# Optimizing Detection of Major Depression Among Patients with Coronary Artery Disease Using the Patient Health Questionnaire: Data from the Heart and Soul Study

*Brett D. Thombs, PhD[1], Roy C. Ziegelstein, MD[2], and Mary A. Whooley, MD[3,4,5]*

[1]Department of Psychiatry, McGill University and Sir Mortimer B. Davis-Jewish General Hospital, Montreal, QC, Canada; [2]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; [3]Department of Medicine, University of California, San Francisco, CA, USA; [4]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA; [5], Veterans Affairs Medical Center, San Francisco, CA, USA.

**BACKGROUND:** Clinical guidelines recommend depression screening in patients with coronary artery disease (CAD), but how to accomplish this is unclear.

**OBJECTIVE:** We evaluated the test characteristics of the two-item Patient Health Questionnaire (PHQ-2), the nine-item Patient Health Questionnaire (PHQ-9), and a two-step screening approach (PHQ-2 then PHQ-9 if positive on PHQ-2), compared with the Computerized Diagnostic Interview Schedule (C-DIS) for major depression. We also evaluated a "PHQ diagnosis" of depression, requiring five of nine symptoms "more than half the days," compared with the C-DIS.

**DESIGN:** Cross-sectional study of 1,024 outpatients with CAD.

**MAIN RESULTS:** Two hundred twenty-four patients (22%) had current major depression. Optimal cutpoints were ≥2 for the PHQ-2 (82% sensitive, 79% specific) and ≥6 for the PHQ-9 (83% sensitive, 76% specific). The two-step screening approach was less sensitive (75%), but more specific (84%), than the PHQ-2 or PHQ-9 alone. The "PHQ diagnosis" had low sensitivity (28%), but high specificity (96%).

**CONCLUSIONS:** Cutpoints of ≥2 on the PHQ-2 and ≥6 on the PHQ-9 had similar test characteristics. A two-step approach using the PHQ-2 followed by the PHQ-9 was no better than either instrument alone. A "PHQ diagnosis" of depression had high specificity, but poor sensitivity.

KEY WORDS: diagnostic accuracy; sensitivity; specificity; cardiovascular disease; depression; screening.

## INTRODUCTION

Major depressive disorder (MDD) is present in approximately 20% of patients with cardiovascular disease (CVD).[1,2] Several clinical guidelines[3,4,5] recommend depression screening in patients with CVD, although none specifies what procedures or instruments should be used.

A recent National Heart, Lung, and Blood Institute (NHLBI) Working Group[6] recommended a two-step approach to screening in research studies in which the two-item version of the Patient Health Questionnaire (PHQ-2)[7] is used as an initial screen, and the nine-item version (PHQ-9)[8] is administered to patients positive on the PHQ-2 to identify patients likely to have MDD based on a structured clinical interview. The PHQ-9 is self-administered and easily scored, maps onto the nine symptoms from the DSM-IV classification for MDD, and can be used to track symptoms.[8,9] Recommended cutoff scores to identify patients in primary care who would likely be positive for MDD based on a clinical interview are ≥10 for the PHQ-9 and ≥3 for the PHQ-2.[7]

It has also been suggested that that a "PHQ diagnosis" of MDD can be obtained from the PHQ-9 based on five of nine depressive symptoms present at least half of the days in the past 2 weeks, including depressed mood or anhedonia.[8] However, two systematic reviews[9,10] reported that this method had similar accuracy to the cutoff score method for identifying patients who met MDD criteria based on a structured clinical interview. These reviews found that the PHQ-9 had good sensitivity (77% and 80%) and specificity (94% and 92%) in primary care settings,[9,10] but one review showed that recommended cutoff scores for the PHQ-9 had poor sensitivity in three of six specialty medicine samples (50% to 69%).[10] McManus et al.[11] found that cutoff points recommended for primary care patients resulted in poor sensitivity for the PHQ-2 (39%) and PHQ-9 (54%), compared to a diagnosis of MDD based on the Computerized Diagnostic Interview Schedule (C-DIS),[12] among 1,024 outpatients with stable coronary artery disease (CAD) from the Heart and Soul Study, but did not identify an optimal screening cutoff. Stafford et al.[13] reported that a cutoff score of ≥6 optimized sensitivity (83%) and specificity (79%) among CAD outpatients, but this was based on a relatively small sample (N=193, MDD=35).

The objective of this study was to assess the test characteristics of the PHQ-2 and PHQ-9 compared to an MDD diagnosis using the C-DIS in CAD patients from the Heart and Soul Study using recommended primary care cutoffs, alternative

cutoffs, the two-step approach recommended by the NHLBI Working Group, and a "PHQ diagnosis."

## METHODS

### Patients and Procedures

Methods of the Heart and Soul Study have been described previously.[11] Eligible patients were identified through administrative databases as having CAD, defined as history of MI, angiographic evidence of ≥50% stenosis in ≥1 coronary vessel, previous evidence of exercise-induced ischemia by cardiac stress testing, history of coronary revascularization, and/or diagnosis of CAD by an internist or cardiologist. Invitations to participate in the study were mailed to 15,438 eligible patients; 2,495 responded by mail and received a follow-up telephone call. Of these, 505 could not be reached, 596 declined participation, and 370 were excluded due to an MI in the prior 6 months, self-assessed inability to walk one block, or pending move from the area. Between September 2000 and December 2002, 1,024 patients were enrolled. At their initial study appointment, patients completed the PHQ-2 and PHQ-9 and were assessed for current (past month) MDD with the C-DIS. The appropriate Institutional Review Boards approved all study procedures, and all participants provided written, informed consent.

The PHQ-9[8,14] includes nine items (scored 0–3; total score range 0 to 27). The PHQ-2 includes the first two items of the PHQ-9 (anhedonia and depressed mood) with a total score range of 0 to 6. For the "PHQ diagnosis," subjects were considered depressed if they reported a total of five of nine PHQ symptoms, including anhedonia or depressed mood, "more than half the days" (thoughts of death counted if present at all).[14] The C-DIS was the gold standard used to assess MDD in the previous month[12,15] by research assistants blind to results of the PHQ.

Sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratios, and area under the receiver-operating characteristic curve (AUC)[16] were calculated. Each of 18 patients was missing one item on the PHQ-9. Missing values were imputed using the SPSS Missing Values Analysis module expectation maximization algorithm (version 15.0, Chicago, IL).

## RESULTS

A total of 224 patients (22%) had MDD diagnoses. Patient characteristics, including MDD prevalence by subgroup, are shown in Table 1. As shown in Table 2, and as previously reported by McManus et al.,[11] cutoffs for the PHQ-2 (≥3) and PHQ-9 (≥10) recommended for primary care resulted in good specificity (93% and 90%, respectively), but poor sensitivity (39% and 54%, respectively). Optimal cutpoints were ≥2 for the PHQ-2 (82% sensitive, 79% specific) and ≥6 for the PHQ-9 (83% sensitive, 76% specific). The two-step procedure (PHQ-9 ≥6 for patients with PHQ-2 ≥2) resulted in somewhat lower sensitivity (75%) and somewhat higher specificity (84%) compared to the PHQ-9 or PHQ-2 alone. The "PHQ diagnosis" approach was highly specific (96%), but poorly sensitive (28%). There were no significant differences in AUC or sensitivity and specificity for the PHQ-2 or PHQ-9 based on sex or age (<70 years versus ≥70 years).

**Table 1. Patient Characteristics and Prevalence of Major Depressive Disorder (MDD) in Each Subgroup**

| Characteristic | Total sample (N = 1,024): number (%) | Number (%) in each subgroup with MDD |
|---|---|---|
| Age | | |
| <50 years | 58 (6%) | 21 (36%) |
| 50–59 years | 200 (20%) | 74 (37%) |
| 60–69 years | 322 (31%) | 78 (24%) |
| 70–79 years | 321 (31%) | 35 (11%) |
| 80 +years | 123 (12%) | 16 (13%) |
| Sex | | |
| Female | 184 (18%) | 67 (36%) |
| Male | 840 (82%) | 157 (19%) |
| Race/ethnicity | | |
| Non-Hispanic White | 615 (60%) | 135 (22%) |
| Hispanic | 89 (9%) | 25 (28%) |
| African American | 169 (16%) | 42 (25%) |
| Asian or Pacific Islander | 118 (12%) | 13 (11%) |
| Other | 33 (3%) | 9 (27%) |
| Education | | |
| Less than high school | 131 (13%) | 26 (20%) |
| High school graduate | 182 (18%) | 38 (21%) |
| Some college/junior college/ vocational school | 354 (35%) | 80 (23%) |
| College degree | 182 (18%) | 38 (21%) |
| Graduate/professional degree | 173 (17%) | 42 (24%) |
| Not reported | 2 (<1%) | 0 (0%) |
| Annual income (US dollars) | | |
| <20,000 | 498 (49%) | 141 (28%) |
| 20,000–29,000 | 138 (13%) | 27 (20%) |
| 30,000–39,000 | 95 (9%) | 16 (17%) |
| 40,000–50,000 | 98 (10%) | 14 (14%) |
| >50,000 | 189 (18%) | 26 (14%) |
| Not reported | 6 (1%) | 0 (0%) |
| Marital status | | |
| Married/partner | 436 (43%) | 71 (16%) |
| Divorced | 239 (23%) | 60 (25%) |
| Single | 191 (19%) | 53 (28%) |
| Widowed | 119 (12%) | 27 (23%) |
| Separated | 36 (4%) | 12 (33%) |
| Not reported | 3 (<1%) | 1 (33%) |
| Medical history: | | |
| Hypertension | 723 (71%) | 158 (22%) |
| Myocardial infarction | 547 (54%) | 110 (20%) |
| Coronary revascularization | 602 (59%) | 111 (18%) |
| Congestive heart failure | 179 (18%) | 40 (22%) |
| Stroke | 148 (15%) | 33 (22%) |
| Diabetes mellitus | 265 (26%) | 68 (26%) |
| Regular alcohol use | 293 (29%) | 67 (23%) |
| Current smoking | 201 (20%) | 64 (32%) |
| NYHA classification | | |
| I | 377 (37%) | 62 (16%) |
| II | 416 (41%) | 95 (23%) |
| III | 181 (18%) | 49 (27%) |
| IV | 49 (5%) | 18 (37%) |
| Left ventricular ejection fraction ≤50% | 113 (11%) | 17 (15%) |

*NYHA = New York Heart Association*

## DISCUSSION

In outpatients with stable CAD, we found that either a PHQ-2 cutpoint of ≥2 or a PHQ-9 cutpoint of ≥6 optimized combined sensitivity and specificity for detecting MDD based on a structured clinical interview. A two-step screening approach using both instruments had similar overall diagnostic accura-

**Table 2.** Test Characteristics of Cutoff Scores for the PHQ-2 and PHQ-9, a Two-step Screening Procedure (PHQ-2≥2 Followed by PHQ-9≥6 for Positive Screens on PHQ-2), and a "PHQ Diagnosis" Approach (Self-report of Five of Nine Symptoms More Than Half the Days in Past 2 Weeks, Including Anhedonia or Depressed Mood, with Thoughts of Death Counted if Present at All)

| Screening method | Sensitivity % (95% CI) | Specificity % (95% CI) | Positive predictive value % (95% CI) | Negative predictive value % (95% CI) | Positive likelihood ratio | Negative likelihood ratio |
|---|---|---|---|---|---|---|
| PHQ-2≥1 | 91 (87–94) | 64 (61–68) | 42 (37–46) | 96 (94–98) | 2.56 | 0.14 |
| PHQ-2≥2 | 82 (76–86) | 79 (76–81) | 52 (46–57) | 94 (92–95) | 3.80 | 0.23 |
| PHQ-2≥3 | 39 (33–45) | 93 (90–94) | 59 (51–67) | 84 (82–87) | 5.18 | 0.66 |
| PHQ-9≥4 | 94 (90–97) | 63 (59–66) | 41 (37–46) | 97 (96–99) | 2.52 | 0.09 |
| PHQ-9≥5 | 89 (84–92) | 71 (67–74) | 46 (41–50) | 96 (94–97) | 3.01 | 0.16 |
| PHQ-9≥6 | 83 (78–87) | 76 (73–79) | 50 (45–55) | 94 (92–96) | 3.51 | 0.22 |
| PHQ-9≥7 | 74 (68–79) | 82 (79–84) | 53 (47–58) | 92 (89–94) | 4.01 | 0.32 |
| PHQ-9≥8 | 69 (62–74) | 84 (82–87) | 55 (49–61) | 91 (88–93) | 4.40 | 0.37 |
| PHQ-9≥9 | 61 (54–67) | 88 (85–90) | 58 (52–64) | 89 (86–91) | 4.96 | 0.45 |
| PHQ-9≥10 | 54 (47–60) | 90 (88–92) | 61 (54–67) | 88 (85–90) | 5.54 | 0.51 |
| Two-step screening | 75 (69–81) | 84 (81–86) | 57 (51–62) | 92 (90–94) | 4.68 | 0.29 |
| "PHQ diagnosis" | 28 (22–34) | 96 (94–97) | 65 (55–73) | 83 (80–85) | 6.51 | 0.76 |

*Area under the receiver-operating characteristic curve for PHQ-9=0.86 (0.84–0.89) and for PHQ-2=0.84 (0.81–0.87)*
*CI = confidence interval; PHQ = Patient Health Questionnaire*

cy to using either alone. As compared with a structured clinical interview for MDD, a "PHQ diagnosis" using the PHQ-9 responses to diagnose MDD was highly specific, but resulted in many false negatives.

Our results build on the work of Stafford et al. who examined a group of individuals in Australia 3 months after discharge from hospitalization for an acute MI or a coronary revascularization procedure.[13] They also found that a PHQ-9 cutoff score of ≥6 optimized sensitivity and specificity. We added to this work by demonstrating that the PHQ-2 performs similarly to the PHQ-9 and that two-step screening with the PHQ-2 followed by the PHQ-9 does not improve results compared to screening with either the PHQ-2 or the PHQ-9 alone. An obvious benefit to using the PHQ-2 is its relative brevity. On the other hand, the PHQ-9 may be a better tool for tracking depressive symptoms over time.

We generated cutoff scores that optimized the balance between sensitivity and specificity. In clinical settings, scores can be used to assess depression severity, to monitor the efficacy of treatment, or to identify patients likely to have a diagnosis of MDD based on further assessment. Cutoff points can also be used for research purposes, but not for the formal diagnosis of MDD. A formal diagnosis of MDD requires a clinical interview that assesses specific symptom patterns, as well as evidence of functional limitations.

As demonstrated in primary care settings, improved depression outcomes are likely to occur only when a collaborative care model is used, including the use of evidence-based protocols for treatment, active collaboration between primary care providers and mental health specialists, active monitoring of adherence to therapy, and access to structured psychotherapy.[17] In the absence of these services, there is no evidence that screening alone is of benefit to patients in CVD settings. It must be noted, however, that this conclusion is made from the perspective of depression care alone and does not take into account the possibility that depression screening may have other potential benefits to patients with CVD. Many studies have now shown that patients with positive depression screens are at increased risk of cardiovascular morbidity and mortality.[18] If depression screening identifies a group of high-risk patients who derive particular benefit from certain cardiac

procedures or from interventions focused on enhancing adherence to medication or to secondary prevention behaviors, for example, then screening may be useful in therapeutic decision making even in the absence of mechanisms for formal depression diagnosis, treatment, and follow-up.

It must also be noted that this analysis is based on data from a study of outpatients with stable CAD, and the degree to which conclusions generalize to patients hospitalized with acute coronary syndromes is unknown. Furthermore, since only 7% of eligible patients actually enrolled in the study, results may not generalize well to other groups of CAD patients, although this response rate is comparable to other large cohort studies, such as the Coronary Artery Disease in Young Adults Study and the Cardiovascular Health Study.[19,20] Additional research is needed on screening in acute care settings. Studies are also needed that examine paradigms, such as multiple positive screens prior to initiating formal evaluation, with the goal of reducing the high number of false positives generated in initial screening. Furthermore, clinical management paradigms are needed to establish whether screening in cardiovascular care settings leads to net benefits for patients.

**Corresponding Author:** *Brett D. Thombs, PhD, Department of Psychiatry, McGill University and Sir Mortimer B. Davis-Jewish*

*General Hospital, 4333 Cote Ste Catherine Road, Montreal, QC H3T 1E4, Canada (e-mail: brett.thombs@mcgill.ca).*

# REFERENCES

1. **Thombs BD, Bass EB, Ford DE, et al.** Prevalence of depression in survivors of acute myocardial infarction. J Gen Intern Med. 2006;21:30–38.
2. **Rudisch B, Nemeroff CB.** Epidemiology of comorbid coronary artery disease and depression. Biol Psychiatry. 2003;54:227–40.
3. **Antman EM, Anbe DT, Armstrong PW, et al.** ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction. J Am Coll Cardiol. 2004;44:E1–E211.
4. **Anderson JL, Adams CD, Antman EM, et al.** ACC/AHA 2007 guidelines for the management of patients with unstable angina/non-ST-elevation myocardial infarction. J Am Coll Cardiol. 2007;50:e1–e157.
5. **Mosca L, Banka CL, Benjamin EJ, et al.** Evidence-based guidelines for cardiovascular disease prevention in women: 2007 update. J Am Coll Cardiol. 2007;49:1230–50.
6. **Davidson KW, Kupfer DJ, Bigger JT, et al.** Assessment and treatment of depression in patients with cardiovascular disease: National Heart, Lung, and Blood Institute Working Group report. Psychosom Med. 2006;68:645–50.
7. **Kroenke K, Spitzer RL, Williams JB.** The Patient Health Questionnaire-2: Validity of a two-item depression screener. Med Care. 2003;41:1284–92.
8. **Kroenke K, Spitzer RL, Williams JB.** The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med. 2001;16:606–13.
9. **Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC.** Diagnostic accuracy of the mood module of the Patient Health Questionnaire: A systematic review. Gen Hosp Psychiatry. 2007;29:388–95.
10. **Gilbody S, Richards D, Brealey S, Hewitt C.** Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): A diagnostic meta-analysis. J Gen Intern Med. 2007;22:1596–1602.
11. **McManus D, Pipkin SS, Whooley MA.** Screening for depression in patients with coronary heart disease (data from the Heart and Soul Study). Am J Cardiol. 2005;96:1076–81.
12. **Robins LN, Helzer JE, Croughan J, Ratcliff KS.** National Institute of Mental Health Diagnostic Interview Schedule. Its history, characteristics, and validity. Arch Gen Psychiatry. 1981;38:381–9.
13. **Stafford L, Berk M, Jackson HJ.** Validity of the Hospital Anxiety and Depression Scale and Patient Health Questionnaire-9 to screen for depression in patients with coronary artery disease. Gen Hosp Psychiatry. 2007;29:417–24.
14. **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. JAMA. 1999;282:1737–44.
15. **Blouin AG, Perez EL, Blouin JH.** Computerized administration of the Diagnostic Interview Schedule. Psychiatry Res. 1988;23:335–44.
16. **Hanley JA, McNeil BJ.** The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29–36.
17. U.S. Preventive Services Task Force. Screening for depression: Recommendations and rationale. Ann Intern Med. 2002;136:760–64.
18. **Nicholson A, Kuper H, Hemingway H.** Depression as an aetiologic and prognostic factor in coronary heart disease: A meta-analysis of 6362 events among 146 538 participants in 54 observational studies. Eur Heart J. 2006;27:2763–74.
19. **Friedman GD, Cutter GR, Donahue RP, et al.** CARDIA: Study design, recruitment and some characteristics of the examined subjects. J Clin Epidemiol. 1988;41:1105–16.
20. **Fried LP, Borhani NO, Enright P, et al.** The Cardiovascular Health Study: design and rationale. Ann Epidemiol. 1991;1:263–76.