Published in final edited form as:
J Affect Disord. 2015 January ; 170: 131–137. doi:10.1016/j.jad.2014.09.002.

A Differential Item Function Analysis of Somatic Symptoms of Depression in People with Cancer

Salene M. W. Jones, **Evette J. Ludman**, **Ruth McCorkle**, **Robert Reid**, **Erin J. Aiello Bowles**, **Robert Penfold**, and **Edward H. Wagner**
Group Health Research Institute, 1730 Minor Ave, Seattle, WA, 98101, USA

Abstract

Background—The overlap of somatic symptoms of depression with symptoms of cancer treatment is widely acknowledged and studied. However, this literature provides little guidance for clinicians as to whether these items should be used in assessing depression. The current study examined the appropriateness of using somatic items for assessment of depression in people with cancer.

Methods—People with newly diagnosed breast, lung or colorectal cancer (n=251) completed the Patient Health Questionnaire-9 (PHQ9) shortly after cancer diagnosis but before cancer treatment (baseline), 4 months later, typically during or shortly after treatment, and 12 months later. Pharmacy data was used to classify participants as having low somatic symptoms or high somatic symptoms. Differential item function (DIF) compared the functioning of the somatic items of the PHQ9 in the low vs. high symptom groups and the chemotherapy vs. no chemotherapy groups at the 4-month assessment.

Results—Significant DIF was not found on any of the four somatic items of the PHQ9 and differences in the item parameters of the somatic items was not consistent across the groups. However, fatigue and sleep indicated only mild depression. Only removing the fatigue item greatly affected the number screening positive for depression at 4 months (8.3%) but removing the other somatic items did not have as large an effect. Only one participant at baseline screened

Conflict of Interest: None.

Author contributions (all authors have approved this manuscript for submission):

Salene M. W. Jones- formulated research question, conducted analyses, primary writer of the manuscript

Evette J. Ludman- design and conduct of the parent study, assisted with formulating research question and drafting manuscript

Ruth McCorkle- designing parent study, assisted with drafting of the manuscript including suggestions for the analyses

Robert Reid- designing parent study, assisted with drafting of the manuscript including suggestions for the analyses

Erin J. Aiello Bowles- designing parent study including data analyses, drafting of the manuscript

Robert Penfold- designing parent study, drafting of manuscript

Edward H. Wagner- principle investigator of the parent study, designed and conducted parent study, drafting of the manuscript

positive for depression by somatic symptoms alone (no psychological symptoms) and no participants screened positive by somatic symptoms alone at 4 months and 12 months.

Limitations—The sample size was small for DIF and consisted of mostly women with breast cancer.

Conclusions—Somatic symptoms of depression can continue to be administered to people with cancer, however the fatigue and sleep items should be used with caution.

Keywords

neoplasm; depression; somatic symptoms

Introduction

Previous research has shown that depressive symptomatology is common in people with cancer (Mitchell et al., 2011); however somatic symptoms of depression overlap with common symptoms of cancer and cancer treatments (Trask, 2004). These overlapping symptoms, manifesting both in depression and cancer, include fatigue, sleep disturbance, appetite changes and perceived cognitive disturbance. Several methods have been proposed to account for the potential symptom overlap (Trask, 2004). The first approach, called the inclusive approach, counts any symptom reported toward a diagnosis of depression, regardless of the cause, while the second approach, the etiologic approach, only counts symptoms if the psychological disorder is clearly contributory. Other approaches are the substitutive approach that replaces somatic symptoms with additional psychological symptoms (brooding, etc.) and the exclusive approach that disregards somatic symptoms without replacement. A specific example of the substitutive approach is the Endicott criteria (Endicott, 1984) and a specific example of the exclusive approach is the Cavanaugh criteria (Cavanaugh, 1995). Most questionnaires of depressive symptoms use an inclusive approach. Most clinical interviews for diagnosis of major depressive disorder (MDD) use either an etiologic or an inclusive approach.

While these alternative criteria have been proposed to compensate for somatic symptom overlap, few studies have actually empirically examined whether the symptom overlap is problematic for assessing depression in people with cancer. These studies suggest that utilizing substitutive or exclusive criteria leads to lower prevalence of MDD than DSM-IV criteria (Grassi & Rosti, 1996; Ryan, Gallagher, Wright, & Cassidy, 2012; Uchitomi et al., 2001). However, some studies suggest somatic symptoms may still provide useful information especially for screening (Akechi et al., 2009; Mitchell, Lord, & Symonds, 2012; Traeger et al., 2011). Psychometric studies comparing medical populations to healthy controls also support the continued use of somatic items of depression in other medical populations including traumatic brain injury (Cook et al., 2011), spinal cord injury (Bombardier, Richards, Krause, Tulsky, & Tate, 2004), HIV (Perkins et al., 1995) and chronic disease (Simon & Von Korff, 2006). Research from primary care populations shows a high prevalence of somatic symptoms in depression (Simon, Gater, Kisely, & Piccinelli, 1996; Simon, VonKorff, Piccinelli, Fullerton, & Ormel, 1999; Tylee & Gandhi, 2005). Reductions in prevalence with alternative criteria does not conclusively support inflated

rates of depression with standard criteria. The reductions could result from true negatives, in which case alternative criteria would be indicated, or the reductions in prevalence could result from false negatives, in which case alternative criteria would not be indicated. The inconclusive literature on measurement of somatic symptoms of depression in people with cancer suggests that further investigation is required before somatic symptoms are abandoned.

Current Study

The purpose of the current study was to use item response theory (IRT), differential item function (DIF), and other analyses to examine empirically what effect somatic symptoms may have on the measurement of depressive symptoms in people recently diagnosed with cancer. Given the high prevalence of depression and adjustment disorders with depressed mood in people with cancer, appropriate screening and monitoring of both depressive and somatic symptoms is important for supportive cancer care. However, if somatic symptoms do inflate the rate of depression, this would indicate that measures of depression excluding somatic symptoms should be used instead of measures that use these symptoms. In this case, either a substitutive or exclusive approach would be indicated. The current study addressed the following questions:

1. Do somatic items function differently in groups with high somatic symptoms compared to those with low somatic symptoms? Do these items function differently in people who receive chemotherapy versus those who do not? If high somatic symptoms do inflate depressive symptoms, somatic items on a depression measure would not indicate the same depression severity in those with high somatic symptoms compared to those with low somatic symptoms. Somatic items would also be less accurate at measuring depressive symptoms in those with high somatic symptoms compared to those with low somatic symptoms.
2. How crucial are the somatic symptoms to depression assessment and diagnosis? Removing items from a measure changes the psychometric properties of the measure. Before excluding all somatic symptoms from depression assessment, it is important to know how many people with cancer report both somatic and psychological symptoms to inform the importance of the symptoms to screening and monitoring of depression.

Methods

Participants and Procedures

People with incident breast, lung or colorectal cancer (n=251) were recruited to participate in a nurse navigator intervention trial. The trial took place at Group Health Cooperative, an integrated healthcare delivery system in Washington State. Participants were recruited after being diagnosed but before cancer treatment began and were assigned to either usual care or the nurse navigator intervention, which has been described elsewhere (Horner et al., 2013; Wagner et al., 2014). Assessments were conducted at baseline, which generally occurred before cancer treatment, 4-months post-baseline, which typically occurred when most participants were in active treatment (chemotherapy, radiation) or had just ended treatment,

and 12-months post-baseline. The study procedures were approved by the Group Health Institutional Review Board before the study was conducted. All participants provided informed consent before entering the study.

Measures

The Patient Health Questionnaire 9 (PHQ9) is one of the most widely used measures of depression (Kroenke, Spitzer, & Williams, 2001; Spitzer, Kroenke, & Williams, 1999). It consists of 9 items corresponding to the symptoms of a major depressive episode as defined in both the Diagnostic and Statistical Manual (DSM) IV and the DSM-5 (American Psychiatric Association, 2000, 2014). Respondents rate each item on a 0 to 3 scale. Numerous studies have supported a unidimensional factor structure for the PHQ9 (Cameron, Crawford, Lawton, & Reid, 2008; Dum, Pickren, Sobell, & Sobell, 2008; Hansson, Chotai, Nordstöm, & Bodlund, 2009; Kalpakjian et al., 2009) and have shown the scale to be reliable and valid (Kroenke, Spitzer, Williams, & Löwe, 2010). A score of 10 or higher has been shown to be a sensitive and specific screen for a major depressive episode (Kroenke et al., 2001).

To classify participants as having high or low somatic symptoms, we collected pharmacy data from participants' electronic medical records during the first six months that each person was in the study. Collecting this data after study enrollment and diagnosis helped ensure that the medication was likely prescribed for treating side effects of treatment. We collected data on the number of fills for medications commonly used to treat effects of cancer and cancer treatment: anemia, nausea and pain. Medications included Epogen and Procrit for anemia, Zofran for nausea and hydrocodone for pain. These effects are also likely to confound depressive symptom measurement either directly due to the symptom or side effect or indirectly due to treatment for the symptom or side effect. The side effects could potentially be fatigue, sleep disturbance and appetite changes and the medications used to treat these side effects could also lead to temporary cognitive changes. Using prescription fills to measure somatic symptoms is not confounded by similar methodology to the PHQ9 (i.e. self-report questionnaire) and also helps to distinguish mild somatic symptoms from somatic symptoms that necessitated treatment. It also provides a more accurate measure of whether the symptoms required treatment. Participants were classified as high somatic symptoms if they filled any anemia medications or any nausea medications, or had 2 or more pain medication fills. All other participants were classified as low somatic symptoms. Whether participants received chemotherapy for their cancer treatment was also pulled from medical records. Participants were coded into never received chemotherapy for this diagnosis (no chemotherapy) or received chemotherapy for this diagnosis at some point (chemotherapy).

Statistical Analyses

Differential Item Function (DIF)—DIF was used to determine whether participants with high levels of somatic symptoms responded to the PHQ9 questions differently than participants with low levels of somatic symptoms. DIF uses item response theory models (IRT) to determine whether item responses are comparable between the groups. Differences between groups on an item could be due to the actual level of the construct, such as

depression, or it could be due to each group responding differently to the item. If groups respond differently to an item, this means that identical scores on an item do not represent the same level of depression across groups and are not comparable. DIF identifies whether groups responded differently to an item while controlling for the actual level of depression.

One IRT model is the two-parameter logistic (2PL) model (Birnbaum, 1968). This model is appropriate for dichotomous items (yes/no) and models the accuracy of the item (also called the slope) for measuring the construct as well as what level of the construct (also called threshold) the item measures. Threshold is usually defined as the level of the construct at which the probability of endorsing the item is 50%. For example, one item could measure low levels of depression (sad mood) while another could measure very high levels of depression (suicidality). The slope and threshold parameter can be used to estimate the probability of reporting a symptom (endorsing an item) at different levels of depression. DIF uses the 2PL to create different measures of slope and threshold between two groups while controlling for the level of the construct (depressive symptoms). The level of the construct is estimated from the other items, called an anchor. The significance of the differences between the accuracy and threshold parameters can then be tested. If the threshold significantly differs between two groups, this indicates the item is measuring different levels of depression in the two groups. If the slope significantly differs, this indicates that the item measures depression more accurately in one group compared to the other group.

DIF was run on the 4-month assessment because that is when most participants are likely to have side effects from treatment and it would increase the possibility of finding DIF. Two DIF analyses were conducted. In the first analysis, participants were divided into low and high somatic symptom groups based on medication use (anemia, nausea, pain). In the second analysis participants were divided into chemotherapy or no chemotherapy. We then used differential item function between the two groups in each analysis to compare the functioning of the PHQ9 items. Due to low numbers in some response categories ($n<5$), responses 1, 2 and 3 were collapsed for all items and a 2-parameter logistic (2PL) model was used to test for DIF. IRTPRO version 2.0 was used to test for sufficient unidimensionality and lack of local dependence (assumptions of IRT) and was used to test for DIF. For this study, we used the Wald statistic (Lord, 1977) to test for significant DIF on the PHQ9. The Wald test calculates a chi-square value on the separate parameter estimates between the low somatic symptom group and the high somatic symptom group. As we were only interested in whether there was DIF on the somatic items, the psychological symptoms (items 1 [mood], 2 [no interest], 6 [guilt], 8 [psychomotor], and 9 [suicidality]) were the anchor items used to estimate the level of depressive symptoms and the somatic items (items 3 [sleep], 4 [fatigue], 5 [appetite] and 7 [cognition]) were the only items tested for DIF.

Longitudinal Analyses of Removing Somatic Items—At each assessment, we examined how many participants were classified as depressed/not depressed with and without the somatic items using the PHQ9≥10 cutoff. Non-parametric tests (McNemar Test) were used to test for significant change in group (depressed, not depressed) with the somatic item compared to without the item. We also examined just the participants who screened positive (≥10) on the PHQ9. Among this subsample, we examined how many people met the ≥10 point criterion with just somatic symptoms but did not report any psychological

symptoms as well as how many participants screened positive by the full PHQ9 but screen negative when all somatic symptoms are removed.

Results

A total of 251 participants enrolled in the study and completed the baseline assessment. At baseline, 13.5% (n=34) screened as depressed using the ≥10 criterion on the PHQ9. At the 4-month assessment, 242 participants completed the assessment and 14.9% (n=36) screened as depressed on the PHQ9. At the 12-month assessment, 230 participants completed the assessment and 8.7% (n=20) screened as depressed on the PHQ9. See Table 1 for baseline disease and demographic variables. Slightly more than half the sample at 4 months was classified as high somatic symptoms (n=137) and the remainder were classified as low somatic symptoms (n=105) based on pharmacy fills. At the 4-month assessment, the high somatic symptom group reported significantly more depressive symptoms on the full PHQ9 (mean=5.92, SD=4.837) than the low somatic symptoms group (mean=3.52, SD=3.372; p<.001). At the 4-month assessment, significantly more people with high somatic symptoms (22%, n=30 out of 137) screened positive for depression than people with low somatic symptoms (6%, n=6 out of 105; χ^2=12.29, p<.001). Of the total sample (n=251), most received some form of chemotherapy or radiation (73%, n=183), with 16% (n=39) receiving only chemotherapy, 32% (n=80) receiving only radiation and 25% receiving both (n=64).

Differential Item Function: Somatic Symptoms

Using a 2PL model on the whole sample at 4-months (n=242), local independence was supported. No LD statistics were over 10 and only one item (#8, psychosomatic retardation or agitation) had a significant S-X^2 statistic (p=.02; all other ps>.40). This indicated that the unidimensional model fit well overall and we proceeded with the DIF analyses comparing people with low somatic symptoms (n=105) and high somatic symptoms (n=137) as defined by use of medication for symptoms and side effects.

Different slope (accuracy) and threshold (level of depressive symptoms) parameters by somatic symptom group are reported in Table 2 as well as parameters for the five anchor items. Although parameters differed slightly between groups, no significant overall DIF was found on any item (all p's>.14). This means somatic items were not more or less accurate in the high somatic symptom group compared to the low somatic symptom group and the symptoms did not indicate different levels of depression between the two groups. Interestingly, the slope parameters, an indicator of how accurately the item measures depressive symptoms, were not uniformly lower in the high somatic symptom group (see Table 2). Slopes were also sometimes higher for the high somatic symptom group (items #4 and #7) whereas others were higher for the low somatic symptom group (items #3 and #5). The threshold parameters were also not uniformly lower or higher in the high somatic symptom group compared to the low somatic symptom group. Items #3 (sleep disturbance) and #4 (fatigue) indicated the high somatic symptom group needed lower levels of depression to endorse the item than the low somatic symptom group whereas for items #5 (appetite changes) and #7 (cognitive changes) the high somatic symptom group needed higher levels of depression before endorsing the item. This indicates that somatic symptoms

overall are neither a better nor worse measure of depression in the high symptom group compared to the low symptom group.

However, some results supported continued caution when using certain somatic symptoms. Item #3, disturbed sleep, had the lowest slope parameter in both samples and both slope parameters were below 1.0. Ideally, slope parameters are at least 1.0 (Hays, Morales, & Reise, 2000). The items with the two lowest threshold parameters were items #3 and #4, fatigue, indicating that these two somatic items measured the lowest level of depression. While there was no significant DIF, caution should be used when using sleep disturbance and fatigue to measure depressive symptoms in cancer patients.

Differential Item Function: Chemotherapy

Different slope (accuracy) and threshold (level of depressive symptoms) parameters by chemotherapy group are reported in Table 3 as well as parameters for the five anchor items in this model. Similar to results for low and high somatic symptoms, no significant DIF was found on the somatic items between the chemotherapy and no chemotherapy groups (all p's>.07). Consistent with the results for somatic symptom groups, some slope and threshold parameters were higher in the chemotherapy groups while others were higher in the no chemotherapy group. Items #4 (fatigue) and #7 (cognition) had higher slope parameters in the chemotherapy group than the no chemotherapy group, indicating more accurate measurement in the chemotherapy group whereas the slopes for items #3 (sleep disturbance) and #5 (appetite) were higher in the no chemotherapy group. Items #3 (sleep disturbance) and #4 (fatigue) had lower threshold parameters in the chemotherapy group than the no chemotherapy group, indicating that those who received chemotherapy need a lower level of depression before reporting the symptom. Items #5 (appetite) and #7 (cognition) showed the reverse pattern for thresholds. It should be noted that, similar to the previous analyses, these results were not significant. However, fatigue and sleep both had low threshold values, indicating those symptoms are markers of mild, but not moderate or severe, depression.

Longitudinal Analyses

Removing the somatic items affected the number of participants classified as depressed. See Table 4. At baseline, removing one of the somatic items resulted in 4.0 to 4.4% of the sample changing classification from depressed to not depressed. At 4 months, removing a somatic item resulted in 5.0 (cognitive changes) to 8.3% (fatigue) changing groups from depressed to not depressed. At 12 months post-diagnosis, the proportion changing from depressed to not depressed ranged from 2.6 (cognitive changes) to 4.3% (sleep) depending on the item removed. All the changes were significant (all p's<.05). While removing the somatic items individually affected the number of participants screened as depressed, removing all the somatic items had a predictably much larger effect.

Analyses on just the participants who screened positive on the total PHQ9 revealed that removing all the somatic items would, understandably, require recalibration of the PHQ cutoff. Only one participant screened positive by somatic symptoms alone (no psychological symptoms reported) at baseline and no participant screened positive by somatic symptoms alone at 4 months and 12 months post-diagnosis. This essentially indicates that false

positives due to overlapping somatic symptoms are likely rare. However, the majority of people who screened positive on the total PHQ9 would no longer screen positive if all four somatic items were removed and the ≥10 criterion was retained. Only 3 people at baseline, 2 at 4 months and 1 person at 12 months could screen positive on psychological symptoms alone, although all of these participants also reported somatic symptoms. The majority of people screening positive on the total PHQ9 had a combination of somatic and psychological symptoms without clearly screening positive based on one set of symptoms alone (baseline n=30; 4-months n=34; 12-months n=19).

Discussion

This study combined self-report data (PHQ9) with pharmacy information to examine somatic symptoms of depression in cancer. Differential item function (DIF) compared people with high somatic symptoms to those with low somatic symptoms and compared people who received chemotherapy with those who did not. As no significant DIF was found, analyses suggested that somatic items could be used in depressive symptom assessment in cancer but sleep and fatigue only indicated mild depression. Removing any one somatic item only affected screening results for a small portion of the total sample and somatic symptoms did not appear to be a substantial problem by 12-months post-diagnosis. However, the largest effect was for the fatigue item during the 4-month assessment (active cancer treatment) when 8.3% of the sample would not screen positive if the item was removed. Analysis of persons who screened positive on the PHQ9 (score of 10 or higher) showed few if any people screened positive on somatic symptoms alone. More commonly, people screening positive had a combination of somatic and psychological symptoms. Given that the majority of this sample underwent some form of adjuvant treatment, the symptom levels were not surprising. Overall, somatic items performed acceptably across people with high and low somatic symptoms although fatigue and sleep may not be the best indicators of depression during active treatment.

Our findings add to the growing literature on the use of somatic symptoms to assess depression in people with cancer. These results support previous studies that showed removing somatic items reduced the rates of depression in people with cancer (Grassi & Rosti, 1996; Ryan et al., 2012; Uchitomi et al., 2001). However, our results support the continued use of somatic symptoms in the assessment of depression with caution when sleep disturbance or fatigue is reported during active treatment. This is consistent with previous work showing that somatic symptoms may also be useful for screening for MDD when combined with other symptoms (Mitchell et al., 2012). Our results showed the majority of people screening positive on the PHQ9 reported both somatic and psychological symptoms. These results further support the use of somatic items and suggest an exclusive approach should not be used. Additionally, removing all somatic items would require recalibration of a measure, both for a screening cutoff and for determining clinically important differences. Given that the majority of gold-standard criteria for MDD or any depressive disorder include somatic symptoms, a new criterion may also need to be developed. Therefore, excluding all somatic symptoms may be detrimental to depression screening and symptom monitoring in people with cancer, especially given the work suggesting the utility of continued use of somatic symptoms.

While these results inform depressive symptom screening and monitoring in those with cancer, this might not translate to those with other conditions. Several studies have examined other measures of depression and these studies show DIF between women with breast cancer and women with MDD but no cancer diagnosis (Waller, Compas, Hollon, & Beckjord, 2005) but no DIF between women with breast cancer and healthy women from a population-based sample (Osborne, Elsworth, Sprangers, Oort, & Hopper, 2004; Zigmond & Snaith, 1983). As mentioned in the introduction, comparisons of the PHQ9 in healthy controls and other medical populations (rehabilitation, chronic disease) found items did not function differently between medical samples and non-medical, healthy controls. This literature suggests that depression measures may function similarly in healthy adults and people with medical conditions but function differently in those with a psychiatric disorder. When these results are considered with the high number of people with cancer who meet criteria for adjustment disorder but not MDD (Mitchell et al., 2011), it seems the type of distress experienced from cancer diagnosis or other chronic illness likely differs from classic MDD but these measures can still identify clinically significant distress.

Limitations

The contributions of the study should be considered within the limitations. First, this was a very small sample for IRT analysis. Even though the models converged, the results should be considered preliminary. Due to the small sample, we were unable to examine all the response categories and could only conduct DIF analyses on whether the symptom was reported or not. DIF may be present when all response categories are examined. Additionally, due to the small sample size, we were unable to compare people on chemotherapy (active treatment) to people who had stopped chemotherapy. The sample consisted of mostly women with breast cancer so results might not generalize to other types of cancer. Also, this study was unable to address validity of these symptoms compared to a gold-standard, criterion diagnosis. We were only able to pull prescription medication fills and not prescription use or use of over-the-counter medication. In non-cancer populations, the use of over-the-counter pain medications can range from 17% to 28% (Qato et al., 2008; Sugumaran, Cohen, & Kacker, 2012). We also did not compare people with cancer to healthy controls but this was one of the first studies to use DIF to examine depressive symptoms by level of somatic symptoms. Although the timing of the assessments provided a longitudinal look at somatic symptoms of depression, the results should be considered preliminary pending replication.

Conclusions

Based on our study results and the research literature, we conclude that continued use of somatic items in measures of depression is acceptable in people with cancer, although sleep disturbance and fatigue are not the strongest markers of depression. Somatic items functioned similarly in people with low somatic symptoms compared to people with high somatic symptoms and in people who received chemotherapy compared to those who did not. Somatic items also did not show a consistent pattern indicating more or less depression or being more or less accurate in low versus high symptom groups. If the somatic symptoms of cancer and cancer treatment did inflate depression assessment, we would expect consistently lower accuracy and items consistently indicating less depression in those with

high somatic symptoms. Somatic symptoms may also provide important clinical information and, combined with the acceptable functioning of these items in cancer, suggest assessment of these symptoms when screening for or monitoring depression in cancer. However, further research with larger samples is needed to better examine the functioning of somatic items of depression in cancer.

Acknowledgments

The authors would like to thank the participants and the providers who helped with this study. Dr. Jones is supported by a fellowship from the National Institute of Aging (T32 AG027677). This study was supported by a grant from the National Cancer Institute (P20CA137219).

Role of funding source: Funding for the study was provided by the National Cancer Institute, National Institutes of Health, USA. Funding for Dr. Jones was provided by the National Institute on Aging, National Institutes of Health, USA. The funders did not influence the analyses or interpretations in the report nor the decision to write the report or submit it for publication.

References

Akechi T, Ietsugu T, Sukigara M, Okamura H, Nakano T, Akizuki N, Uchitomi Y. Symptom indicator of severity of depression in cancer patients: a comparison of the DSM-IV criteria with alternative diagnostic criteria. Gen Hosp Psychiatry. 2009; 31(3):225–232.10.1016/j.genhosppsych.2008.12.004 [PubMed: 19410101]

American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-IV-TR. 4. Washington, DC: American Psychiatric Association; 2000. text revision

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5. Arlington, VA: American Psychiatric Publishing; 2014.

Birnbaum, A. Some latenet trait models and their use in inferring an examinee's ability. In: Lord, FM.; Novick, MR., editors. Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley; 1968. p. 395-479.

Bombardier CH, Richards JS, Krause JS, Tulsky D, Tate DG. Symptoms of major depression in people with spinal cord injury: implications for screening. Arch Phys Med Rehabil. 2004; 85(11):1749–1756. S0003999304010846 [pii]. [PubMed: 15520969]

Cameron IM, Crawford JR, Lawton K, Reid IC. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. Br J Gen Pract. 2008; 58(546):32–36.10.3399/bjgp08X263794 [PubMed: 18186994]

Cavanaugh S. Depression in the medically ill: Critical issues indiagnostic assessment. Psychosomatics. 1995; 36:48–59. [PubMed: 7871134]

Cook KF, Bombardier CH, Bamer AM, Choi SW, Kroenke K, Fann JR. Do somatic and cognitive symptoms of traumatic brain injury confound depression screening? Arch Phys Med Rehabil. 2011; 92(5):818–823.10.1016/j.apmr.2010.12.008 [PubMed: 21530731]

Dum M, Pickren J, Sobell LC, Sobell MB. Comparing the BDI-II and the PHQ-9 with outpatient substance abusers. Addict Behav. 2008; 33(2):381–387.10.1016/j.addbeh.2007.09.017 [PubMed: 17964079]

Endicott J. Measurement of depression in patients with cancer. Cancer. 1984; 53(10 Suppl):2243–2249. [PubMed: 6704912]

Grassi L, Rosti G. Psychiatric and psychosocial concomitants of abnormal illness behaviour in patients with cancer. Psychother Psychosom. 1996; 65(5):246–252. [PubMed: 8893325]

Hansson M, Chotai J, Nordstöm A, Bodlund O. Comparison of two self-rating scales to detect depression: HADS and PHQ-9. Br J Gen Pract. 2009; 59(566):e283–288.10.3399/bjgp09X454070 [PubMed: 19761655]

Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. Med Care. 2000; 38(9 Suppl):II28–42. [PubMed: 10982088]

Horner K, Ludman EJ, McCorkle R, Canfield E, Flaherty L, Min J, Wagner EH. An oncology nurse navigator program designed to eliminate gaps in early cancer care. Clin J Oncol Nurs. 2013; 17(1): 43–48.10.1188/13.CJON.43-48 [PubMed: 23372095]

Kalpakjian CZ, Toussaint LL, Albright KJ, Bombardier CH, Krause JK, Tate DG. Patient health Questionnaire-9 in spinal cord injury: an examination of factor structure as related to gender. J Spinal Cord Med. 2009; 32(2):147–156. [PubMed: 19569462]

Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001; 16(9):606–613. [PubMed: 11556941]

Kroenke K, Spitzer RL, Williams JB, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen Hosp Psychiatry. 2010; 32(4):345–359.10.1016/j.genhosppsych.2010.03.006 [PubMed: 20633738]

Lord, FM. A study of item bias, using item characteristic curve theory. In: Portinga, YH., editor. Basic problems in cross-cultural psychology. Amsterdam: Swets and Zeitlinger; 1977. p. 19-29.

Mitchell AJ, Chan M, Bhatti H, Halton M, Grassi L, Johansen C, Meader N. Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a meta-analysis of 94 interview-based studies. Lancet Oncology. 2011; 12(2):160–174.10.1016/S1470-2045(11)70002-X

Mitchell AJ, Lord K, Symonds P. Which symptoms are indicative of DSMIV depression in cancer settings? An analysis of the diagnostic significance of somatic and non-somatic symptoms. J Affect Disord. 2012; 138(1–2):137–148.10.1016/j.jad.2011.11.009 [PubMed: 22310033]

Osborne RH, Elsworth GR, Sprangers MA, Oort FJ, Hopper JL. The value of the Hospital Anxiety and Depression Scale (HADS) for comparing women with early onset breast cancer with population-based reference women. Qual Life Res. 2004; 13(1):191–206. [PubMed: 15058800]

Perkins DO, Leserman J, Stern RA, Baum SF, Liao D, Golden RN, Evans DL. Somatic symptoms and HIV infection: relationship to depressive symptoms and indicators of HIV disease. [Research Support, U.S. Gov't, P.H.S.]. Am J Psychiatry. 1995; 152(12):1776–1781. [PubMed: 8526245]

Qato DM, Alexander GC, Conti RM, Johnson M, Schumm P, Lindau ST. Use of prescription and over-the-counter medications and dietary supplements among older adults in the United States. JAMA. 2008; 300(24):2867–2878.10.1001/jama.2008.892 [PubMed: 19109115]

Ryan DA, Gallagher P, Wright S, Cassidy EM. Sensitivity and specificity of the Distress Thermometer and a two-item depression screen (Patient Health Questionnaire-2) with a 'help' question for psychological distress and psychiatric morbidity in patients with advanced cancer. Psychooncology. 2012; 21(12):1275–1284.10.1002/pon.2042 [PubMed: 21919118]

Simon GE, Gater R, Kisely S, Piccinelli M. Somatic symptoms of distress: an international primary care study. Psychosom Med. 1996; 58(5):481–488. [PubMed: 8902899]

Simon GE, Von Korff M. Medical co-morbidity and validity of DSM-IV depression criteria. Psychol Med. 2006; 36(1):27–36.10.1017/S0033291705006136 [PubMed: 16202189]

Simon GE, VonKorff M, Piccinelli M, Fullerton C, Ormel J. An international study of the relation between somatic symptoms and depression. N Engl J Med. 1999; 341(18):1329–1335.10.1056/NEJM199910283411801 [PubMed: 10536124]

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(18):1737–1744. [PubMed: 10568646]

Sugumaran M, Cohen JC, Kacker A. Prevalence of over-the-counter and complementary medication use in the otolaryngology preoperative patient: a patient safety initiative. Laryngoscope. 2012; 122(7):1489–1492.10.1002/lary.23370 [PubMed: 22539239]

Traeger L, Braun IM, Greer JA, Temel JS, Cashavelly B, Pirl WF. Parsing depression from fatigue in patients with cancer using the fatigue symptom inventory. J Pain Symptom Manage. 2011; 42(1): 52–59.10.1016/j.jpainsymman.2010.10.262 [PubMed: 21398085]

Trask PC. Assessment of depression in cancer patients. J Natl Cancer Inst Monogr. 2004; (32):80–92.10.1093/jncimonographs/lgh013 [PubMed: 15263046]

Tylee A, Gandhi P. The importance of somatic symptoms in depression in primary care. Prim Care Companion J Clin Psychiatry. 2005; 7(4):167–176. [PubMed: 16163400]

Uchitomi Y, Kugaya A, Akechi T, Nakano T, Wenner M, Okamura H, Yamawaki S. Three sets of diagnostic criteria for major depression and correlations with serotonin-induced platelet calcium mobilization in cancer patients. Psychopharmacology (Berl). 2001; 153(2):244–248. [PubMed: 11205426]

Wagner EH, Ludman E, Bowles EJ, Penfold R, Rutter C, Chubak J, McCorkle R. Nurse navigators in early cancer care: A randomized, controlled trial. Journal of Clinical Oncology. 2014; 32(1):12–18. [PubMed: 24276777]

Waller NG, Compas BE, Hollon SD, Beckjord E. Measurement of depressive symptoms in women with breast cancer and women with clinical depression: A differential item functioning analysis. Journal of Clinical Psychology in Medical Settings. 2005; 12(2):127–141.

Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983; 67(6):361–370. [PubMed: 6880820]

Table 1

Descriptive statistics by PHQ9 classification as likely depressed or not likely depressed at the baseline assessment.

Variable	Mean (SD) or % (n)	
	Depressed at baseline (n=34)	**Not depressed at baseline (n=217)**
Stage		
0-I	10.9% (14)	89.1% (115)
II	13.5% (10)	86.5% (64)
III	25.7% (9)	74.3% (26)
IV	8.3% (1)	91.7% (11)
Unknown	0.0% (0)	100.0% (1)
Site		
Breast	12.1% (23)	87.9% (167)
Lung	6.7% (2)	93.3% (28)
Colorectal	29.0% (9)	71.0% (22)
Chemotherapy		
Yes	13.6% (14)	86.4% (89)
No	13.5% (20)	86.5% (128)
Radiation		
Yes	9.7% (14)	90.3% (130)
No	18.7% (20)	81.3% (87)
Gender		
Female	13.5% (30)	86.5% (193)
Male	14.3% (4)	85.7% (24)
Marital Status		
Married or partnered	12.1% (21)	87.9% (153)
Not married or partnered	16.9% (13)	83.1% (64)
Charlson		
0	9.8% (16)	90.2% (148)
1+	18.8% (12)	81.2% (52)
Unknown	26.1% (6)	73.9% (17)
Education		
% with some college or less	15.5% (18)	84.5% (98)
% with bachelor’s or higher	11.9% (16)	88.1% (119)
Age at baseline	58.79 (12.49)	62.84 (11.79)
Race/Ethnicity		
Caucasian	11.2% (23)	88.8% (183)
African American	40.0% (4)	60.0% (6)
Asian	26.7% (4)	73.3% (11)
Native American	25.0% (3)	75.0% (9)
Hispanic	0.0% (0)	100.0% (9)
Other	0.0% (0)	100.0% (6)

Table 2

Differential Item Function (DIF) results for low and high somatic symptoms. Item response theory parameters for participants with low and high physical symptoms as measured by side effect medication use.

	Slope	Threshold parameter	Wald X^2 statistic	p-value for overall DIF
Item #1 Mood	3.76	1.28	-	-
Item #2 No interest	1.42	1.64	-	-
Item #3 Sleep				
Low somatic sxs	.99	.11	1.5	.48
High somatic sxs	.64	−.42		
Item #4 Fatigue				
Low somatic sxs	1.06	−.37	3.9	.15
High somatic sxs	1.47	−.90		
Item #5 Appetite				
Low somatic sxs	2.80	.76	1.3	.51
High somatic sxs	1.41	.93		
Item #6 Guilt	1.45	2.17	-	-
Item #7 Cognition				
Low somatic sxs	1.13	1.28	3.7	.16
High somatic sxs	2.18	1.40		
Item #8 Psychomotor	1.30	2.20	-	-
Item #9 Suicidality	1.22	3.76	-	-

Table 3

Differential Item Function (DIF) results for no chemotherapy and received chemotherapy.

	Slope	Threshold parameter	Wald X^2 statistic	p-value for overall DIF
Item #1 Mood	4.24	1.00	-	-
Item #2 No interest	1.71	1.29	-	-
Item #3 Sleep				
No chemotherapy	1.03	−.07	1.0	.59
Chemotherapy	.67	−.62		
Item #4 Fatigue				
No chemotherapy	1.24	−.48	5.1	.08
Chemotherapy	1.75	−1.13		
Item #5 Appetite				
No chemotherapy	2.95	.57	2.2	.33
Chemotherapy	1.43	.73		
Item #6 Guilt	1.78	1.72	-	-
Item #7 Cognition				
No chemotherapy	1.63	.99	.5	.79
Chemotherapy	2.11	1.06		
Item #8 Psychomotor	1.55	1.76	-	-
Item #9 Suicidality	1.64	2.88	-	-

Table 4

Proportion of participants with probable major depression with physical symptoms removed. The percent depressed by full PHQ9 was 13.5% (n=34) at baseline, 14.9% (n=36) at four months and 8.7% (n=20) and is the comparison used for the significance tests.

	% (n) classified as depressed with item but not depressed without item	% (n) still depressed by PHQ9 with item removed	P-value testing standard PHQ9 scoring with scoring with item removed
Baseline, PHQ9>=10			
Sleep item removed	4.4% (11)	9.2% (23)	.001
Fatigue item removed	4.4% (11)	9.2% (23)	.001
Appetite item removed	4.0% (10)	9.6% (24)	.002
Cognition item removed	4.0% (10)	9.6% (24)	.002
4-month, PHQ9>=10			
Sleep item removed	7.0% (17)	7.9% (19)	<.001
Fatigue item removed	8.3% (20)	6.6% (16)	<.001
Appetite item removed	7.0% (17)	7.9% (19)	<.001
Cognition item removed	5.0% (12)	9.9% (24)	<.001
12-month, PHQ9>=10			
Sleep item removed	4.3% (10)	4.3% (10)	.002
Fatigue item removed	3.9% (9)	4.8% (11)	.004
Appetite item removed	3.0% (7)	5.7% (13)	.016
Cognition item removed	2.6% (6)	6.1% (14)	.031