



Published in final edited form as:

Gynecol Oncol. 2017 June ; 145(3): 562–568. doi:10.1016/j.ygyno.2017.03.024.

Comparison of four brief depression screening instruments in ovarian cancer patients: diagnostic accuracy using traditional versus alternative cutpoints

Eileen H. Shinn, Ph.D.¹, Alan Valentine, M.D.², George Baum¹, Cindy Carmack, Ph.D.³, Kelly Kilgore³, Diane Bodurka, M.P.H., M.D.⁴, Karen Basen-Engquist, M.P.H., Ph.D.¹

Alan Valentine: avalenti@mdanderson.org; George Baum: gpbaum@mdanderson.org; Cindy Carmack: ccarmack@mdanderson.org; Kelly Kilgore: kkilgore@mdanderson.org; Diane Bodurka: dcbodurka@mdanderson.org; Karen Basen-Engquist: kbasenen@mdanderson.org

¹Department of Behavioral Science, The University of Texas M. D. Anderson Cancer Center: Cancer Prevention Building - Unit 1330, 1155 Pressler St. Houston, TX 77030

²Department of Psychiatry, The University of Texas MD Anderson Cancer Center: T. Boone Pickens Academic Tower - Unit 1454, 1515 Holcombe Blvd. Houston, TX 77030

³Department of Palliative Care and Rehabilitation Medicine, The University of Texas MD Anderson Cancer Center: T. Boone Pickens Academic Tower - Unit 1414, 1515 Holcombe Blvd. Houston, TX 77030

⁴Department of Gynecologic Oncology and Reproductive Medicine, The University of Texas M. D. Anderson Cancer Center: Mid Campus Building 1 - Unit 1727, 1515 Holcombe Blvd. Houston, TX 77030

Abstract

Objectives—We compared the diagnostic accuracy of 4 depression screening scales, using traditional and alternative scoring methods, to the gold standard Structured Clinical Interview-DSM IV major depressive episode (MDE) in ovarian cancer patients on active treatment.

Methods—At the beginning of a new chemotherapy regimen, ovarian cancer patients completed the following surveys on the same day: the Center for Epidemiological Studies Depression Scale (CES-D), the Beck Depression Inventory Fast-Screen for Primary Care (BDI-FastScreen), the Patient Health Questionnaire-9 (PHQ-9), and a 1-item screener (“Are you depressed?”). Each instrument's sensitivity, specificity, positive predictive value (PPV) and negative predictive value were calculated with respect to major depression. To control for antidepressant use, the analyses were re-run for a subsample of patients who were not on antidepressants.

Corresponding Author Contact Information: Eileen H. Shinn, Ph.D., Department of Behavioral Science, 1550 Hermann Pressler, Unit 1330, Houston, Texas 77230, (713) 792-0919; (713) 745-5726, eshinn@mdanderson.org.

Conflict of Interest Statement: We have no conflicts of interest to disclose.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Results—One hundred fifty-three ovarian cancer patients were enrolled into the study. Only fourteen participants met SCID criteria for current MDE (9%). When evaluating all patients regardless of whether they were already being treated with antidepressants, the two-phase scoring approach with an alternate cutpoint of 6 on the PHQ-9 had the best positive predictive value (PPV=32%). Using a traditional cutpoint of 16 on the CES-D resulted in the lowest PPV (5%); using a more stringent cutpoint of 22 resulted in a slightly improved but still poor PPV, 7%.

Conclusions—Screening with a two-phase PHQ-9 proved best overall, and its accuracy was improved when used with patients who were not already being treated with antidepressants.

Keywords

depression screening; sensitivity; specificity; SCID

Introduction

Untreated major depression is a critical issue in cancer patient care and survivorship. Research has shown that untreated depression is associated with longer hospital stay,[1] increased pain,[2] reduced adherence to treatment,[3,4] compromised immune functioning, [5,6] and possibly decreased length of survival.[7]

When compared with liaison psychiatrists consulting the same patients, oncologists tend to miss most cases of major depression, with study concordance rates of 23%.[8] With respect to oncologists' attitudes toward depression screening, studies consistently show that oncologists lack confidence in their ability to distinguish between the somatic-based symptoms of depression (loss of appetite, fatigue, and psychomotor retardation) and side effects of cancer treatment and the disease itself.[9] Another frequently cited barrier is the lack of time during oncology treatment visits.[9]

There is a need for an efficient method to reliably detect clinically significant depressive disorders. However, since clinicians lack the training and time to conduct rigorous DSM-based interviews with all of their patients, the next best option may be to use a screening instrument as a first-line approach to detect previously undiagnosed cases of depression. Screening tools are designed to maximize sensitivity, i.e., the likelihood of detecting the presence of a condition among all screened patients. By maximizing sensitivity, actual cases of a condition are not mistakenly missed within the screened population. The IOM spell out and NCCN spell out currently recommend routine screening of all cancer patients for distress and depression, provided follow-up care systems are available. However distress screening in oncology settings has not been widely implemented due to a) most screening instruments have high false positive rates, b) lack of consensus as to the best screening instrument, and c) lack of resources for follow-up after a positive screen test result.

To counter the problem of false positive test results and the unnecessary medical costs that are subsequently engendered, screening instruments should not only have high sensitivity, but also high positive predictive value, (PPV). PPV is a critical parameter for determining the accuracy of a screening test and is defined as the likelihood that a person with a positive test result in fact truly has the disease. Positive predictive values are based upon both the

screening test's sensitivity and, indirectly, its specificity. Specificity refers to the ability of a test to correctly rule out the presence of disease among all patients who are screened. While high sensitivity is to be desired in a screening tool, specificity is also important, in that the more specific the screening method, the less likely an individual who is in actuality disease-free will be falsely identified as having the disease and subsequently referred for additional diagnostic testing.

Few studies compare the performance of various screening instruments against a gold-standard, Structured Clinical Interview using Diagnostic Statistical Manual criteria (SCID)-derived diagnosis of major depressive episode (MDE). As a whole, screening instruments for depression are not interchangeable and have considerable variance in sensitivity, specificity and positive predictive values, which in turn have yielded wide ranges in estimates of probable major depression in cancer patients, 1 to 53%. [10] Other sources of measurement variability include variability in cancer site, and timing of depression assessment. [11,12]

To investigate the performance of brief and ultra-brief screening methods in an oncological setting, we tested the performance of 4 brief screening instruments representing different approaches (the Fastscreen BDI, the CESD, a simple 1-item question, "Are you depressed?" and the PHQ-9) against a diagnosis of MDE using the DSM-IV SCID in patients who were attending treatment or surveillance visits for ovarian cancer. Because a significant proportion of our sample (25%) were already being treated with antidepressants at the time of enrollment, and since previous studies had shown that screening efficiency decrease as rates of antidepressant treatment increase within the screened sample, [13,14] all analyses were then repeated among the subsample of ovarian cancer patients who were not already being treated with antidepressant medication.

We chose to study depression in ovarian cancer patients because treatment and symptom profiles for ovarian cancer overlap with many of the risk factors for depression in cancer: Most women are first diagnosed at an advanced stage of disease, when 5-year survival rates are severely compromised. Treatment for ovarian cancer is often aggressive, requiring repeated regimens of chemotherapy [15-17]. Some studies have found significantly higher levels of depression in ovarian cancer patients compared with patients that have other gynecological cancers [18]. While a few studies have found elevated distress ranging from 23-33% in ovarian cancer patients [16,19,20], the prevalence of major depression in ovarian cancer patients using the gold standard of clinical interviews has not been reported.

Participants

Following Institutional Review Board Approval, ovarian cancer patients beginning a new chemotherapy regimen were enrolled into the study. Patients were eligible if they: a) were beginning a new chemotherapy treatment regimen for ovarian cancer; b) at least 18 years of age; c) spoke and read English; d) were oriented; e) had no other cancer diagnoses; and f) had a Zubrod performance status of 0–3.

Design

Patients were identified prior to their first chemotherapy appointment of a new cycle through online medical record. At the time of the clinical consultation with their gynecologic

oncologist or nurse, the patient was approached for recruitment either in the waiting room. After eligibility was confirmed, the rationale and description of the study were presented and informed consent was obtained if the participant agreed to participate. Participants were prospectively enrolled within the first 3 weeks of a new chemotherapy regimen, which typically lasted 4.5 months.

Telephone SCID interviews were scheduled in advance with the participant, usually 1-2 weeks after the initial consent so that participants would have a chance to recover from the first administration of chemotherapy. The sequence of the depression screening instruments was randomized according to a computerized randomization program (packets were prepared in advance, with screeners placed in randomized order). All screening tools were administered on paper and the SCID was administered via telephone interview on the same day. While participants were allowed to complete other parts of the questionnaire before or after the scheduled SCID telephone interview, they were asked to complete the screening portion of the questionnaire on the same day. If participants indicated that they had not completed the screening instruments on the day of the telephone call, the interviewer gave them 15-20 minutes to complete before calling back to initiate the SCID- depression modules. If the patient did not have time to complete the SCID interview in person at the clinic, an appointment was made to call the patient at home and administer the SCID depression modules over the telephone. Previous studies have shown concordance of telephone-administered diagnostic interviews with face-to-face interviews for assessment of depression.[21] To control for experimenter bias, the interviewer was blinded to the results of the depression screening instruments.

Diagnostic and screening instruments for MDE

1-item depression screening instrument—“*Are you depressed (yes or no)?*” has been reported to have 100% sensitivity and 100% specificity in validation studies done with terminally ill patients.[22] However, this one-item measure has not been validated in ambulatory cancer patients. This screening method was scored dichotomously.

PHQ-9—The PHQ-9 is a 9-item self-administered version of the DSM-based Prime-MD assessing the nine criteria for major depression. [23] For each of the items, patients indicate whether during the previous 2 weeks, the symptoms occurred “not at all,” “several days,” “more than half the days,” or “nearly every day.”

PHQ-9 two-phase scoring—In our study, we scored the PHQ-9 two ways: (1) the conventional scoring strategy of totaling all points and using a cutoff of 10; and (2) a more restrictive 2-phase scoring strategy.[24,25] For the 2-phase scoring, the first two items on the PHQ-9 were scored first to determine for presence of sad mood or anhedonia. The remaining 7 items were scored only if one or both of the items in the first phase was scored positively. A cutoff of 8 points total was used for the resulting score.

Center for Epidemiological Studies Depression Scale (CES-D)—The CES-D is a well-validated, widely-used 20-item self-report measure with possible scores ranging from 0 to 60. A cutpoint of 16 and above is commonly used to indicate clinically significant levels

of depression. It has high internal consistency ($\alpha = .84$ to $.90$) and moderate reliability ($\kappa = .51$ to $.70$). It has good construct validity with other measures of depression.[26] Although the majority of items are devoted to the affective component of MDE (sadness, hopelessness), the CES-D does contain four items assessing somatic issues which are also common in cancer patients without depression (e.g., poor appetite, restless sleep). In our study, we scored the CES-D using the traditional cutpoint of 16 and a higher, more stringent cutpoint of 22.[27,28]

Beck Depression Inventory FastScreen for Primary Care (BDI-FastScreen)—

The BDI-FastScreen is a seven-item self-report measure with a 4-point Likert response scale ranging from 0 to 3. It is distinguished from the traditional Beck Depression Inventory by its omission of somatic items. The BDI-FastScreen sensitivity and specificity rates for identifying major depressive disorder in medical patients have ranged between 82 to 99% and specificity rates of 82 to 97% [29,30], a range considerably higher than average screening instruments' performance.[31] It has high internal consistency ($\alpha = .85$) and has been validated against a DSM-IV based clinical interview, the PRIME-MD Mood module. [32]

SCID-Major Depression portion of the Mood Module—The Structured Clinical Interview for DSM-IV diagnosis of current MDE was used as the gold standard in the calculation of sensitivity and specificity of the screening strategies for depression. The SCID has good inter-rater reliability (kappas ranging from $.70$ to 1.00) and validity.[33] It is frequently used as a gold standard when evaluating the sensitivity and specificity of psychiatric screening instruments.[1] All SCID interviews were conducted by a SCID-certified Ph.D. or Master's level counselor. As part of training, interviewers reviewed the SCID user's guide, received feedback on audiotaped interviews and were scored by second raters until 95% agreement on presence of symptoms was reliably achieved. Current MDE were scored as present if participants scored a 3 on at least five criteria for MDE, with at least one symptom being either sadness or anhedonia.

Antidepressant use—All current medications, reason for use, and dose were collected via participant self-report questionnaire and confirmed by the psychiatrist on the study team (AV). Antidepressant use was scored as present if the participant reported taking any of the following on a daily basis: paroxetine, sertraline, fluoxetine, escitalopram, citalopram, bupropion, venlafaxine, amitriptyline, and nortriptyline. Alprazolam, zolpidem and other anxiety and sleep aids were not counted as antidepressants.

Analysis

For each of the screening approaches, the sensitivity, specificity, positive predictive value and negative predictive values were calculated in detecting: 1) MDE as defined by the SCID and 2) SCID-defined MDE within the subsample of patients who were not already taking antidepressants at the time of assessment. Receiver operating characteristic curve analysis was performed to determine how well the different screening instruments discriminated between participants who were classified as depressed or nondepressed on the SCID for each of the four main analyses. If applicable, alternative cutpoints were chosen from the

ROC analysis so as to maximize the sensitivity and specificity values. These cutpoints were compared to the standard cutpoints found in the literature for each of the screening approaches.

Using an allocation ratio of approximately 10 to 1 (10 nondepressed to every 1 depressed), a sample size of 153 yielded 80% power to detect area under the curve of .7 or greater (the ROC value ranges from 0 to 1.0, with higher values indicating that the scale itself is a good screening tool, with high positive predictive value). We did not do use any compensatory procedures.

Results

Two hundred and thirty-seven eligible patients were approached for enrollment into the study and 160 were enrolled and consented onto the study (68%); 132 completed at least four of the depression screening instruments on the same day as their completion of the telephone-administered SCID and 127 answered all five of the depression screening instruments. The average age was 58 years, 65% of the sample had had at least some college education, and 80% of the sample was non-Hispanic white. Fourteen participants met SCID criteria for having current MDE (9%). For the secondary analysis, after excluding those participants who were already on antidepressants ($n=37$; 24.5% of the sample), five participants met SCID criteria for having current MDE (4.3%).

Results using MDE as a gold standard

PHQ-9—The PHQ-9's reliability for the entire sample was good, $\alpha = .84$. Using the standard cutpoint of 10 for the PHQ yielded a sensitivity of .33, a specificity of .84, a positive predictive value (PPV) of .17 and a negative predictive value (NPV) of .93. The ROC analysis gave an area under the curve of .673, indicating that the PHQ was a poor predictor of SCID-derived MDE, and indicated a best cutpoint of 7 ($\geq 7 =$ depressed). Using this new cutpoint the sensitivity was .58, specificity was .75, the PPV was .18 and the NPV was .95 (Table 2).

PHQ-9 two-phase scoring—The 2-phase scoring method generated a sensitivity of .5, specificity of .89, PPV of .30 and NPV of .95. The ROC analysis area under the curve was .893, indicating that the alternate scoring method was a much stronger predictor of MDE. ROC analysis yielded a best total score cutpoint of 6 for major depressive episode, yielding an area under the curve of .893. Using the new cutpoint, the sensitivity was .83, specificity was .83, PPV was .32 and NPV was .98.

CES-D—The CES-D's reliability for the entire sample was high, $\alpha = .89$. Using the traditional cutpoint 16 generated a sensitivity of 1.0, specificity of .59, PPV of .07 and NPV of 1.0. The second cutpoint 22 generated a lower sensitivity of .75, but higher specificity of .78, PPV of .10 and NPV of .99. ROC analysis indicated a best cutpoint of 22 for depressed, yielding an area under the curve of .829.

BDI-FastScreen—The BDI FastScreen's reliability for the entire sample was acceptable, $\alpha = .78$. Using the recommended cutpoint 4 generated a sensitivity of .60, specificity

of .88, PPV of .17 and NPV of .98. The ROC analysis area under the curve was .920, the highest of all the screening methods. ROC analysis yielded an alternative cutpoint of 3. Using the new cutpoint, sensitivity was improved to 1.0, specificity was .82, PPV was .18 and NPV 1.0.

Results using MDE as a gold standard within patients not taking antidepressant medication

In general, the results were unchanged as to whether the screening approaches were more accurate in predicting MDE in patients who were not already treated with antidepressants. While the ROC area under the curve for the two-phase PHQ-9 approach and for the BDI-FastScreen were improved (.931 and .944 respectively, Table 2), positive predictive values for these two approaches were slightly worse when compared with the PPV's in the full sample.

Results using major or minor depression as the gold standard

Because so few patients met the criteria for MDE, additional analyses were performed to compare the screening methods against SCID-derived diagnoses of either major or minor depression. As with previous analyses, the comparisons of the screening methods for major or minor depressive episode were repeated within the subsample of patients who were not already taking antidepressants. Minor depression was scored as present if participants scored a 3 on either the anhedonia or sadness elements and a 3 on at least one additional criteria for MDE, but did not have a past history of major depressive disorder. An additional two participants met SCID criteria for minor depressive episode, for a total of 16 participants (10%). When comparing with the performance of the screening methods in predicting major depression alone, the overall results, both for the full and untreated patient samples, were quite similar (Table 2).

Discussion

The main findings of the study was that the base rate of major depressive disorder in the overall sample was 9%, which is within the prevalence range found with previous studies using DSM-based clinical interviews to measure major depression. Derogatis's study of 215 cancer patients randomly selected from 3 cancer centers found that 12 patients were diagnosed with major depression (5.5%).[11] In one of the few studies measuring depression with DSM clinical interview, Rhondali et al. found that the prevalence of major depression in a sample of elderly (average age= 78) previously untreated advanced ovarian cancer patients was 15%. These patients, 40% of whom were above the age of 80 at assessment, were much older as a group than our sample, whose average age was 58.[34] Of note, when compared to 22 other cancer patient populations (n=8,265), Brintzenhofe-Szoc et al.'s study found that ovarian cancer patients were found to have the lowest prevalence rate of depression, as measured by the Brief Symptom Index.[35]

Another main finding is that the two-phase scoring algorithm for the PHQ-9 had the best overall performance, slightly more so among participants who were not already taking daily antidepressant medication. On the other hand, when using a cutpoint of 16, the widely-used

CES-D performed the worst. Positive predictive value (PPV) is a critical measure of a screening method's performance and indicates the likelihood that a person with a positive test result in fact truly has the disease. Regardless of whether the full or untreated sample was analyzed or whether MDE alone vs major/minor depression was used as the gold standard, the PPV was invariably highest for the PHQ-9 using the two-phase scoring method compared to the other scoring methods. PPV ranged from .29 to .32, depending on the total score cutpoint and type of analysis, meaning that 29- 32% of those who scored above the cutpoint on the PHQ-9 two-phase scoring method were also diagnosed as having depression (either MDE or MDE/minor depression) on the SCID. In contrast, PPVs were consistently lowest for the CES-D using the standard cutpoint of 16, ranging from .05 to .11. Using a higher cutpoint of 22 did improve the CES-D's performance, but the false positive rate was still high.

While the overall diagnostic accuracy of a screening tool is dependent on the base rate of cases within the population, the performance of these screeners were within range of those found in the literature. Thekkumpurath et al., found that a cutoff of 10 on the PHQ-9 had an overall PPV of 33% with a large mixed-cancer patient sample oversampled to include probably cases of depression.[36] Whitney et al. found that scoring the PHQ-9 as a total score rather than using a two-phase scoring approach resulted in a higher number identified as depressed, but did not test the diagnostic accuracy against a gold-standard.[37]

Because many symptoms of cancer treatment (fatigue, lack of appetite, disturbed sleep) overlap with some criteria for major depression, researchers have argued that these types of somatic items should be omitted from depression screening in cancer patients. Our results did not support this argument. We found that the PHQ, which included somatic items, PHQ performed just as well if not better than the scales which omitted all somatic items (BDI FASTSCREEN). The PHQ-9 two-phase was effective in detecting true cases of MDE from nondepressed women who had high levels of fatigue or poor appetite, perhaps because the first phase assesses for the presence of either of two essential criteria for Major Depressive Episode: feelings of depression or anhedonia. Patients who did not exhibit either of these criteria did not move on to the second phase of the PHQ-9 and received a negative depression screening test result. Thus, the somatic items of sleeplessness, fatigue and poor appetite were not asked unless patients had already endorsed sadness or anhedonia or both.

While the CES-D cutpoint of 16 had perfect sensitivity, its false positive rate was 95% for major depression and 89% for major and minor depression, rendering it an inefficient and costly screening method to detect clinical depression among ovarian cancer patients in clinical settings. In other studies, the positive predictive values for the CES-D of 16 has been found to range from a low of .26 in a sample of 79 breast cancer patients[39] to .55 in a sample of 60 head and neck cancer patients.[40]

The one-item screener, "Are you depressed? Yes or no" is often employed in clinical settings by time-pressed healthcare providers. While this approach is efficient in correctly ruling out patients who were not truly depressed (NPV's ranged from .96 to .99), it missed an alarmingly high rate (80%) of all true cases of MDE or MDE/minor depression, suggesting that it is an oversimplified measure of depression. As the other screening methods were

similarly efficient in identifying patients who were truly not depressed, the one-item screener did not exhibit any special performance characteristics to justify its use. We were unable to find other studies that measured the diagnostic accuracy of the one-item screener. Akizuki et. al tested a one-item screening tool “Please grade your mood during the past week by assigning a score from 0 to 100” against a DSM-based clinical interview; Using a cutoff of 20%, the one-item screener had a sensitivity of 80% and specificity of 61% and PPV of .34 in the detection of major depression and adjustment disorder combined.[38]

Finally, while the BDI-FastScreen outperformed the PHQ-9 when using a traditional cutpoint of 10, its sensitivity and positive predictive value were not as strong as the PHQ-9 two-phase scoring method, and the BDI Fast-Screen's performance metrics were reduced significantly when assessing participants who were not already taking antidepressants.

Importantly, the PHQ-9 two-phase screening instrument had superior diagnostic performance compared with the other screening methods. It should be noted that the two-phase scoring method for the PHQ-9 can be easily implemented using computerized platforms, such as REDCap, using skip patterns so that the patient would not have to answer the second phase of questions if he or she did not score above the threshold on the first phase of the PHQ-9. This in turn may lead to higher completion rates since the majority of patients would only have to complete 2 questions rather than all 9. If the patient completes a paper-based version of the PHQ-9, the two-phase scoring method rather than the traditional summation of all 9 items can be applied to gain a more accurate screening result for major depressive episode.

With very few exceptions, existing studies do not exclude cases that are already detected and in treatment.[13,14] While the goal of screening for depression should be to detect patients whose depression is undetected and untreated, it is interesting to note that 9 of the 14 true cases were already being treated with antidepressants, yet still met criteria for major depression. That patients were being treated with antidepressants and still meeting DSM criteria for a major depressive episode indicates that a substantial number of patients who were depressed and receiving antidepressant medication were not being treated optimally; either higher doses of medication may have been needed or perhaps their medication needed to be combined with psychotherapy. [41]

Limitations

Our results should not be generalized to all cancer patients, as this study restricted its sample to patients with ovarian cancer at a comprehensive cancer center. Second, the prevalence of MDE in our sample, 9%, was lower than that found in other studies assessing DSM-derived diagnoses of major depression in gynecologic cancers, 15-23%.[42] It was also lower than an overall estimate, 14.3%, cited in Mitchell et al.'s meta analytic review of 4007 oncological palliative care patients.[43] This low prevalence rate had the effect of suppressing the range of PPV's within our study, which again, were lower than those found in other evaluations of depression screening instruments' accuracy.[31]

Conclusion

Using a two-phase scoring method for the PHQ-9, with an alternate cutpoint of 8 resulted in the best diagnostic performance among the four screening instruments, whether a) the gold standard was major depressive episode, vs. MDE and minor depressive episode, or b) the full sample or untreated participants alone were analyzed. The traditional cutpoint of 16 on the CES-D and the one-item screener were among the worst methods and are not recommended as first-line screening methods in oncological settings. Diagnostic performance of the screening methods was slightly improved when screening patients who were currently being treated with antidepressants. The next steps in this research would be to test the utility of the two-phase screening approach in a larger, more diverse sample than typically found at M. D. Anderson, to determine whether the testing process is burdensome for staff and patients, and whether it results in reasonable rate of uptake of depression treatment in patients identified as depressed.

Acknowledgments

Funding Sources: This study was supported by the following grants: NCI K07 CA 093512 and the Lance Armstrong Foundation.

References

1. Pearson S, Katzelnick D, Simon G, Manning W, Helstad C, et al. 1999; Depression among high utilizers of managed care. *Journal of General Internal Medicine*. 14:461–468. [PubMed: 10491229]
2. Wells K, Stewart A, Hays R, Burnam A, Rogers W, et al. 1989; The functioning and well-being of depressed patients. *Journal of the American Medical Association*. 262:914–919. [PubMed: 2754791]
3. Stoudemire A, Thompson T. 1983; Medication noncompliance: systematic approaches to evaluation and intervention. *General Hospital Psychiatry*. 5:233–239. [PubMed: 6662352]
4. Ayres A, Hoon P, Franzoni J, Matheny K, Cotanch P, et al. 1994; Influence of mood and adjustment to cancer on compliance with chemotherapy among breast cancer patients. *J of Psychos Res*. 38:393–402.
5. Miller G, Cohen S, Hebert T. 1999; Pathways linking major depression and immunity in ambulatory female patients. *Psychosomatic Medicine*. 60:850–860.
6. Fawzy F, Fawzy N, Arndt L, Pasnau R. 1995; Critical review of psychosocial interventions in cancer care. *Archives of General Psychiatry*. 52:100–113. [PubMed: 7848046]
7. NCI. NCI Common Scientific Outline [Online source]. NCI; 2000.
8. Berard R, Boormeester F, Viljoen G. 1998; Depressive disorders in an out-patient oncology setting: prevalence, assessment, and management. *Psycho-Oncology*. 7:112–120. [PubMed: 9589509]
9. Passik S, Dugan W, McDonald M, Rosenfeld B, Theobald D, et al. 1998; Oncologists' recognition of depression in their patients with cancer. *Journal of Clinical Oncology*. 16:1594–1600. [PubMed: 9552071]
10. Trask, P. *Journal of the National Cancer Institute Monographs*. 2004. Assessment of depression in cancer patients; 80–92.
11. Derogatis L, Morrow G, Fetting J. 1983; The prevalence of psychiatric disorders among cancer patients. *Journal of the American Medical Association*. 249:751–757. [PubMed: 6823028]
12. Zabora J, Brintzenhofesoc K, Curbow B, Hooker C, Piantadosi S. 2001; The prevalence of psychological distress by cancer site. *Psycho-Oncology*. 10:19–28. [PubMed: 11180574]
13. Coyne J, Palmer S, Shapiro P, Thompson R, DeMichele A. 2004; Distress, psychiatric morbidity, and prescriptions for psychotropic medication in a breast cancer waiting room sample. *General Hospital Psychiatry*. 26:121–128. [PubMed: 15038929]

14. Palmer S, Taggi A, DeMichele A, Coyne J. 2012; Is screening effective in detecting untreated psychiatric disorders among newly diagnosed breast cancer patients? *Cancer*. 118:2735–2743. [PubMed: 21989608]
15. Rubin, SC, Benjamin, I, Berek, JS. Secondary cytoreductive surgery. In: Gershenson, D, McGuire, WP, editors. *Ovarian Cancer: Controversies in Management*. New York, NY: Churchill Livingstone, Inc; 1998. 101–114.
16. Kornblith A, Thaler H, Wong G, Vlamis V, LePore J, et al. 1995; Quality of life of women with ovarian cancer. *Gynecologic Oncology*. 59:231–242. [PubMed: 7590479]
17. Sugimoto, A, Thomas, G. Early-stage ovarian carcinoma. In: Kavanah, J, Singletary, S, Einhorn, N, DePetrillo, A, editors. *Cancer in Women*. Malden: Blackwell Science; 1998.
18. Cain EN, Kohorn EI, Quinlan DM, Schwartz PE, Latimer K, et al. 1983; Psychosocial reactions to the diagnosis of gynecologic cancer. *Obstetrics & Gynecology*. 62:635–641. [PubMed: 6621954]
19. Bodurka-Bevers D, Basen-Engquist K, Carmack C, Fitzgerald M, Wolf J, et al. 2000; Depression, anxiety and quality of life in patients with epithelial ovarian cancer. *Gynecologic Oncology*. 78:302–308. [PubMed: 10985884]
20. Portenoy R, Kornblith A, Wong G, Vlamis V, Lepore J, et al. 1994; Pain in ovarian cancer patients: prevalence, characteristics, and associated symptoms. *Cancer*. 74:907–915. [PubMed: 8039118]
21. Wells K, Leake B, Robins L. 1988; Agreement between face-to-face and telephone administered versions of the depression section of NIMH Diagnostic Interview Schedule. *Journal of Psychiatric Research*. 22:207–230. [PubMed: 3225790]
22. Chochinov H, Wilson K, Enns M, Lander S. 1997; “Are you depressed?” screening for depression in the terminally ill. *American Journal of Psychiatry*. 154:674–676.
23. Spitzer RL, Kroenke K, Williams JB. group tPHQs. 1999; Validation and utility of a self-report version of PRIME-MD: The PHQ Primary Care Study. *Journal of the American Medical Association*. 282:1737–1744. [PubMed: 10568646]
24. Rost K, Duan N, Rubenstein L, Ford D, Sherbourne C, et al. 2001; The Quality Improvement for Depression collaboration: General analytic strategies for a coordinated study of quality improvement in depression care. *General Hospital Psychiatry*. 23:239–253. [PubMed: 11600165]
25. Rost K, Nutting P, Smith J, Coyne J. 2000; The role of competing demands in the treatment provided for primary care patients with major depression. *Archives of Family Medicine*. 9:150–154. [PubMed: 10693732]
26. Radloff LS. 1977; The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*. 1:385–401.
27. Gilbertson-White S, Aouizerat B, Jahan T, Paul S, West C, et al. 2012; Determination of cutpoints for low and high number of symptoms in patients with advanced cancer. *Journal of Palliative Medicine*. 15:1027–1036. [PubMed: 22853731]
28. Krebber A, Buffart L, Kleijn G, Riepma I, de Bree R, et al. 2014; Prevalence of depression in cancer patients: a meta-analysis of diagnostic interviews and self-report instruments. *Psycho-Oncology*. 23:121–130. [PubMed: 24105788]
29. Beck A, Guth D, Steer R, Ball R. 1997; Screening for major depression in medical inpatients with the Beck depression inventory for primary care. *Behavioral Research and Therapy*. 35:785–791.
30. Steer R, Cavalieri T, Leonard D, Beck A. 1999; Use of the Beck Depression Inventory for Primary Care to screen for major depression in disorders. *General Hospital Psychiatry*. 21:106–111. [PubMed: 10228890]
31. Mulrow C, Williams J, Gerety M, Ramirez G, Montiel O, et al. 1995; Case-finding instruments for depression in primary care settings. *Annals of Internal Medicine*. 122:913–921. [PubMed: 7755226]
32. Beck, A, Steer, R. *BDI-Fast Screen for Medical Patients Manual*. San Antonio: Harcourt Assessment Company; 2000.
33. First, M, Gibbon, M, Spitzer, R, Williams, J. *User’s guide for the structured clinical interview for DSM-IV Axis I Disorders—research version*. New York: Biometrics Research; 1996.
34. Rhondali W, Freyer G, Adam V, Filbet M, Derzelle M, et al. 2015; Agreement for depression diagnosis between DSM-IV-TR criteria, three validated scales, oncologist assessment, and

- psychiatric clinical interview in elderly patients with advanced ovarian cancer. *Clin Interv Aging*. 10:1155–1162. [PubMed: 26203235]
35. Brintzenhofe-Szoc KM, Levin TT, Li Y, Kissane DW, Zabora JR. 2009; Mixed anxiety/depression symptoms in a large cancer cohort: prevalence by cancer type. *Psychosomatics*. 50:383–391. [PubMed: 19687179]
36. Thekkumpurath P, Walker J, Butcher I, Hodges L, Kleiboer A, et al. 2011; Screening for major depression in cancer outpatients: the diagnostic accuracy of the 9-item patient health questionnaire. *Cancer*. 117:218–227. [PubMed: 20737537]
37. Whitney KA, Steiner AR, Lysaker PH, Estes DD, Hanna NH. 2010; Dimensional Versus Categorical Use of the PHQ-9 Depression Scale Among Persons with Non-Small-Cell Lung Cancer: A Pilot Study Including Quality-of-Life Comparisons. *The Journal of Supportive Oncology*. 8:219–226. [PubMed: 21086881]
38. Akizuki N, Akechi T, Nakanishi T, Yoshikawa E, Okamura M, et al. 2003; Development of a brief screening interview for adjustment disorders and major depression in patients with cancer. *Cancer*. 97:2605–2613. [PubMed: 12733160]
39. Pasacrete J. 1997; Depressive phenomena, physical symptom distress, and functional status among women with breast cancer. *Nursing Res*. 46:214–221.
40. Katz MR, Kopek N, Waldron J, Devins GM, Tomlinson G. 2004; Screening for depression in head and neck cancer. *Psychooncology*. 13:269–280. [PubMed: 15054731]
41. Thombs B, Arthurs E, El-Baalkbaki G, Meijer A, Ziegelstein R, et al. 2011; Risk of bias from inclusion of patients who already have diagnosis of or are undergoing treatment for depression in diagnostic accuracy studies of screening tools for depression: systematic review. *BMJ*. 343:d4825. [PubMed: 21852353]
42. Evans D, McCartney C, Nemeroff C, Raft D, Quade D, et al. 1986; Depression in women treated for gynecological cancer: Clinical and neuroendocrine assessment. *American Journal of Psychiatry*. 143:447–452.
43. Mitchell A, Chan M, Bhatti H, Halton M, Grassi L, et al. 2011; Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a meta-analysis of 94 interview-based studies. *Lancet Oncology*. 12:160–174. [PubMed: 21251875]

Highlights

- The diagnostic accuracy of different depression screening methods was compared
- A two-phase scoring approach using a cutpoint of 6 on the PHQ-9 performed best.
- The CES-D (cutpoint = 16) performed worst, with a positive predictive value of 5%.
- The one-item screener “Are you depressed?” missed 80% of all true depressed cases.
- The results were similar when analyzed with patients not on antidepressants

Table 1

Demographics (n=153)

Category	
Age (Mean, Std Dev)	58, 11
	N (%)
Current Antidepressant Use	
Yes	39 (25)
No	114 (75)
Race	
Hispanic White	16 (10)
Non Hispanic Black	13 (8)
Non Hispanic White	123 (80)
Non Hispanic Asian	2 (1)
Education	
No HS diploma	7 (5)
HS/GED/Vocational	46 (30)
Some college or 2-year college degree	38 (25)
College degree or Higher	62 (40)
Marital/partner status	
Married/Living with Partner	102 (66)
Single/Divorced/Widowed	51 (34)
Religious Group Membership	
Yes	124 (81)
No	29 (19)
AJCC Stage	
I-II	22 (14)
III-IVb	131 (86)
Newly Diagnosed	
Yes	57 (37)
No	96 (63)

Table 2

Screening Performance using Major Depression as the Gold Standard, for All Participants in the Study and for Participants not Treated with Antidepressants.

- a. Major Depressive Episode, All Participants
- b. Major Depressive Episode, Subsample of Participants Not on Antidepressants
- c. Major + Minor Depressive Episode, All Subjects
- d. Major + Minor Depressive Episode, Subsample of Participants Not on Antidepressants

Instrument	ROC Area under curve (SE)	Performance with traditional cutpoint				Performance with ROC suggested best cutpoint					
		ROC suggested Cut point	Sen (95%CI)	Spec (95%CI)	PPV (95%CI)	NPV (95%CI)	Cut point	Sen (95%CI)	Spec (95%CI)	PPV (95%CI)	NPV (95%CI)
Major Depressive Episode, All Participants											
PHQ-9	.673 (.084)	10	.33 (.11,.65)	.84 (.76,.90)	.17 (.05,.38)	.93 (.86,.97)	7	.58 (.29,.4)	.75 (.66,.82)	.18 (.08,.34)	.95 (.88,.98)
PHQ-9 2-Phase Scoring ^a	.893 (.029)	8	.5 (.22,.78)	.89 (.82,.94)	.3 (.13,.54)	.95 (.89,.98)	6	0.83 (.51,.97)	0.83 (.76,.89)	.32 (.17,.51)	.98 (.93,.996)
CESD 16	.829 (.081)	16	1 (.40, 1)	.59 (.49,.67)	.07 (.02,.18)	1 (.94,1)	22	.75 (.22,.99)	.78 (.70,.85)	.10 (.03,.28)	.99 (.94,.999)
CESD 22	.829 (.081)	22	.75 (.22,.99)	.78 (.70,.85)	.10 (.03,.28)	.99 (.94,.99)	22	.75 (.22,.99)	.78 (.70,.85)	.10 (.03,.28)	.99 (.94,.999)
BDI	.920 (.035)	4	.6 (.17,.93)	.88 (.81,.93)	.17 (.04,.42)	.98 (.93,.999)	3	1 (.46,1)	.82 (.74,.88)	.18 (.07,.38)	1 (.96,1)
One-Item	No ROC yes/no variable		.2 (.01,.70)	.96 (.90,.99)	.17 (.01,.64)	.97 (.91,.99)					
Major Depressive Episode, Subsample of Participants Not on Antidepressants											
PHQ-9	.606 (.149)	10	.4 (.07,.83)	.84 (.75,.90)	.11 (.02,.36)	.96 (.89,.99)	No good cut point				
PHQ-9 2-Phase scoring ^a	.931 (.033)	8	.80 (.30,.99)	.88 (.79,.93)	.25 (.08,.53)	.99 (.93,.99)	7	1 (.46,1)	.88 (.79,.93)	.29 (.11,.56)	1 (.95,1)
CESD 16	.822 (.130)	16	1 (.20,1)	.63 (.52,.72)	.05 (.01,.20)	1 (.92,1)	16	1 (.20,1)	.63 (.52,.72)	.05 (.01,.20)	1 (.92,1)
CESD 22	.822 (.130)	22	.5 (.03,.97)	.79 (.69,.86)	.05 (.002,.26)	.99 (.92,.999)	16	1 (.20,1)	.63 (.52,.72)	.05 (.01,.20)	1 (.92,1)
BDI	.944 (.045)	4	.5 (.03,.97)	.92 (.84,.96)	.11 (.01,.49)	.999 (.93,.999)	3	1 (.20,1)	.86 (.77,.92)	.13 (.02,.40)	1 (.95,1)

Instrument	ROC Area under curve (SE)	Performance with traditional cutpoint						Performance with ROC suggested best cutpoint					
		ROC suggested Cut point	Sen (95%CI)	Spec (95%CI)	PPV (95%CI)	NPV (95%CI)	Cut point	Sen (95%CI)	Spec (95%CI)	PPV (95%CI)	NPV (95%CI)		
One-Item	No ROC yes/no variable		.5 (.03,.97)	.96 (.89,.99)	.2 (.01,.70)	.99 (.93,.999)							
Major + Minor Depressive Episode, All Subjects													
PHQ-9	.655 (.075)	10	.29 (.10,.58)	.84 (.76,.90)	.17 (.05,.38)	.91 (.84,.96)	No good cut point						
PHQ-9 2-Phase Scoring ^A	.843 (.050)	8	.43 (.19,.70)	.89 (.82,.94)	.3 (.13,.54)	.93 (.87,.97)	6	.71 (.42,.90)	.83 (.75,.89)	.32 (.17,.51)	.96 (.90,.99)		
CESD Cutpoint 16	.817 (.059)	16	1 (.52,1)	.60 (.50,.68)	.11 (.05,.23)	1 (.94,1)	19	.83 (.36,.99)	.72 (.63,.80)	.13 (.05,.28)	.99 (.93,.999)		
CESD Cutpoint 22	.817 (.059)	22	.67 (.24,.94)	.79 (.70,.85)	.13 (.04,.32)	.98 (.92,.996)	19	.83 (.36,.99)	.72 (.63,.80)	.13 (.05,.28)	.99 (.93,.999)		
BDI**	.802 (.092)	4	.43 (.12,.80)	.88 (.81,.93)	.17 (.04,.42)	.97 (.91,.999)	3	.71 (.30,.95)	.82 (.73,.88)	.18 (.07,.38)	.98 (.93,.996)		
One-Item	No ROC yes/no variable		.17 (.01,.64)	.96 (.90,.98)	.17 (.01,.64)	.96 (.90,.98)							
Major + Minor Depressive Episode, Subsample of Participants Not on Antidepressants													
PHQ-9	.597 (.110)	10	.29 (.05,.70)	.83 (.74,.90)	.11 (.02,.36)	.94 (.86,.98)	No good cut point						
PHQ-9 2-Phase Scoring	.821 (.088)	8	.57 (.20,.89)	.88 (.79,.93)	.25 (.08,.53)	.97 (.90,.99)	7	.71 (.30,.95)	.88 (.79,.93)	.29 (.11,.56)	.98 (.91,.99)		
CESD 16	.815 (.070)	16	1 (.40,1)	.64 (.53,.74)	.11 (.04,.26)	1 (.92,1)	19	.75 (.22,.99)	.75 (.65,.83)	.12 (.03,.31)	.99 (.91,.999)		
CESD 22	.815 (.070)	22	.5 (.09,.91)	.79 (.69,.87)	.10 (.02,.32)	.97 (.90,.999)	19	.75 (.22,.99)	.75 (.65,.83)	.12 (.03,.31)	.99 (.91,.999)		
BDI	.745 (.143)	4	.25 (.01,.78)	.92 (.84,.96)	.11 (.01,.49)	.97 (.90,.99)	0	.75 (.22,.99)	.59 (.49,.69)	.07 (.02,.21)	.98 (.90,.999)		
One-Item	No ROC yes/no variable		.33 (.02,.87)	.96 (.89,.99)	.20 (.01,.70)	.98 (.92,.999)							

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

PHQ traditional= items 1 – 9 are added, scores of 10 are depressed.

PHQ 2 phase alternate scoring method: If PHQ 1 or 2 is 1 or greater, add PHQ 3 – 9, this is the score. If PHQ 1 or 2 is 0 then the score is 0.

CESD cut at 16: Scores of 16 are depressed

CESD cut at 22: Scores of 22 are depressed

BDI: Sum of BDI 1 – 7. Scores of 4 are depressed"