The PHQ-9 versus the PHQ-8 — Is item 9 useful for assessing suicide risk in coronary artery disease patients? Data from the Heart and Soul Study

Ilya Razykov a,b, Roy C. Ziegelstein f, Mary A. Whooley g,h,i, Brett D. Thombs a,b,c,d,e,*

a Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada
b Department of Psychiatry, McGill University, Montréal, Québec, Canada
c Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montréal, Québec, Canada
d School of Nursing, McGill University, Montréal, Québec, Canada
e Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States
f Department of Medicine, University of California, San Francisco, United States
g Department of Epidemiology & Biostatistics, University of California, San Francisco, United States
h Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montréal, Québec, Canada
i Department of Psychiatry, McGill University, Montréal, Québec, Canada

A R T I C L E   I N F O

Article history:
Received 26 March 2012
Received in revised form 7 June 2012
Accepted 8 June 2012

Keywords:
Coronary artery disease
Depression
PHQ-8
PHQ-9
Suicide

A B S T R A C T

Objective: Item 9 of the Patient Health Questionnaire—9 (PHQ-9), which inquires about both passive thoughts of death and active ideas of self-harm, has been used to assess suicide risk. The objectives of this study were (1) to determine the proportion of patients who responded “yes” to Item 9 who endorsed active suicidal ideation in response to more direct questions from a structured clinical interview and (2) to compare the sensitivity and specificity for detecting cases of depression of the PHQ-9 and the PHQ-8, which does not include Item 9, as well as the correlation between the PHQ-8 and PHQ-9.

Methods: Coronary artery disease (CAD) outpatients were administered the PHQ-9 and the Computerized Diagnostic Interview Schedule (C-DIS). Item 9 responses were compared to suicidal ideation and intent in the last year based on the C-DIS. Scores on the PHQ-8 were obtained by eliminating Item 9 from the PHQ-9. Test characteristics of the PHQ-9 and PHQ-8 were compared.

Results: Of 1022 patients, 110 (10.8%) endorsed Item 9. Of those, only 22 (19.8%) reported thoughts about committing suicide, and only 9 of those (8.1%) reported a suicide plan any time in the last year based on the C-DIS. Correlation between PHQ-9 and PHQ-8 scores was r = 0.997. Sensitivity and specificity for the PHQ-9 (54%, 90%) and PHQ-8 (50%, 91%) to detect major depression were similar.

Conclusion: Item 9 does not appear to be an accurate suicide screen. The PHQ-8 may be a better option than the PHQ-9 in CAD patients.

© 2012 Elsevier Inc. All rights reserved.

Introduction

The Patient Health Questionnaire—9 (PHQ-9) [1] is a self-administered and easily scored measure of depressive symptoms that is comprised of 9 items that map onto the 9 DSM-IV criteria for Major Depressive Disorder (MDD). It is commonly used to assess depressive symptoms among patients in medical settings [2–4], including patients with coronary artery disease (CAD) [5–9]. A US National Heart, Lung, and Blood Institute Working Group [8] recommended using the PHQ-9 in research to identify CAD patients likely to have MDD based on a structured clinical interview, and an American Heart Association (AHA) Science Advisory recommended its use for routine depression screening among patients with CAD in clinical settings [9].

A recent population-based case–control study from Denmark found that patients with and without psychiatric illness were at increased risk of suicide following a myocardial infarction (MI), and the authors concluded that MI patients should be screened for suicidal ideation [10]. A growing number of studies have used Item 9 of the PHQ-9 to estimate the prevalence of suicidal ideation in patients with chronic illness. The prevalence of suicidal ideation based on Item 9 has been reported to be 29% in multiple sclerosis [11], 14% in HIV [12], 8% in cancer [13], and 6% in arthritis [14]. Two studies among patients with heart disease reported that 12% of 886 general cardiology outpatients [15] and 17% of 294 outpatients with congestive heart failure [16] endorsed Item 9. Several authors [11,12,14–16] have recommended that the PHQ-9 be used to detect patients at risk of suicide, and the AHA Science Advisory [9] recommended that all patients who answer “yes” to Item 9 of the PHQ-9 be immediately evaluated for acute suicidality.

Item 9 asks patients “How often have you been bothered by thoughts that you would be better off dead or of hurting yourself in...
some way?” with response options ranging from “not at all” to “nearly every day.” It consists of two parts — thoughts of being better off dead, which is a passive concept that is not necessarily related to self-harm, and thoughts of hurting oneself in some way, which is a more active concept. In medical settings, however, most patients who endorse Item 9 appear to be agreeing with the first part only, passive thoughts about “being better off dead” [17,18], and this appears to be the case even in clinical trials of depressed patients from medical settings [19,20]. In two samples of depressed patients, only 6% of patients who responded “yes” to Item 9 or a similar item indicated that they were “somewhat likely” or “very likely” to attempt suicide without identifying factors preventing them from taking action when they were questioned more specifically [20].

Because of the lack of clarity regarding what Item 9 of the PHQ-9 is assessing and because it may identify many medical patients who have had passive thoughts about death or about self-harm, but a relatively small proportion who have actually considered self-harm, the PHQ-8, which omits Item 9, has been proposed for use in medical populations [19]. Removal of Item 9 has been found to have only a minimal effect on overall scoring. In one study of over 1000 patients from a US Department of Veterans Affairs primary care setting, for instance, PHQ-9 and PHQ-8 scores correlated at r = 0.998. Only 3 patients at or above the standard cutoff score of 10 to identify possible depression on the PHQ-9 had a PHQ-8 score below 10 [17]. In two other validation studies with over 6000 patients total, the PHQ-8 and PHQ-9 correlated at r = 0.997 [19].

If a large proportion of CAD patients who respond “yes” to Item 9 of the PHQ-9 are not deemed to be at risk of suicide based on more specific questions and if the PHQ-8 and PHQ-9 perform similarly with respect to the detection of possible cases of depression in CAD patients, then the PHQ-8 (with a cutpoint of ≥ 10) may be a more appropriate measure for use with CAD patients. If the “suicide item” on the PHQ-9 does not accurately assess suicide risk, then using the PHQ-8 will avoid referring for immediate psychiatric evaluation the many patients with passive thoughts of death or self-harm who are not at risk for suicide. Therefore, the objectives of this study were to (1) to determine the proportion of patients who responded “yes” to Item 9 of the PHQ-9 who also endorsed active suicidal ideation when more specific questions from a structured clinical interview were asked; and (2) to determine the degree of association of PHQ-9 and PHQ-8 scores and the test characteristics of each compared to MDD diagnoses.

Method

Patients and procedures

Methods of the Heart and Soul Study have been described previously [21]. Eligible patients were identified through administrative databases as having CAD, defined as history of myocardial infarction, 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; 2495 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 a myocardial infarction in the prior 6 months, self-assessed inability to walk 1 block, or pending move from the area. Between September 2000 and December 2002, 1024 patients were enrolled. At their initial study appointment, patients completed the PHQ-9 and were assessed for current (past month), 12-month, and lifetime MDD with the Computerized Diagnostic Interview Schedule (C-DIS). The PHQ-9 was self-administered, and the C-DIS was administered by a trained research assistant under the supervision of a clinician. The order of administration varied across participants because 3 participants per day rotated through the 3 different parts of the baseline exam (echo/treadmill, questionnaires, psychiatric interview). When one participant was doing an echo/treadmill, the second would be interviewed, and the 3rd would be completing questionnaires. The appropriate Institutional Review Boards approved all study procedures, and all participants provided written informed consent.

Measures

The PHQ-9 [22] is a 9-item measure for assessment of depressive symptoms that is commonly used in medical populations [2-4], including patients with CAD [5-9]. The 9 items reflect the criteria for MDD as described in the DSM-IV [23] and ask about the presence of symptoms in the past 2 weeks. Items are scored from 0 to 3 with response options “not at all” (score 0), “several days” (score 1), “more than half the days” (score 2) and “nearly every day” (score 3). The maximum total score is 27. Higher scores represent increased severity of depressive symptoms, and the standard cutoff score to identify possible depression is 10 [1-3]. Item 9 asks “Over the last two weeks how often have you been bothered by this problem: thoughts that you would be better off dead or hurting yourself in some way?” In the present study, consistent with previous studies [11-20] and the recent AHA Science Advisory [9], any response other than “not at all” on Item 9 was considered to reflect possible suicidal ideation. PHQ-8 scores were derived from responses to the PHQ-9, but eliminating Item 9.

The C-DIS [24,25] is a structured clinical interview designed for lay administration to assess current, 12-month, and lifetime psychiatric disorders in the general population based on the diagnostic criteria of the DSM-IV [23]. The C-DIS is not a “gold standard” measure of suicidal ideation or suicide risk, but includes a series of questions explicitly directed at suicidal ideation and plan. For patients with a depressive episode in the last 12 months, the C-DIS queries about suicidal thoughts and intent in the worst two-week depressive period in the past 12 months, including: “Did you think about committing suicide?” (yes/no) and “Did you make a plan as to how you might do it?” (yes/no). For the purpose of the current analysis, responses for patients who did not have a depressive episode in the past 12 months and, thus, were not administered these items, were coded as “no.”

Because the C-DIS queries about thoughts of committing suicide and plan for a suicide attempt in the last 12 months, rather than in the last 2 weeks as with the PHQ-9, it does not allow for a direct test of concordance between the two methods. However, it does allow for an assessment of whether Item 9 of the PHQ-9 would generate a high number of positive responses compared to specific questions about suicide, even when the time period for the specific questions on the C-DIS is substantially longer.

Analyses

The number of patients who reported active suicidal thoughts or a plan during a depressive episode during the last 12 months was compared to responses on the PHQ-9 Item 9 using cross-tabulations. Whether or not there was an association between a positive response on Item 9 and reporting thoughts of suicide or a plan based on the C-DIS was assessed with McNemar’s test.

To compare the measurement characteristics of the PHQ-9 and PHQ-8, a Pearson correlation was calculated for the total scores. In addition, using a cutpoint of ≥ 10 for both instruments, sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic curve (AUC) compared to a diagnosis of MDD were calculated [26] with 95% confidence intervals [27] for each measure. Responses on item 9 of the PHQ-9 (“not at all” versus “several days” or more frequently) were compared to responses to items about suicidal ideation and intent from the C-DIS. There were 18 patients who had 1 item missing on
an item other than Item 9 of the PHQ-9. Missing values for the 1 missing item for those patients were imputed using the SPSS Missing Values Analysis module expectation maximization algorithm. No patients were missing data on Item 9. All statistical analyses were done with SPSS statistical software version 20.0 (Chicago, IL).

Results

Patient characteristics

Complete data on all depression assessments and suicide-related inquiries were available for 1022 of 1024 patients in the Heart and Soul Study. Patient sociodemographic and disease characteristics are shown in Table 1. Of the 1022 patients, 223 were diagnosed with current MDD (21.8%).

PHQ-9 Item 9 responses

Of the 1022 patients, 110 (10.8%) responded something other than “not at all” on Item 9 of the PHQ-9 including 83 (8.1%) who responded “several days,” 19 (1.9%) who responded “more than half the days,” and 8 (0.8%) who responded “nearly every day” in the past 2 weeks. Each of the other items of the PHQ-9 was endorsed (any response other than “not at all”) by at least 22.0% of patients. Among the 110 patients with positive responses on Item 9, 66 scored at least 10 on the PHQ-8 (60.0%), compared to 120 other than

The mean PHQ-9 score was 5.2 (standard deviation = 5.4; range 0 to 26), and 198 patients (18.4%) scored 10 or greater. For the PHQ-8, the mean score was 5.0 (standard deviation = 5.2; range 0 to 24), and 186 (18.2%) scored at least 10 (Table 4). The correlation between PHQ-8 and PHQ-9 scores was r = 0.997. Of the 223 patients with current MDD, 120 (53.8%) scored at least 10 on the PHQ-9 versus 112 (50.8%) on the PHQ-8. Thus, 8 patients detected by the PHQ-9 were not detected by the PHQ-8. Of the 790 patients who did not have a current diagnosis of MDD, 721 scored less than 10 on the PHQ-9 (90.2%) versus 725 (90.7%) on the PHQ-8. As shown in Table 5, test characteristics for the PHQ-9 and PHQ-8, using the standard cutoff score of 10 or greater, were very similar. The positive or negative status based on a cutoff of 10 changed for only 12 of 1022 patients if the PHQ-8 rather than the PHQ-9 was used.

Discussion

Only a small portion of patients who endorsed Item 9 of the PHQ-9, which queries thoughts about death or self-harm in the past 2 weeks without distinguishing between the two, reported suicidal thoughts or a plan at some point during a depressive episode over the last 12 months based on questions from a structured clinical interview. Additionally, there was little difference in the performances of the PHQ-8 and PHQ-9 in a sample of patients who completed both the PHQ-9 and PHQ-8. The P4, which is designed to assess suicide risk [20] has important implications for research and clinical care. By using the PHQ-9, urgent attention may be inappropriate diverted to patients who have passive thoughts of death or self-harm but who are not actually at risk of suicide. If the PHQ-9 is used to identify patients who may be depressed or to assess the severity of depressive symptoms, but not as part of a program to assess suicide risk, a positive response on Item 9 could be thought

of as an incidental finding. Incidental findings are discoveries that occur during a clinical investigation or research, which, while potentially important, were not the target of the investigation. When physicians or researchers administer tests or conduct procedures to obtain specific information, they may sometimes unintentionally find themselves in possession of additional information that may not be easily ignored [29–31]. Incidental findings present difficult dilemmas for clinicians and researchers. They offer the possibility of benefit to patients, but often reflect false-positive signals that can set off a cascade of further tests, which are often costly and which pose their own set of potential risks and burdens to patients [29,31].

The decision of whether or not to interpret and follow-up on any incidental findings, including positive responses to Item 9 of the PHQ-9, should be based on a careful consideration of potential benefits and risks of doing so [29,31]. In terms of possible benefit, the results of our study suggest that Item 9 provides ambiguous and not easily interpreted information, and that if one were to use questionnaires or single items to assess suicide risk for research or clinical purposes, Item 9 of the PHQ-9 would be a poor choice. Furthermore, to date, no trials have established that screening for suicide would prevent suicide attempts or even improve mental health. Preventing suicide by identifying individuals at the population level who are at risk is not a straightforward task, and it has been suggested that not enough is known yet to do this effectively [32]. Consistent with this, the United States Preventive Services Task Force has concluded that there is no evidence to support screening for suicide risk [33,34]. Given the content of Item 9, one might wonder if patients who respond positively to Item 9 might benefit from a referral for mental health assessment, regardless of whether there is risk of suicide. Whether or not this might benefit patients in excess of potential harms would similarly need to be studied in a properly conducted trial. However, as described in a recent analysis, there are a number of reasons why it should not simply be assumed that patients would benefit [35].

On the other hand, reviewing and following up on Item 9 responses would certainly result in substantial costs, as well as possible harms or inconvenience to patients. If outpatients with stable CAD were screened for depression using the PHQ-9, this study shows that follow-up on positive responses to Item 9 would require that more than 1 of every 10 of these patients be referred for immediate evaluation to assess suicide risk. It has been suggested that this would not be sustainable [28], and it could divert resources from caring for patients who are known to have mental health care needs, but are not able to readily access services. In a recent study, for example, investigators in Boston posed as depressed patients who had been discharged from an emergency department with instructions to obtain a psychiatric appointment within 2 weeks. Even though these “patients” had excellent health insurance and the study was in a well-resourced area, only 4 of 64 sites called for follow-up appointments offered an appointment within 2 weeks [36].

There are many cases where knowledge of incidental findings has not been shown to benefit patients and may cause harm or place a significant burden on existing resources. It has been argued that patient care may even be improved when access is limited to unsolicited diagnostic information, which is easily misinterpreted and not likely to benefit patients [31]. Consistent with this, a growing number of published studies have used the PHQ-8 rather than the PHQ-9, including studies of pregnant women [37], breast cancer survivors, [38] arthritis patients [14], pulmonary arterial hypertension patients [39], diabetes patients [40–42], HMO patients [43] Veterans Affairs patients [17] and general population samples [44,45]. Two studies [17,19] have compared continuous PHQ-8 and PHQ-9 scores and, consistent with the results of the present study, the sensitivity and specificity of the PHQ-8 and PHQ-9 compared to a diagnosis of MDD based on a validated diagnostic interview were virtually identical for the PHQ-8 and PHQ-9 in a sample of over 1000 Veterans Affairs patients [17].

Our study has limitations that should be considered in interpreting results. First, because the C-DIS is a structured clinical interview for depression, only patients who had a major depressive episode in the last 12 months were assessed for suicidal ideation and intent. Thus, patients who had suicidal ideation or intent outside of the context of an episode of depression would not have been identified, which would have underestimated suicidal ideation and intent. Conversely, since the C-DIS assesses suicidal ideation and intent over the last 12 months, we may have overestimated current suicidal ideation and intent compared to the PHQ-9, which assesses the previous 2 weeks. The PHQ-9 was administered on a self-report basis, whereas the C-DIS was an interview. It is possible that the difference in how questions were delivered may have influenced results. Similarly, it is possible that the C-DIS interview, which was delivered by trained lay interviewers, may have generated somewhat different results than would have been obtained if a clinician had interviewed patients. However, the results that were obtained with these methods were generally similar to results that were obtained in other studies that have accessed whether PHQ-9 Item 9 may be overly general to

<table>
<thead>
<tr>
<th>C-DIS question</th>
<th>PHQ-9 Item 9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
</tr>
<tr>
<td></td>
<td>(n=912)</td>
</tr>
<tr>
<td>Thought about committing suicide</td>
<td>13 (1.4%)</td>
</tr>
<tr>
<td>Made plan for committing suicide</td>
<td>7 (0.8%)</td>
</tr>
<tr>
<td>No thought of or plan for suicide</td>
<td>892 (97.8%)</td>
</tr>
</tbody>
</table>

Table 2
Number of patients endorsing PHQ-9 Item 9 compared to responses for suicide ideation or plan during sad period in past 12 months from the Computerized Diagnostic Interview Schedule (C-DIS)

<table>
<thead>
<tr>
<th>Positive responses on the C-DIS</th>
<th>PHQ-8 score (N=110)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;10 (N=43)</td>
</tr>
<tr>
<td>Thought about committing suicide</td>
<td>3 (7.0%)</td>
</tr>
<tr>
<td>Made plan for committing suicide</td>
<td>2 (4.7%)</td>
</tr>
</tbody>
</table>

Table 3
Number of patients screening positive for depressive symptoms on the PHQ-8 compared to responses for suicide ideation or plan during sad period in past 12 months from the Computerized Diagnostic Interview Schedule (C-DIS)

<table>
<thead>
<tr>
<th>PHQ-8</th>
<th>PHQ-9</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥10</td>
<td>&lt;10</td>
</tr>
<tr>
<td>PHQ-8</td>
<td>186 (18.2%)</td>
</tr>
<tr>
<td>PHQ-9</td>
<td>12 (1.2%)</td>
</tr>
</tbody>
</table>
identify patients at risk of suicide [19,20]. The order of administration of the PHQ-9 and C-DIS varied, so it was not possible to assess if there may have been order effects. Finally, this study did not assess actual suicide outcomes subsequent to PHQ-9 Item 9 and C-DIS responses in the context of a suicide screening program.

With respect to the study sample, this analysis was based on data from a study of outpatients with stable CAD, and the degree to which conclusions generalize to patients in other cardiac settings, patients with non-cardiac illnesses, women or other ethnic groups is unknown. Only 7% of patients who were contacted by mail actually enrolled in the study, a response rate comparable to other large cohort studies, such as the Coronary Artery Disease in Young Adults Study and the Cardiovascular Health Study [46,47]. Of the 1620 patients who responded to the mailing and were confirmed eligible, 1024 (63%) enrolled.

In summary, results from this study show Item 9 of the PHQ-9 is not an accurate means of identifying patients at risk of suicide. Because Item 9 of the PHQ-9 does not appear to differentiate potentially benign thoughts about dying and thoughts of self-harm, acting upon responses to Item 9 could potentially lead to large-scale utilization of mental health care resources so that many patients who have not considered suicide are further evaluated. There is no evidence that doing this would reduce risk of suicide in the very small number of patients who endorse this item who appear to have considered suicide. Given the potential downside of administering Item 9 and because the PHQ-8 and PHQ-9 perform similarly among outpatient CAD patients, the PHQ-8 may be a preferred measure in research and clinical applications in this setting.

Acknowledgments

Mr. Razykov’s work was supported by a McGill University Faculty of Medicine Alexander McFee Fellowship Award. Dr. Ziegelstein was supported by the National Center for Complementary and Alternative Medicine (Grant no. R24AT004641) and the Miller Family Scholar Program of the Johns Hopkins Center for Innovative Medicine. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Complementary and Alternative Medicine or the National Institutes of Health. The Heart and Soul Study was funded by the Department of Veterans Affairs, Washington, DC, the National Heart Lung and Blood Institute (R01 HL797235), Bethesda, MD, the American Federation for Aging Research (Paul Beeson Scholars Program), New York, NY, the Robert Wood Johnson Foundation (Faculty Scholars Program), Princeton, NJ, and the Somatic Research and Education Foundation, South San Francisco, CA.

References


Table 5

<table>
<thead>
<tr>
<th>PHQ version</th>
<th>Sensitivity % (95% CI)</th>
<th>Specificity % (95% CI)</th>
<th>Positive predictive value % (95% CI)</th>
<th>Negative predictive value % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHQ-9</td>
<td>54 (47–60)</td>
<td>90 (88–92)</td>
<td>61 (54–67)</td>
<td>88 (85–90)</td>
</tr>
<tr>
<td>PHQ-8</td>
<td>50 (44–57)</td>
<td>91 (89–93)</td>
<td>60 (53–67)</td>
<td>87 (84–89)</td>
</tr>
</tbody>
</table>


