VALIDATION OF THE DISTRESS THERMOMETER WITH BONE MARROW TRANSPLANT PATIENTS

SEAN RANSOM, PAUL B. JACOBSEN* and MARGARET BOOTH-JONES
University of South Florida, USA

SUMMARY

The Distress Thermometer (DT) is a one-item screening measure of psychological distress in cancer patients. This study examines the operating characteristics of the DT in patients about to undergo bone marrow transplant (BMT). Patients (N = 491) completed the DT, the Center for Epidemiological Studies-Depression Scale (CES-D), the State-Trait Anxiety Inventory-State Version (STAI-S), the ECOG Performance Status Scale, and the Patient Problem List. DT scores were related to higher depression, higher anxiety, and poorer performance status. Receiver operating characteristic (ROC) curve analyses of DT scores yielded area under the curve estimates of 0.75 when compared to the CES-D cutoff score of 16, suggesting the DT has acceptable overall accuracy. The DT cutoff score of 4 had the greatest sensitivity and specificity when compared to the CES-D cutoff score. Patients above this cutoff score reported worse ECOG scores and more practical, family, emotional, and physical problems (all \( p's < 0.05 \)) than those below the cutoff. The findings suggest the DT is a useful tool for screening for distress in BMT patients. The optimal DT cutoff score of 4 found here was identical to that found in another study using different criteria. This cutoff score also identified patients with problems likely to reflect psychological distress. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: bone marrow transplantation; psychological distress; cancer; depression; anxiety; screening; oncology

INTRODUCTION

Health care professionals often fail to recognize cancer patients’ psychological distress (e.g. Passik et al., 1998; McDonald et al., 1999; Sollner et al., 2001). This under-recognition of patient distress is unfortunate since such distress can impact the patient’s quality of life (Skarstein et al., 2000) and prevents them from receiving treatment for their distress (Newell et al., 2002). Distressed patients are often less treatment adherent (Kennard et al., 2004) and tend to be unhappier with their quality of care (Von Essen et al., 2002).

Because of this, National Comprehensive Cancer Network guidelines recommend that clinicians routinely screen cancer patients for distress (National Comprehensive Cancer Network, 2003). A number of well-validated measures exist that can be used as distress screeners, including the Hospital Anxiety and Depression Scale (HADS; Zigmond and Snaith, 1983), the 18-item version of the Brief Symptom Inventory (BSI-18; Derogatis, 2000), and the Center for Epidemiological Studies-Depression Scale (CES-D; Radloff, 1977), but these measures require time commitments that can limit their use in busy medical clinics.

The recently developed distress thermometer (DT; National Comprehensive Cancer Network, 2003; Roth et al., 1998) addresses the need for a quick, easily understood measure of distress in medical populations. As shown in Figure 1, the measure consists of a single item that asks patients to rate their distress in the past week on a 0 (‘No distress’) to 10 (‘Extreme distress’) scale.

Research suggests the DT fares well when compared to longer distress-related instruments. In a recent multi-center validation study, Jacobsen

*Correspondence to: Psychosocial and Palliative Care Program, Moffitt Cancer Center, 12902 Magnolia Drive, MRC-PSY, Tampa, FL 33612, USA. E-mail: jacobsen@moffitt.usf.edu

Copyright © 2005 John Wiley & Sons, Ltd.
et al. (2005) compared the DT with the HADS and the BSI-18 in a sample of patients with mixed cancer diagnoses. Using the HADS and BSI-18 cutoff scores as criteria of clinically significant distress, these researchers calculated receiver operating characteristics (ROC) curves to determine which cutoff score on the DT was most effective in distinguishing distressed patients from those who were not distressed. Findings indicated that a DT cutoff score of 4 was the most sensitive (0.77) and specific (0.68) in identifying patients who met or surpassed the HADS cutoff score of 14. When the BSI-18 cutoff scores of 10 (for males) and 14 (for females) were used as a criterion, the optimal DT
cutoff score was also 4. Again, this score was acceptably sensitive (0.70) and specific (0.70). The DT's area under the curve (AUC), a representation of the measure's accuracy in discriminating distressed patients from non-distressed ones, was 0.80 when the HADS cutoff score was used as a criterion and 0.77 when the BSI-18 cutoff was used, both indicating good accuracy.

At least four other studies have also examined the DT. In the article that introduced the DT, Roth et al. (1998) adopted a DT cutoff score of 5 and reported the concordance of this score with the established HADS cutoff score of 15 in 93 men with prostate cancer. The researchers found a concordance rate of 74.4% between the two measures, with 28.6% of patients meeting the DT cutoff and 13% of patients meeting the HADS cutoff. The ability of the DT to sensitively and specifically distinguish those patients who surpassed the HADS cutoff score was not reported.

Using the same DT cutoff score of 5, Trask et al. (2002) compared the DT with the Anxiety and Depression Scales of the HADS in a sample of 50 candidates for bone marrow transplantation (BMT). These researchers found that 50% of the patients met the cutoff of 5 on the DT, 51% of patients met the established cutoff of 8 for the HADS anxiety subscale, and less than 20% met the cutoff of 8 for the depression subscale. No data on the concordance of the measures was presented.

The remaining two studies used ROC curve analysis to examine the DT. In a sample of 275 people with mixed types of cancers, Akizuki et al. (2003) used a Japanese version of the DT to test its ability to determine which patients had been diagnosed with adjustment disorder or major depressive disorder based on a psychiatric interview. ROC analyses indicated that a DT cutoff of 5 could optimally detect patients who had received either of the two diagnoses. This score had a sensitivity of 0.81 and a specificity of 0.61. However, the reliability of the psychiatric diagnoses is unclear as the interviews were apparently non-standardized and no reliability information on the diagnoses was presented.

Finally, Hoffman et al. (2004) used ROC curve analysis to determine whether the DT could distinguish patients who met the cutoff scores of the BSI and the BSI-18. Using a sample of 68 cancer patients with mixed diagnoses, the researchers were unable to reach a conclusion as to which DT cutoff score was most effective in detecting patients who met the criteria. They reported that the traditional cutoff score of 5 yielded a sensitivity of 0.59 and a specificity of 0.71 when using BSI cutoff scores, and a sensitivity of 0.70 and a specificity of 0.64 when using the BSI-18 cutoff scores.

This study seeks to further explore and clarify the psychometric characteristics of the DT by examining the DT's relationship with the Center for Epidemiological Studies-Depression Scale (CES-D). Using responses from a large sample (N = 491) of BMT candidates, ROC curve analysis will be used to determine the DT's ability to distinguish those with clinically significant distress from those without based on the previously established CES-D cutoff score of 16. In addition, we will also note the DT's relationships with measures of anxiety, performance status, and problems experienced by the patient.

**METHODS**

**Participants**

Participants were patients undergoing evaluation for possible BMT at the H. Lee Moffitt Cancer Center (HLMCC) in Tampa, Florida who: (1) were between the ages of 18 and 75 years old; (2) had no untreated or unstable major medical conditions; (3) had no major psychiatric or neurological disorders that could potentially invalidate assessment; (4) completed the 8th grade or its equivalent; (5) could speak and read standard English; and (6) intended to return to HLMCC for post-transplant follow-up appointments.

**Procedure**

BMT candidates were introduced to the study following the completion of the standard psychosocial evaluation required by HLMCC prior to transplantation. After providing informed consent, participants were provided a questionnaire packet that included the DT, the Patient Problem List, the state version of the State-Trait Anxiety Inventory (STAI-S), the Center for Epidemiological Studies-Depression Scale (CES-D), and questions assessing demographic and clinical information. Participants were given the choice of completing the measures on site or returning the
measures by mail. The measures were included in a larger packet of questionnaires intended for a long-term study of psychosocial and neuropsychological outcomes of BMT.

**Measures**

**Demographic and clinical information.** A standard demographic form was used to collect self-report data on age, gender, race/ethnicity, marital status, education, and annual household income. Performance status was assessed using a patient self-report version of the ECOG Performance Status Scale (Oken et al., 1982), and the patient’s cancer type was obtained via medical chart review.

**Distress thermometer (DT).** The DT (National Comprehensive Cancer Network, 2003) is a single-item, self-report measure of psychological distress (see Figure 1). The standard research version of the DT consists of an 11-point scale with endpoints labeled ‘No distress’ (0) and ‘Extreme distress’ (10). Participants are instructed to circle the number that describes their level of distress in the past seven days. Previous psychometric evaluation of the DT indicated that a cutoff point of four and above yielded the best sensitivity and specificity in distinguishing between individuals who were distressed and those who were not, as judged by the established cutoff scores on the HADS and BSI-18 (Jacobsen et al., 2005).

**Patient Problem List.** The Patient Problem List was developed by the Distress Management Guidelines Panel of the National Comprehensive Cancer Network (National Comprehensive Cancer Network, 2003). The list comprises problems commonly experienced by cancer patients. The problems are grouped into five categories—practical problems, family problems, emotional problems, spiritual/religious concerns, and physical problems. Respondents indicate whether or not (yes/no) they have experienced any of the problems in the past week. The version used in the current study consisted of 33 problems.

**Center for Epidemiological Studies-Depression Scale (CES-D).** The CES-D (Radloff, 1977) is a 20-item self-report measure of depressive symptomatology that has been well validated and is widely used. Participants rate the degree to which they experienced depressive symptoms during the past week on a 0 (‘Rarely or none of the time’) to 3 (‘Most or all of the time’) scale. Previous research indicates that a cutoff score of 16 best distinguishes people with clinical levels of depressive symptoms from those without (Myers and Weissman, 1980; Radloff, 1977).

**State-Trait Anxiety Inventory-State Version (STAI-S).** The STAI-S (Spielberger et al., 1983) is a 20-item measure that assesses current anxious symptoms (i.e. worry and nervousness) on a 0 (‘Not at all’) to 3 (‘Very much so’) scale. The STAI-S has been widely used, including in samples of BMT patients (e.g. Rodrigue et al., 1993).

**RESULTS**

**Demographic and clinical characteristics**

A total of 491 patients participated in the study. As shown in Table 1, the sample averaged approximately 50 years of age and had slightly more males (54%) than females (46%). Most participants were White (82%) and the majority was married (67%). The average level of education approached 14 years. A variety of cancer diagnoses were represented in the sample, with the greatest number of patients diagnosed with multiple myeloma (43%), followed by non-Hodgkins lymphoma (18%), and acute myeloid leukemia (11%). Most patients (68%) had a self-reported ECOG Performance Status of 0 (‘Fully active’) or 1 (‘Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature’). The majority of patients (76%) were being evaluated for autologous transplant.

**Relationships among DT, CES-D, and STAI-S scores**

Patients averaged 10.9 on the CES-D (S.D. = 8.5), 39.4 on the STAI-S (S.D. = 9.6), and 3.4 on the DT (S.D. = 2.7). Using the CES-D cutoff score of 16 as the criterion, 24% of participants exhibited clinically significant depressive symptoms.
The DT was significantly correlated with the CES-D \((r = 0.59, \ p < 0.0001)\) and the STAI-S \((r = 0.58, \ p < 0.0001)\), with results indicating that the DT shares substantial variance with each of the other two measures.

### Relationships of DT, STAI-S, and CES-D scores to demographic and clinical characteristics

To determine whether patients differed in their responses on the CES-D, STAI-S, and DT based on demographic and clinical characteristics, correlational analyses were conducted (see Table 2). With regard to demographic characteristics, these analyses found that females scored significantly higher than males on the CES-D and the STAI-S. Non-Whites scored significantly higher than Whites on the CES-D. Correlational analyses showed that years of education were significantly negatively associated with scores on the CES-D and the STAI-S. None of the demographic variables assessed were associated with responses on the DT. With regard to clinical characteristics, participants in the sample’s largest diagnostic category (multiple myeloma) were compared with participants who did not share this diagnosis. These patients did not differ significantly from other patients in the study on the CES-D, STAI-S, or DT. The procedure was repeated for participants in the second largest diagnostic category (non-Hodgkins lymphoma) and, again, no significant relationships were found. Bone marrow transplant type (autologous vs allogeneic/MUD) was also unrelated to variables of interest. Lastly, correlational analyses indicated that poorer ECOG Performance Status was significantly related to higher scores on the CES-D, STAI-S, and DT.

### Establishment of a distress thermometer cutoff score

The frequency distribution of DT scores is reported in Table 3. ROC analysis was used to determine whether scores on the DT could validly distinguish whether a patient met criteria for clinically significant depression as measured by the CES-D cutoff score of 16.\(^1\) To do this, sensitivity and 1-specificity of each score in the DT’s range was calculated and used to determine how well the score distinguished individuals who surpassed the CES-D cutoff score from those who did not. The ROC curve graphically represents the sensitivity and specificity coefficients that would be generated using each possible cutoff score in the DT’s range, with the accuracy of the cutoff score determined by calculating the AUC. Calculating AUC generates scores from 1.0 (perfect accuracy) to 0.5 (accuracy no better than chance).

Examination of the DT’s ROC curve indicated an AUC of 0.75. A cutoff score of 4 yielded the optimal ratio of sensitivity (0.80) to specificity (0.70) when the CES-D cutoff score of 16 is used as a criterion (see Figure 2).
The frequencies of true positives, true negatives, false positives, and false negatives generated using the DT cutoff score of 4 were examined (see Table 4). Results indicated that 354 people (72.1%) were correctly classified, including 96 (19.6%) true positives—those identified by both the DT and the CES-D as experiencing clinically significant symptomatology and 258 (52.6%) true negatives—those identified by both the DT and the CES-D as not experiencing clinically significant symptomatology. Of those incorrectly classified, 24 (4.9%) were false negatives—individuals identified by the CES-D as having clinically significant depressive symptomatology, but who scored below 4 on the DT. The majority of those misclassified (113 individuals, 23.0%) were false positives—individuals who scored at least 4 on the DT, but less than 16 on the CES-D.

### Table 2. Correlations between CES-D, STAI-S, and DT continuous and cutoff scores and demographic and clinical variables

<table>
<thead>
<tr>
<th></th>
<th>CES-D</th>
<th>STAI-S</th>
<th>DT (0 to 10)</th>
<th>DT cutoff (0 = below DT cutoff; 1 = at or above DT cutoff)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-0.05</td>
<td>-0.07</td>
<td>-0.05</td>
<td>-0.02</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1 = male 2 = female)</td>
<td>0.09*</td>
<td>0.11**</td>
<td>-0.04</td>
<td>-0.08</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1 = White 2 = non-White)</td>
<td>0.11**</td>
<td>0.04</td>
<td>-0.04</td>
<td>-0.08</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1 = married; 2 = not married)</td>
<td>-0.05</td>
<td>-0.07</td>
<td>-0.04</td>
<td>-0.06</td>
</tr>
<tr>
<td>Education</td>
<td>-0.11**</td>
<td>-0.10*</td>
<td>-0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Performance status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0 = no; 1 = yes)</td>
<td>0.31***</td>
<td>0.22***</td>
<td>0.22***</td>
<td>0.20***</td>
</tr>
<tr>
<td>Myeloma</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0 = no; 1 = yes)</td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.05</td>
<td>-0.01</td>
</tr>
<tr>
<td>NHL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0 = no; 1 = yes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMT type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1 = autologous; 2 = allogeneic/MUD)</td>
<td>-0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

*p < 0.05; **p < 0.01; ***p < 0.001.

### Table 3. Frequency distribution of distress thermometer scores

<table>
<thead>
<tr>
<th>N</th>
<th>%</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>68</td>
<td>14</td>
</tr>
<tr>
<td>1</td>
<td>72</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>77</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>65</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>51</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>41</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>38</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>36</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>2</td>
</tr>
</tbody>
</table>

The frequencies of true positives, true negatives, false positives, and false negatives generated using the DT cutoff score of 4 were examined (see Table 4). Results indicated that 354 people (72.1%) were correctly classified, including 96 (19.6%) true positives—those identified by both the DT and the CES-D as experiencing clinically significant symptomatology—and 258 (52.6%) true negatives—those identified by both the DT and the CES-D as not experiencing clinically significant symptomatology. Of those incorrectly classified, 24 (4.9%) were false negatives—individuals identified by the CES-D as having clinically significant depressive symptomatology, but who scored below 4 on the DT. The majority of those misclassified (113 individuals, 23.0%) were false positives—individuals who scored at least 4 on the DT, but less than 16 on the CES-D.

### Relationship of distress thermometer cutoff score to demographic and clinical variables and the Patient Problem List

Correlation coefficients were used to determine the relationship between demographic variables and the scores that met or did not meet the DT cutoff score of 4. Of demographic variables measured, no significant relationships were found with the DT cutoff score (see Table 2).

Among clinical variables, scores above or below the DT cutoff score of 4 were significantly related to performance status ($r = 0.20$, $p < 0.0001$). Greater distress was associated with worse performance status.

Lastly, $\chi^2$ analyses were used to determine whether the scores above or below the DT cutoff score of 4 were associated with the presence of each of the problems found on the Patient Problem List. Of the 33 problems listed, the DT cutoff score was significantly associated with 32 of them at the $p < 0.05$ level, with only mouth sores being non-significant.
DISCUSSION

As with other studies (e.g. Jacobsen et al., 2005), the findings here indicate that the single-item DT compares well with longer measures to assess psychological distress. When patients completed both the CES-D and the DT, the DT distinguished those with clinically significant depressive symptoms from those without such symptoms. Using a DT cutoff score of 4 made this determination most successfully.

Additionally, people who scored higher on the DT had a strong tendency to endorse measures indicating greater anxiety (STAI-S), greater depression (CES-D), and poorer performance status (ECOG). Those scoring at or above the cutoff of 4 also endorsed a wide range of physical, emotional, practical, and family problems. Taken together, the findings provide converging evidence that the DT is an efficacious screener for the presence of psychological distress.

The DT cutoff score found here corresponds with the cutoff score found by a multi-site study of cancer patients (Jacobsen et al., 2005) but differs from previous research (e.g. Akizuki et al., 2003) and from NCCN guidelines for the use of the DT (National Comprehensive Cancer Network, 2003). Currently, the Distress Management Guidelines published by the NCCN suggest that patients scoring 5 or above on the DT be referred for a broader distress assessment. This study and the Jacobsen et al. study indicate that those who score as low as 4 on the DT may also have significant distress and should be considered for a psychosocial referral.

The present study has significant strengths. It uses a sample size (n = 491) larger than other studies that have examined the DT. It also uses a patient sample—BMT candidates—for whom distress is known to be a particularly significant concern (e.g. Fife et al., 2000). Finally, this study was the first to compare the DT with the well-validated and widely used CES-D. However, the participants in this study were predominately Caucasian and college educated. Further research on the DT when used in culturally and socio-economically diverse samples is needed.

<table>
<thead>
<tr>
<th>Cutoff score</th>
<th>Below CES-D</th>
<th>Above CES-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below DT</td>
<td>258 (52.5%)</td>
<td>24 (4.9%)</td>
</tr>
<tr>
<td>Above DT</td>
<td>113 (23.0%)</td>
<td>96 (19.6%)</td>
</tr>
<tr>
<td></td>
<td>371 (75.6%)</td>
<td>120 (24.4%)</td>
</tr>
<tr>
<td></td>
<td>491 (100%)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Frequency of correct and incorrect classifications when using the DT cutoff score of 4, with the CES-D cutoff score of 16 used as a classification criterion.

Figure 2. ROC curve analysis comparing Distress Thermometer scores to established CES-D cutoff score.

The lack of a diverse patient sample in this study precluded careful analysis of the DT in different socio-cultural groups and may explain why the findings of this study differ from a previous study on the DT by Akizuki et al. (2003) that found an optimal DT cutoff score of 5 in a sample of Japanese cancer patients. Although differences in the studies’ methodologies could explain the difference in their findings, the differing cutoff scores could also indicate that the interpretation of the DT may differ based on the culture of origin of the respondent.

The use of the Patient Problem List in this sample indicated that the problems experienced by those who score 4 or above on the DT are wide and varied. Of the 33 problems on the measure, 32 of them were more likely to be reported by patients whose DT scores were above the cutoff point. As would be expected, emotional problems (i.e. worry, fears, sadness, depression and nervousness) showed the greatest relationship with the DT, as demonstrated by the \( \chi^2 \) statistic.

Although it is becoming clearer that the DT can efficaciously screen for distress in samples of cancer patients, it is yet unknown how effective the DT would be as a routine aspect of clinical practice. As with any measure, the DT generates a number of false positives—in the present study 113 patients, or 23% of the entire sample, were flagged by the DT as being distressed despite having CES-D scores less than the clinically significant cutoff of 16. The number of false positives found here exceeded the 96 true positives that were generated by the comparison of the two measures’ scores. The ideal DT cutoff score found here appears to err on the side of caution, at least in terms of the CES-D.

Previous studies have also found relatively high false positive rates for suggested DT cutoff scores. For example, in a study that compared a DT cutoff score of 4 to accepted HADS and BSI-18 cutoff scores, the false positive rates ranged from 20 to 24% (Jacobsen et al., 2005). These findings suggest that a sizable percentage of patients who do not meet the criteria for distress as defined by longer measures will surpass the DT cutoff score. On the other hand, the false negative rate for the DT found here was low (4.9%). This suggests that most of those in this study who scored above the CES-D’s threshold for clinically significant depressive symptomatology were identified using the cutoff score recommended here. Although a low false negative rate helps ensure that maximum number of distressed patients would be identified using this score, clinicians who consider using the DT as a screening measure should be aware that the number of false positives generated by the DT may attenuate the amount of time saved by the measure’s ease of use and interpretability.

NOTES

1. The DT was not compared with the STAI-S to determine the DT’s operating characteristics because there is no widely accepted STAI-S cutoff score for clinically significant anxiety.

REFERENCES


