Sensitivity and specificity of the Distress Thermometer in screening for distress in long-term nasopharyngeal cancer survivors

J.S. Hong PhD* and J. Tian PhD†

ABSTRACT

Background

The Distress Thermometer (dt) is a screening tool recommended to quickly identify cancer patients with distress. Our study aimed to examine the sensitivity and specificity of the dt in detecting psychological distress in long-term Chinese nasopharyngeal cancer (NPC) survivors.

Methods

Data for the 442 participating NPC survivors were collected through a self-administered questionnaire based on the dt and the Hospital Anxiety and Depression Scale (HADS). The HADS was used to define cases of psychological distress. Positive and negative groups were defined based on 4 HADS criteria (Anxiety, Depression, Anxiety or Depression, and overall score). Receiver operating characteristic (ROC) curves were used to examine the ability of all possible cut-off values of the dt to detect positive and negative cases. For each ROC curve, the area under the curve (AUC) was used as an indicator of the overall accuracy of the dt to identify positive cases of distress.

Results

The positive AUC values [with 95% confidence intervals (CI)] for the 4 HADS criteria were 0.715 (95% CI: 0.667 to 0.764), 0.714 (95% CI: 0.661 to 0.768), 0.724 (95% CI: 0.677 to 0.771), and 0.724 (95% CI: 0.664 to 0.775) respectively. At a cut-off score of 4, the sensitivity of the dt to the four HADS criteria was, respectively, 0.366 (95% CI: 0.296 to 0.436), 0.448 (95% CI: 0.364 to 0.532), 0.362 (95% CI: 0.299 to 0.425), and 0.421 (95% CI: 0.339 to 0.502), and the specificity of the dt to the 4 HADS criteria was, respectively, 0.860 (95% CI: 0.818 to 0.902), 0.860 (95% CI: 0.821 to 0.899), 0.854 (95% CI: 0.814 to 0.894), and 0.854 (95% CI: 0.814 to 0.894). At a cut-off score of 5, the corresponding sensitivities were lower than those at the cut-off score of 4. All potential cut-off scores showed poor sensitivity (<0.90).

Conclusions

The ROC analysis showed poor discrimination. No potential dt cut-off score had an acceptable sensitivity. The dt showed poor sensitivity in NPC survivors. Thus, the dt might not be a valid scale for psychological distress screening in long-term Chinese NPC survivors.

KEY WORDS

Distress, nasopharyngeal cancer survivors, validity, Distress Thermometer

1. INTRODUCTION

Cancer can cause patients to experience serious psychological distress 1. During diagnosis and treatment of the disease, patients not only experience a series of physical problems, such as pain, fatigue, nausea, vomiting, and sleep disorders, but also a series of mental health issues, including fear, sadness, anxiety, and depression 2. These physical and mental problems cause psychological distress in patients 3.

Distress is defined as an unpleasant emotional, psychological, social, or spiritual experience that interferes with the ability to cope with cancer treatment. The distress experience extends along a continuum, from normal feelings of vulnerability, sadness, and fear, to disabling problems such as true depression, anxiety, panic, and feelings of being isolated or being in a spiritual crisis 3,4.

For cancer patients, psychological distress is likely to be present at any stage from diagnosis to survivorship, thereby affecting various aspects of life 5–7. Psychological distress has been linked to
decreased social functioning, further physical and cognitive impairment \(^1,^8,^9\), and nonadherence to treatments and health-promoting behaviors \(^10,^11,^12\). Patients with psychological distress have a high probability of tumour recurrence \(^13,^14\), a low survival rate \(^14,^15\), and poor performance status and quality of life \(^15,^16,^17\). Because survival and rehabilitation can be affected by psychological distress, the timely discovery of severe distress in cancer patients and cancer survivors and the availability of psychological care and interventions should be a focus of health care providers.

The U.S. National Comprehensive Cancer Network (nccn) has suggested that assessment of psychological distress in cancer patients should be added to the standard program of cancer treatment \(^18,^19\). The nccn recommends that care providers distinguish a normal emotional reaction after cancer diagnosis from psychological distress, estimate the degree of potential distress in cancer patients, and identify patients with severe distress. Assessment of a high risk of psychological distress in patients allows health care providers to deliver timely psychological interventions and treatment.

Various screening tools have been developed to quickly identify people who may be psychologically distressed \(^18,^20,^21\). The Distress Thermometer (dt) is recommended by the nccn \(^18\). The dt is a single-item scale used to score individuals based on their current situation. A high score expresses a high degree of distress. The potential advantages of the dt over other screening tools are its brevity and ease of administration and scoring. The dt is therefore routinely used in numerous studies of cancer patients to obtain a self-measure of psychological distress. The ability of the dt to detect distress in cancer patients has been widely reported (sensitivity: 0.56–0.83; specificity: 0.49–0.85) \(^22–^27\).

Nasopharyngeal carcinoma (npc) is a common cancer in southern China and Southeast Asia. With advances in cancer treatment, the 5-year survival rate in npc patients has improved \(^28\). Hence, various studies have focused on determining long-term quality of life and psychological adjustment in npc survivors \(^29,^30\). Assessing distress in npc survivors and providing them psychosocial care should be part of routine follow-up. However, application of the dt to cancer survivors has not been widely investigated \(^31\). Little research has been done on the validity of dt in npc survivors. Accordingly, the present study aimed to evaluate the sensitivity and specificity of the dt in detecting psychological distress in long-term Chinese npc survivors. We selected the Hospital Anxiety and Depression Scale (hads) as the comparison measure, because it is among the most widely used tools to screen for symptoms of anxiety and depression in oncology populations \(^20,^32\). It has also been recommended as the tool of choice in recent reviews \(^33,^34\).

2. METHODS

2.1 Participants

Participants were patients visiting the Fujian Tumor Hospital from January 2011 to December 2012. Eligible survivors were those who had been diagnosed with primary npc more than a year earlier, who were between 18 and 70 years of age, who were without mental or psychological disease, and who understood their cancer diagnosis.

All participants provided written informed consent. The study was approved by the relevant institutional review boards for human research at Fujian Medical University.

2.2 Measures

The nccn’s dt was used to measure distress as self-reported by participants. The dt is a visual analog scale on which participants rate their level of distress in the preceding 7 days from 0 (none) to 10 (extreme). Some studies have suggested that a cut-off score of 4 or more indicates distress \(^35,^36\); others support a cut-off score of 5 or more \(^24\). The validity of the Chinese version of the dt was confirmed in a previous study \(^37\). The hads, a 14-item questionnaire (7 items for the Anxiety subscale, and 7 for the Depression subscale) \(^30\), was used to evaluate anxiety and depression in the study patients. The point range for each item is 0 to 3. Patients score the items based on their current situation. The scores for the Anxiety and the Depression subscales both range from 0 to 21, with 0–7 indicating “asymptomatic,” 8–10 indicating “suspicous symptoms,” and 11–21 indicating “certainly existing symptoms” \(^20\). The Chinese version of the hads has been confirmed to be suitable for Chinese patients \(^32\). In the present study, the Chinese version of the hads had reliabilities of 0.89 for anxiety and 0.92 for depression.

2.3 Procedure

Trained graduate students from Fujian Medical University introduced each eligible patient to the study. After obtaining informed consent, the investigator asked the patient to complete a self-administered questionnaire consisting of demographic questions, the dt, and the hads.

2.4 Statistical Analysis

The hads was used to define cases of psychological distress. The participants were divided into two groups (positive and negative for distress) based on four hads criteria:

- Anxiety subscale scores of 8 or higher (positive group) and less than 8 (negative group)
• Depression subscale scores of 8 or higher (positive group) and less than 8 (negative group)
• An Anxiety or Depression subscale score of 8 or higher (positive group) and both subscales less than 8 (negative group)
• A total score of 16 or higher (positive group) and less than 16 (negative group)

Receiver operating characteristic (ROC) curves were used to examine the ability of all possible cut-off values of the DT to detect positive and negative cases of distress. For each ROC curve, the area under the curve (AUC) was used as an indicator of the overall accuracy of the DT in identifying positive cases of distress. The AUC values vary from 0 to 1, with values of 0.80 or more indicating good discrimination. Sensitivity (the true positive rate), specificity (the true negative rate), and 95% confidence intervals (CIs) were calculated for each DT score.

Statistical analysis was performed using SAS for Windows (version 9.0: SAS Institute, Cary, NC, U.S.A.).

3. RESULTS

3.1 Participants

Of the 552 NPC survivors eligible for the study, 107 did not consent to complete the survey, and 3 did not finish the questionnaire. The 442 NPC survivors who completed the survey included 318 men (72%) and 124 women (28%) with a mean age of 46.4 ± 10.7 years. The proportion of the participants with primary school, middle school, high school, and college education were 20.4%, 36.2%, 24.2%, and 19.2% respectively. Time since diagnosis was 1–2.9, 3–4.9, 5–6.9, and 7 years in 42 (9.5%), 281 (63.6%), 116 (26.2%), and 3 patients (0.7%) respectively.

3.2 Areas Under the ROC Curve

Figure 1 shows the ROC curves for the 4 positive HADS criteria; Table i shows the AUCs and their corresponding 95% CIs. The upper limits of the 95% CIs are lower than 0.8. Hence, the ROC AUCs for the HADS-defined positive cases are less than 0.8 for every criterion (p < 0.05), indicating that the DT provided poor discrimination in long-term NPC survivors.

3.3 Sensitivity and Specificity

Each DT score was treated as a potential cut-off score, and the sensitivities and specificities according to the 4 HADS criteria were calculated (Table ii). For a cut-off score of 4, the sensitivity of the DT for the 4 HADS criteria was 0.366 (95% CI: 0.296 to 0.436), 0.448 (95% CI: 0.364 to 0.532), 0.362 (95% CI: 0.299 to 0.425), and 0.421 (95% CI: 0.339 to 0.502) respectively, and the specificity of the DT for the 4 HADS criteria was 0.634 (95% CI: 0.547 to 0.721), 0.572 (95% CI: 0.487 to 0.657), 0.638 (95% CI: 0.553 to 0.723), and 0.580 (95% CI: 0.480 to 0.679) respectively.

**FIGURE 1** Analysis of the receiver operating characteristic curves for comparisons of Distress Thermometer scores with criteria from the Hospital Anxiety and Depression Scale (HADS). (A) HADS Anxiety score of 8 or higher. (B) HADS Depression score of 8 or higher. (C) HADS Anxiety or Depression score of 8 or higher. (D) Total HADS score of 16 or higher.
HONG and TIAN showed poor agreement with the in cancer patients in the: 0.814 to 0.894) respectively. The sensitivities corresponding to a cut-off score of 5 were much lower than those corresponding to a cut-off score of 4, but the specificities were higher. The results shown in Table II suggest that all potential cut-off scores reflect poor sensitivity (<0.90). If a sensitivity of 0.9 or more were to be used as the criterion for evaluating the ΔT, no possible ΔT cut-off score can detect distress in long-term NPC survivors. Similarly, when using the ΔT to detect psychological distress, regardless of whether 4 or 5 is used as the cut-off score, sensitivity did not exceed 0.5.

4. DISCUSSION AND CONCLUSIONS

In this study of a sample of long-term NPC survivors, the ΔT showed poor agreement with the HADS, which is a more detailed measure of psychological distress. Using the NCCN-recommended cut-off score of 5, the ΔT identified fewer than 30% of the NPC survivors who were psychologically distressed according to the HADS criteria. In addition, ROC analysis showed poor discrimination, and no potential ΔT cut-off score had an acceptable sensitivity. The results suggest that the ΔT has poor sensitivity in Chinese NPC survivors. Hence, the ΔT might not be a valid screening tool for psychological distress in long-term NPC survivors in China.

The NCCN recommends the ΔT as an initial screening tool for assessing distress in cancer patients. The validity of the ΔT in cancer patients in the treatment phase has been confirmed by many studies. However, data about the utility of the ΔT for individuals in the survivorship phase are limited. In a study comparing the ΔT with the Brief Symptom Inventory 18 in a sample of cancer survivors 2 years after diagnosis, and in a study comparing the ΔT with the HADS in a sample of colorectal cancer survivors, the ΔT demonstrated poor sensitivity. In another study comparing the ΔT with the Symptom Checklist-90-Revised in a sample of adult survivors of childhood cancer, the results also do not support the validity of ΔT. Our findings strongly suggest that, in screening for distress, the ΔT is less valid in cancer survivors than in cancer patients.

Cultural differences may also contribute to the low sensitivity of the ΔT in the present study. Many Chinese people think that psychological problems are disgraceful, and people with psychological distress often consider themselves to be “weak.” Cancer survivors might therefore be reluctant to admit to psychological distress and might give low scores to hide their true feelings. However, Tang et al. suggested that the Chinese version of the ΔT has a good sensitivity and specificity in Chinese patients with cancer, which indicates that cultural differences may not be the main reason for the low sensitivity of the ΔT in our cohort of cancer survivors. However, the cohort studied by Tang et al. were hospitalized patients with cancer; our participants were outpatient survivors of NPC for at least 1 year after diagnosis, with 90% having survived for more than 3 years.

Some researchers believe that differences in the performance of the ΔT between cancer patients and cancer survivors might reflect differences in cancer treatment, emotional status, and understanding of distress. That is, the sensitivity of the ΔT in measuring psychological distress is associated with the individual’s disease stage. Compared with newly diagnosed patients, NPC survivors with a survival duration of more than 1 year are generally not so fearful and worried; hence, the survivors might not think that they are in distress. However, they might in fact still be experiencing anxiety or depression concerning disease recurrence. That anxiety or depression creates inconsistency between the ΔT and HADS scores, leading to the poor sensitivity observed for the ΔT. In long-term cancer survivors, the HADS may be better than the ΔT as a tool in screening for distress.

Another possible reason for the poor sensitivity of the ΔT might be the understanding of the ΔT score by patients. Higher ΔT scores indicate higher distress. However, if the meaning of each score is not clearly understood, the patient will not be sure of how the score describes their situation. Unlike the ΔT, the HADS...
items are assigned scores from 0 to 3 that correspond to specific degrees. For example, for the item “I feel tense or wound up,” the options are 0 (not at all), 1 (sometimes), 2 (frequently), and 3 (all the time). The HADS is therefore easy to score and administer. The scores do not attach a description to the degree of distress, which might lead to an inconsistent result between the DT and the HADS. When using the DT to detect distress in cancer survivors, providing an explanation for the cut-off score may help to increase the sensitivity of the DT.

Although the HADS has been well validated, its high rate of false positives for the detection of anxiety and depression must be mentioned. Using a threshold

### Table II

Sensitivity and specificity of Distress Thermometer (DT) cut-off scores by the four Hospital Anxiety and Depression Scale (HADS) criteria

<table>
<thead>
<tr>
<th>HADS criterion</th>
<th>DT</th>
<th>Sensitivity</th>
<th>95% CI</th>
<th>Specificity</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety subscale score ≥8</td>
<td>≥1</td>
<td>0.760</td>
<td>0.698 to 0.822</td>
<td>0.597</td>
<td>0.537 to 0.657</td>
</tr>
<tr>
<td></td>
<td>≥2</td>
<td>0.612</td>
<td>0.542 to 0.682</td>
<td>0.694</td>
<td>0.638 to 0.750</td>
</tr>
<tr>
<td></td>
<td>≥3</td>
<td>0.426</td>
<td>0.355 to 0.497</td>
<td>0.818</td>
<td>0.771 to 0.865</td>
</tr>
<tr>
<td></td>
<td>≥4</td>
<td>0.366</td>
<td>0.296 to 0.436</td>
<td>0.860</td>
<td>0.818 to 0.902</td>
</tr>
<tr>
<td></td>
<td>≥5</td>
<td>0.186</td>
<td>0.130 to 0.242</td>
<td>0.942</td>
<td>0.913 to 0.971</td>
</tr>
<tr>
<td></td>
<td>≥6</td>
<td>0.120</td>
<td>0.073 to 0.167</td>
<td>0.969</td>
<td>0.948 to 0.990</td>
</tr>
<tr>
<td></td>
<td>≥7</td>
<td>0.066</td>
<td>0.030 to 0.102</td>
<td>0.988</td>
<td>0.975 to 1.000</td>
</tr>
<tr>
<td></td>
<td>≥8</td>
<td>0.027</td>
<td>0.004 to 0.050</td>
<td>0.992</td>
<td>0.981 to 1.000</td>
</tr>
<tr>
<td></td>
<td>≥9</td>
<td>0.016</td>
<td>0.000 to 0.034</td>
<td>0.996</td>
<td>0.988 to 1.000</td>
</tr>
<tr>
<td>Depression subscale score ≥8</td>
<td>≥1</td>
<td>0.791</td>
<td>0.722 to 0.859</td>
<td>0.554</td>
<td>0.498 to 0.610</td>
</tr>
<tr>
<td></td>
<td>≥2</td>
<td>0.642</td>
<td>0.561 to 0.723</td>
<td>0.658</td>
<td>0.605 to 0.711</td>
</tr>
<tr>
<td></td>
<td>≥3</td>
<td>0.515</td>
<td>0.431 to 0.599</td>
<td>0.818</td>
<td>0.775 to 0.861</td>
</tr>
<tr>
<td></td>
<td>≥4</td>
<td>0.448</td>
<td>0.364 to 0.532</td>
<td>0.860</td>
<td>0.821 to 0.899</td>
</tr>
<tr>
<td></td>
<td>≥5</td>
<td>0.216</td>
<td>0.146 to 0.285</td>
<td>0.935</td>
<td>0.907 to 0.962</td>
</tr>
<tr>
<td></td>
<td>≥6</td>
<td>0.134</td>
<td>0.076 to 0.191</td>
<td>0.961</td>
<td>0.939 to 0.983</td>
</tr>
<tr>
<td></td>
<td>≥7</td>
<td>0.082</td>
<td>0.036 to 0.128</td>
<td>0.987</td>
<td>0.974 to 0.999</td>
</tr>
<tr>
<td></td>
<td>≥8</td>
<td>0.037</td>
<td>0.005 to 0.069</td>
<td>0.993</td>
<td>0.984 to 1.000</td>
</tr>
<tr>
<td></td>
<td>≥9</td>
<td>0.022</td>
<td>0.000 to 0.047</td>
<td>0.997</td>
<td>0.991 to 1.000</td>
</tr>
<tr>
<td>Anxiety or Depression subscale score ≥8</td>
<td>≥1</td>
<td>0.732</td>
<td>0.674 to 0.790</td>
<td>0.636</td>
<td>0.572 to 0.700</td>
</tr>
<tr>
<td></td>
<td>≥2</td>
<td>0.585</td>
<td>0.521 to 0.649</td>
<td>0.724</td>
<td>0.664 to 0.783</td>
</tr>
<tr>
<td></td>
<td>≥3</td>
<td>0.420</td>
<td>0.355 to 0.484</td>
<td>0.857</td>
<td>0.810 to 0.903</td>
</tr>
<tr>
<td></td>
<td>≥4</td>
<td>0.362</td>
<td>0.299 to 0.425</td>
<td>0.899</td>
<td>0.859 to 0.939</td>
</tr>
<tr>
<td></td>
<td>≥5</td>
<td>0.183</td>
<td>0.132 to 0.233</td>
<td>0.963</td>
<td>0.938 to 0.988</td>
</tr>
<tr>
<td></td>
<td>≥6</td>
<td>0.107</td>
<td>0.067 to 0.147</td>
<td>0.972</td>
<td>0.950 to 0.994</td>
</tr>
<tr>
<td></td>
<td>≥7</td>
<td>0.058</td>
<td>0.027 to 0.088</td>
<td>0.991</td>
<td>0.978 to 1.000</td>
</tr>
<tr>
<td></td>
<td>≥8</td>
<td>0.027</td>
<td>0.006 to 0.048</td>
<td>0.995</td>
<td>0.985 to 1.000</td>
</tr>
<tr>
<td></td>
<td>≥9</td>
<td>0.013</td>
<td>0.000 to 0.028</td>
<td>0.995</td>
<td>0.985 to 1.000</td>
</tr>
<tr>
<td>Total score ≥16</td>
<td>≥1</td>
<td>0.807</td>
<td>0.742 to 0.872</td>
<td>0.568</td>
<td>0.512 to 0.624</td>
</tr>
<tr>
<td></td>
<td>≥2</td>
<td>0.664</td>
<td>0.586 to 0.742</td>
<td>0.674</td>
<td>0.621 to 0.727</td>
</tr>
<tr>
<td></td>
<td>≥3</td>
<td>0.486</td>
<td>0.403 to 0.568</td>
<td>0.811</td>
<td>0.767 to 0.855</td>
</tr>
<tr>
<td></td>
<td>≥4</td>
<td>0.421</td>
<td>0.339 to 0.502</td>
<td>0.854</td>
<td>0.814 to 0.894</td>
</tr>
<tr>
<td></td>
<td>≥5</td>
<td>0.193</td>
<td>0.128 to 0.258</td>
<td>0.927</td>
<td>0.898 to 0.956</td>
</tr>
<tr>
<td></td>
<td>≥6</td>
<td>0.136</td>
<td>0.079 to 0.192</td>
<td>0.963</td>
<td>0.942 to 0.984</td>
</tr>
<tr>
<td></td>
<td>≥7</td>
<td>0.071</td>
<td>0.029 to 0.113</td>
<td>0.983</td>
<td>0.968 to 0.997</td>
</tr>
<tr>
<td></td>
<td>≥8</td>
<td>0.029</td>
<td>0.001 to 0.057</td>
<td>0.990</td>
<td>0.979 to 1.000</td>
</tr>
<tr>
<td></td>
<td>≥9</td>
<td>0.021</td>
<td>0.000 to 0.045</td>
<td>0.997</td>
<td>0.991 to 1.000</td>
</tr>
</tbody>
</table>
score of 8+, the sensitivity and specificity of the HADS for detecting anxiety and depression usually range from 0.70 to 0.90 \(^{43}\), which suggests that the HADS has a false positive rate of 10%—30%. When the HADS is used as the sole criterion for defining cases of psychological distress, it will overestimate the number of individuals with anxiety or depression. The high false positive rate of the HADS might therefore also partly explain the low sensitivity of the DT in our study. Further research should be conducted to evaluate the validity of the DT when compared with clinical diagnostic criteria for anxiety and depression.

It should be noted that all of the specificity values for the DT in our study were high (≥85%). A high specificity suggests that the DT has a low false-positive rate in screening for psychological distress. Community health nurses should therefore provide psychological nursing and interventions to individuals identified by the DT as experiencing psychological distress.

Our study has limitations. First, only one criterion was used to validate DT. In addition, only survivors visiting the hospital during the study period were investigated; survivors not attending the hospital were ignored. That sample bias might have contributed to the poor sensitivity of DT observed in the study. Further research should be conducted using additional measures of psychological distress to evaluate the validity of the DT in NPC survivors in the community.

5. ACKNOWLEDGMENTS

The authors thank Katherine Casey–Sawicki for editing and preparing this manuscript for publication.

6. CONFLICT OF INTEREST DISCLOSURES

The present study has no financial relationship with any sponsoring organization, and the authors have no financial conflicts of interest to declare.

7. REFERENCES


Correspondence to: Jun Tian, Department of Epidemiology and Health Statistics, Fujian Medical University, Fuzhou, Fuzhou 350004, Fujian Province, PR China.
E-mail: tianjunfjmu@126.com

* Department of Radiation Oncology, Department of Radiation Biology, First Affiliated Hospital of Fujian Medical University, Fuzhou, PR China.
† Department of Epidemiology and Health Statistics, Fujian Medical University, Fuzhou, PR China.