Published in final edited form as:

Psychooncology. 2008 June ; 17(6): 556–560. doi:10.1002/pon.1289.

Sensitivity and specificity of the Distress Thermometer for depression in newly diagnosed breast cancer patients

Mark T. Hegel^{1,*}, E. Dale Collins², Stephen Kearing¹, Karen L. Gillock¹, Caroline P. Moore², and Tim A. Ahles³

¹Dartmouth Medical School, Hanover, NH, USA

²Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA

³Memorial Sloan-Kettering Cancer Center, New York, NY, USA

Abstract

Background—Receiving a new diagnosis of breast cancer is a distressing experience that may precipitate an episode of major depressive disorder. Efficient screening methods for detecting depression in the oncology setting are needed. This study evaluated the receiver operating characteristics (ROC) of the single-item Distress Thermometer (DT) for detecting depression in women newly diagnosed with Stage I–III breast cancer.

Methods—We assessed 321 patients (of 345 consecutive patients) at the time of their pre-surgical consultation at a Comprehensive Breast Cancer Program. Patients were administered the DT along with the Patient Health Questionnaire 9-Item Depression Module (PHQ-9) as a gold standard diagnostic assessment of depression status.

Results—Mean DT scores (11-point scale, 0–10) were significantly higher for depressed versus non-depressed patients (8.1 versus 4.4). In ROC analyses the DT showed strong discriminatory power relative to the PHQ-9-derived diagnosis of depression, with an area under the curve of 0.87. Patient age, education, marital status and stage of disease resulted in similar operating characteristics. A score of 7 represented the optimal trade-off between sensitivity (0.81) and specificity (0.85) characteristics for detecting depression.

Conclusions—The single-item DT performs satisfactorily relative to the PHQ-9 for detecting depression in newly diagnosed breast cancer patients. A cutoff score of 7 on the DT possesses the optimal sensitivity and specificity characteristics. The strength of these findings suggests that a careful psychosocial evaluation should follow a positive screen.

Keywords

medical oncology; breast neoplasms; depression; sensitivity and specificity; psychological tests

Introduction

Receiving a diagnosis of breast cancer, contemplating the uncertainty of the future, and anticipating the daunting process of choosing and undergoing treatments can induce significant distress. For a significant percentage of new breast cancer patients, initial distress

can be extreme and persistent, often presaging a psychiatric disorder, such as major depression [1].

About one-third of breast cancer patients experience significant emotional distress and functional impairment [2–6]. This is similar to prevalence rates in other cancer patient populations [7–11]. Of the breast cancer patients surveyed by Coyne *et al.* [3], 29% evidenced some level of distress, with 9% meeting the criteria for major depression. Recently, Hegel *et al.* documented a clinically elevated distress level in 47% of newly diagnosed breast cancer patients, with 11% screening positive for depression [1]. Depression often tracks with and complicates other cancer-related symptoms, such as pain and fatigue [12]; therefore, it is crucial to identify patients early in their cancer treatment and initiate early intervention. A longitudinal assessment of breast cancer patients' psychological adjustment revealed that 26% continued to meet the criteria for a depressive illness at six months post-treatment [4].

The NCCN [7,13] and the American Psychosocial Oncology Society [14] have both issued consensus panel statements in which they recommend the development of improved screening mechanisms to enhance the detection and management of emotional distress in cancer patients. The guidelines recommend that screening tools be brief, easy to administer, non-stigmatizing and easy to score and interpret. The Distress Thermometer (DT), piloted at Memorial Sloan-Kettering Cancer Center by Roth and colleagues [11], meets these criteria and has been endorsed by the NCCN Distress Practice Guidelines panel [13].

The DT has been examined for its performance against other more lengthy measures of general distress [15,16]. Gil and colleagues assessed a mixed sample of European cancer patients at various disease stages and found an optimal DT cutoff of 4 (sensitivity = 0.65, specificity = 0.79) [17] relative to the Hospital Anxiety and Depression Scale (HADS) total score [18]. Jacobsen *et al.* studied a mixed sample of long-term cancer survivors and found a DT cutoff of 4 to have 0.77 sensitivity and 0.68 specificity relative to the HADS total score and 0.70 sensitivity and 0.70 specificity relative to the Brief Symptom Inventory short form (BSI-18) [19]. In a study of a mixed sample of long-term cancer survivors, Hoffman and colleagues [20] found that no specific DT score optimized sensitivity or specificity relative to the full version of the BSI [21] or the BSI-18 [22].

To our knowledge, the sensitivity and specificity of the DT for specifically detecting clinical depression have been examined in only one report. Akizuki *et al.* compared the DT to an unstructured psychiatric interview in a mixed sample of Japanese cancer patients. They found a cutoff of 4/5 for optimal sensitivity (0.84) and specificity (0.61) relative to the presence or absence of adjustment disorder or major depressive disorder (MDD) [23].

The DT is quickly finding its way into routine use in oncology clinics, and it would therefore be useful to further develop its utility for multiple screening purposes. In this paper we report the sensitivity and specificity of the DT, using receiver operating characteristics (ROC) curve analyses for depression screening in newly diagnosed breast cancer patients. The Patient Health Questionnaire depression module (PHQ-9) [24], a validated self-report instrument for diagnosing depression, was used as the gold standard for the analyses. The PHQ-9 was chosen as the depression measure because its items correspond directly to Diagnostic and Statistical Manual of Mental Disorder-version IV (DSM-IV) criteria for depression [25], thus allowing it to yield a diagnosis.

Methods

Study participants

Participants were female patients from Northern New England consecutively attending a comprehensive breast cancer clinic at the Norris Cotton Cancer Center of Dartmouth-Hitchcock Medical Center. All potential participants ($N=345$) were newly diagnosed after having undergone a biopsy resulting in a positive diagnosis of breast cancer and were judged eligible for either lumpectomy or mastectomy. Definitive staging was not completed at the time of the assessment. Subsequent chart review showed six patients were Stage 0, six were Stage IV and seven were 'unknown' due to their surgery being performed at another hospital. These 19 patients were excluded from the analyses. Patients were eligible to complete the assessments if they could read, write and speak English, and were capable of making health-care decisions for themselves. Only one woman was ineligible on these grounds, and four women failed to complete the intake due to scheduling problems, leaving a sample of 321 participants with Stages I–III breast cancer. The Dartmouth College Institutional Review Board approved this study and deemed informed consent as unnecessary due to the use of de-identified health information.

Procedure

Between April 2004 and December 2006, patients with a positive biopsy for breast cancer completed a self-report assessment of psychosocial status (part of a larger assessment of medical and treatment decision-making variables) as standard clinic procedure prior to their initial consultation with a surgical oncologist. Measures were administered via a touch pad interface, which requires responses to all items, eliminating missing data. The DT was always completed prior to the PHQ-9, which was sequentially separated from the DT by several questionnaires regarding general health information.

Measures

DT—The DT is a one-item self-report screening tool for measuring psychological distress in cancer patients [26]. The DT measures distress levels over the past week using a thermometer-like Likert scale with scores from 0 (no distress) to 10 (extreme distress) and a midpoint anchor labeled 'moderate distress' [7].

Patient Health Questionnaire 9-Item Depression Module (PHQ-9)—The PHQ-9 is a depression subscale derived from the Patient Health Questionnaire [24]. The PHQ-9 is a self-administered instrument that measures both the presence and severity of the nine symptoms of major depression as defined by the DSM-IV [25]. Responses for each question range from 'not at all' (scored as 0) to 'nearly every day' (scored as 3), with a highest possible score of 27.

The PHQ-9 is capable of yielding a preliminary diagnosis of MDD [27]. As with the DSM-IV criteria, for a diagnosis of depression patients must endorse at least one of the first two items (anhedonia and/or depressed mood) and three to four remaining items as being present for 'more than half the days' during the past two weeks. The cutoff score used in the current study ($\text{PHQ-9} > 10$) has been shown to have a diagnostic specificity of 89–94% and sensitivity of 73–91% for major depression in medical patients compared with a structured diagnostic interview [27,28]. The PHQ-9 has also shown strong convergent validity with other established measures of depression and function and strong construct validity for the general population [24,25,27–29].

Analysis

The PHQ-9 was used as the diagnostic tool for depression. ROC analyses were used to assess the discriminatory power of the DT as a screening tool for depression. We compared effects of age, education, marital status and stage of disease on the DT operating characteristics. Sensitivities, specificities and likelihood ratios were calculated at each distress cutoff for identifying patients with depression (PHQ-9 screening score >10). Positive and negative predictive values were calculated using our sample prevalence for major depression. Demographic and clinical characteristics were compared for participants meeting the indicated cutoff on the DT using the Pearson χ^2 test. SAS version 9.1 (SAS Institute, Cary, NC) was used for all analyses.

Results

Demographic and clinical characteristics

The population ($n = 321$) was female (100%), primarily white, non-Hispanic (97%) and well educated (69% at least some college) (Table 1).

DT score and depression

The mean DT score at intake for all patients ($n = 321$) was 4.7 (SD = 2.6), with a range of 0.0–10.0. Scores were skewed toward the upper half of the DT for depressed patients, while non-depressed patients showed a more normal distribution (Figure 1). Mean DT scores were significantly higher ($P < 0.001$) for depressed (8.1, SD = 1.5) than for non-depressed patients (4.4, SD = 2.4).

Of the 321 patients completing the assessment, 32 (10%) screened positive for depression with a PHQ-9 score >10. The DT showed strong discriminatory power relative to the PHQ-9 for patients with depression, with an area under the curve of 0.86 (Figure 2). Operating characteristics were similar among older (0.87) versus younger (0.87) patients, college (0.85) versus high school educated (0.94) and married (0.86) versus unmarried (0.89) and stage of disease (0.84–0.98).

Table 2 demonstrates the association of DT scores and screening positive for depression using the PHQ-9. For this cohort, a score of 7 represents the optimal trade-off between sensitivity (0.81) and specificity (0.85) characteristics for depression. The corresponding positive likelihood ratio was 5.3 (i.e. a score of 7 on the DT is 5.3 times as likely to occur in a patient scoring >10 on the PHQ9 as opposed to a patient screening negative for depression using the PHQ9). Using a cutoff of DT = 7, no differences were observed for age ($P = 0.14$), education ($P = 0.22$), marital status ($P = 0.14$) or stage of disease ($P = 0.53$).

Discussion

Our prevalence of 10% for depression based on the PHQ-9 screening instrument falls within the reported range for depression in cancer patients in general [3,23,30]. For these patients, the DT performed well relative to the PHQ-9 for predicting depression and did not differ by age, education, marital status or stage of disease. A DT cutoff of 7 optimized sensitivity and specificity for detecting depression.

Akizuki *et al.* [23] reported a lower DT cutoff of 4/5 for detecting depression. However, their study was notably different from the current study in (1) using a sub-sample of patients already referred for psychiatric care prior to screening; (2) conducting a combined analysis with major depression and adjustment disorder (a less severe disorder); and (3) not utilizing a standardized instrument for psychiatric diagnosis (i.e. diagnosis was derived from a non-

structured clinical interview without reliability checks). Therefore, their screening performance for the DT may have been inflated due to the potential oversampling of depressed patients, including a less severe diagnostic category, and the lack of a validated diagnostic procedure.

Following a positive screen for depression on the DT, clinicians may wish to follow up with a more elaborate depression screen, such as using the PHQ-9. In our setting patients with high DT scores in the absence of depression are followed by the social work care managers, whereas those subsequently screening positive for depression on the PHQ-9 are offered referral to specialty mental health.

Limitations

There are limitations to the conclusions that can be reached from this patient cohort. Most notably, cross-cultural comparisons to other breast cancer patients may be limited due to the homogeneous make-up (mostly white, non-Hispanic, well-educated women from Vermont and New Hampshire) of the study population. Second, the DT and PHQ-9 were administered concurrently, although not consecutively, which may contribute nominally to the observed agreement between the scales [31].

Conclusions

The single-item DT performs satisfactorily, with an optimal cut-off of 7 (sensitivity = 0.81, specificity = 0.85, LR = 5.3) relative to the PHQ-9 for detecting depression in newly diagnosed breast cancer patients. With increased time pressures and shorter clinical appointments now the norm, brief screening instruments such as the DT, followed by a comprehensive assessment or referral protocol for those who screen positive, are necessary to adequately detect and manage depression for this vulnerable population. Future research should evaluate the sensitivity of the DT to change as a function of treatment that is provided in the oncology setting or via referral to specialty mental health care.

Acknowledgments

Preparation of this paper was supported in part by the Behavioral Science Shared Resource of the Norris Cotton Cancer Center core grant P30CA23108, National Cancer Institute, Bethesda, MD, for M. T. H. and T. A. A., and by the Foundation for Informed Medical Decision Making for E. D. C., S. K. and C. P. M.

Informed consent: The Dartmouth College Institutional Review Board approved this study and deemed informed consent as unnecessary due to use of de-identified health information.

References

1. Hegel MT, Moore CP, Collins ED, et al. Distress, psychiatric syndromes and functional impairment in women with newly diagnosed breast cancer. *Cancer*. 2006; 107(12):2924–2931. [PubMed: 17103381]
2. Andrykowski MA, Carpenter JS, Studts JL, et al. Psychological impact of benign breast biopsy: a longitudinal, comparative study. *Health Psychol*. 2002; 21(5):485–494.
3. Coyne JC, Palmer SC, Shapiro PJ, Thompson R, DeMichele A. Distress, psychiatric morbidity, and prescriptions for psychotropic medication in a breast cancer waiting room sample. *Gen Hosp Psychiatry*. 2004; 26:121–128. [PubMed: 15038929]
4. Epping-Jordan JE, Compas BE, Osowiecki DM, et al. Psychological adjustment in breast cancer: processes of emotional distress. *Health Psychol*. 1999; 18(4):315–326. [PubMed: 10431932]
5. Kornblith AB, Herndon JE II, Weiss RB, et al. Long-term adjustment of survivors of early stage breast carcinoma, 20 years after adjuvant chemotherapy. *Cancer*. 2003; 98(4):679–689. [PubMed: 12910510]

6. Luecken LJ, Compas BE. Stress, coping, and immune function in breast cancer. *Ann Behav Med*. 2002; 24(4):336–344. [PubMed: 12434945]
7. Holland JC. Update: NCCN practice guidelines for the management of psychosocial distress. *Oncology*. 1999; 13(11A):459–507.
8. Kangas M, Henry JL, Bryant RA. Posttraumatic stress disorder following cancer: a conceptual and empirical review. *Clin Psychol Rev*. 2002; 22:499–524.
9. Aapro M, Cull A. Depression in breast cancer patients: the need for treatment. *Ann Oncol*. 1999; 10:627–636. [PubMed: 10442183]
10. Nezu, AM.; Nezu, CM.; Felgoise, SH.; Zwick, ML. Psychosocial oncology. In: Nezu, AM.; Nezu, CM.; Geller, PA., editors. *Health Psychology*. Vol. 9. Wiley; New York: 2003. p. 267–292. *The Handbook of Psychology*
11. Roth AJ, Kornblith AB, Batel-Copel L, Peabody E, Scher HI, Holland JC. Rapid screening for psychologic distress in men with prostate carcinoma. *Cancer*. 1998; 82(10):1904–1908.
12. Fleishman SB. Treatment of symptom clusters: pain, depression, and fatigue. *J Natl Cancer Inst (Monographs)*. 2004; 32:119–123.
13. Holland JC, Jacobsen PB, Riba MB. NCCN distress management. *Cancer Control*. 2001; 8(6 Suppl 2):88–93.
14. APOS. APOS Standards of Care for the Management of Distress in Patients with Cancer. APOS; 2000.
15. Murphy JM, Berwick DM, Weinstein MC, Borus JF, Budman SH, Klerman GL. Performance of screening and diagnostic tests. Application of receiver operating characteristic analysis. *Arch Gen Psychiatry*. 1987; 44:550–555. [PubMed: 3579501]
16. Zweig MH, Campbell G. Receiver-operating characteristics (ROC) plots: a fundamental evaluation in clinical medicine. *Clin Chem*. 1993; 39:561–577.
17. Gil F, Grassi L, Travado L, Tomamichel M, Gonzalez JR. Use of distress and depression thermometers to measure psychosocial morbidity among southern European cancer patients. *Support Care Cancer*. 2005; 13:600–606.
18. Zigmond A, Snaith R. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand*. 1983; 67:361–370. [PubMed: 6880820]
19. Jacobsen PB, Donovan KA, Trask PC, et al. Screening for psychologic distress in ambulatory cancer patients: a multicenter evaluation of the distress thermometer. *Cancer*. 2005; 103:1494–1502.
20. Hoffman BM, Zevon MA, D'Arrigo MC, Cecchini TB. Screening for distress in cancer patients: the NCCN rapid-screening measure. *Psycho-Oncology*. 2004; 13:792–799.
21. Derogatis, LR. *The Brief Symptom Inventory Administration, Scoring, and Procedures Manual*. Vol. II. Clinical Psychometric Research, Inc; New York: 1992.
22. Derogatis, LR. *Brief Symptom Inventory (BSI) 18 Administration, Scoring, and Procedures Manual*. National Computer Systems, Inc; Minneapolis, MS: 2000.
23. Akizuki N, Akechi T, Nakanishi T, et al. Development of a brief screening interview for adjustment disorders and major depression in patients with cancer. *Cancer*. 2003; 97(10):2605–2613.
24. Spitzer RL, Kroenke K, Williams JBW. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *J Am Med Assoc*. 1999; 282(18):1737–1744.
25. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4. American Psychiatric Association; Washington, DC: 1994.
26. Perry, MC., editor. *An Algorithm for Rapid Assessment and Referral of Distressed Patients*. American Society of Clinical Oncology; Alexandria, VA: 2000.
27. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001; 16:606–613. [PubMed: 11556941]
28. Spitzer R, Williams J, Kroenke K, et al. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 Study. *J Am Med Assoc*. 1994; 14(272):1749–1756.

29. Martin A, Fief W, Klaiberg A, Braehler E. Validity of the brief patient health questionnaire mood scale (PHQ-9) in the general population. *Gen Hosp Psychiatry*. 2006; 28:71–77. [PubMed: 16377369]
30. Ganz PA, Kwan L, Stanton AL, et al. Quality of life at the end of primary treatment of breast cancer: first results from the moving beyond cancer randomized trial. *J Natl Cancer Inst*. 2004; 96(5):376–387.
31. Rost K, Burnam MA, Smith GR. Development of screeners for depressive disorders and substance disorder history. *Med Care*. 1993; 31(3):189–200.

Figure 1. Distribution (%) of distress thermometer scores by depression screening outcome; PHQ9 negative (score ≤ 10)/PHQ9 positive (score >10)

Figure 2.
Receiver operating characteristic curve for distress thermometer versus PHQ-9 scores for major depression diagnosis

Table 1

Patient characteristics

	<i>n</i>	(%)
<i>Demographic characteristics</i>		
Age (mean $\pm$ SD, range) 57.8 $\pm$ 12.6 years (25–88)		
Female	321	(100)
Race ethnicity		
White, non-Hispanic	310	(97)
Other	11	(3)
Education		
High school or less	100	(31)
Some college	96	(30)
College graduate	125	(39)
Marital status		
Married	210	(65)
Unmarried	111	(35)
<i>Clinical characteristics</i>		
Stage I	150	(47)
Stage II	130	(40)
Stage III	41	(13)

Sensitivity, specificity, likelihood ratio (LR), positive predictive value (PPV) and negative predictive value (NPV) of the Distress Thermometer as a screening tool for depression

Table 2

Distress cutoff	Sensitivity (%)	Specificity (%)	LR (positive)	PPV (%)	NPV (%)
10	0	100			
9	25	98	12.0	57	92
8	50	94	8.0	47	94
7	81	85	5.3	37	98
6	84	73	3.2	26	98
5	94	64	2.6	22	99
4	100	45	1.8	17	100
3	100	37	1.6	15	100
2	100	24	1.3	13	100
1	100	9	1.1	11	100
0	100	2	1.0	10	100