorder[mh] OR depression[tiab] OR depressive symptom*[tiab] OR mood disorder[tiab] OR mental disorder[tiab] OR psychiatric disorder[tiab]) AND eng[la] NOT (animal[mh] NOT human[mh])

Cochrane

(myocardial next infarction) and (depression)

EMBASE

'acute heart infarction'/exp OR 'heart infarct'/exp OR 'heart infarction'/exp OR (myocardial AND ('infarct'/exp OR 'infarction'/exp)) AND ('depression'/exp OR 'mood disorder'/exp OR ((mental OR 'mood'/exp OR psychiatric) AND (disorder))) AND [english]/lim AND [humans]/lim AND [embase]/lim

CINAHL

(((myocardial or myocardiac) and (infarct*)) and ((depression) or (mental disorder) or (mood disorder) or (psychiatric disorder) or (depressive symptom))) and (ZL "ENGLISH")

PsychInfo

((myocardial infarct*) and ((depression) or (mental disorder) or (psychiatric disorder) or (depressive symptom))) and (ZL "ENGLISH")

APPENDIX B: Journals Included in Hand Searching

American Heart Journal
American Journal of Cardiology
American Journal of Medicine
American Journal of Psychiatry
Annals of Behavioral Medicine
Archives of General Psychiatry
Archives of Internal Medicine
Biological Psychiatry
Circulation
Health Psychology
JAMA
Journal of Behavioral Medicine
Journal of Cardiopulmonary Rehabilitation
Journal of the American College of Cardiology
Psychosomatic Medicine
Psychosomatics