Assessment of a new self-rating scale for post-traumatic stress disorder


From the Department of Psychiatry and Behavioral Sciences, Duke University Medical Center
Department of Psychology, Duke University and Veterans Administration Medical Center, Durham, NC; Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina; SC; Department of Psychiatry, University of Miami, FL; and Ciba-Geigy Pharmaceuticals, Summit, NJ, USA

ABSTRACT

Background. In post-traumatic stress disorder (PTSD) there is a need for self-rating scales that are sensitive to treatment effects and have been tested in a broad range of trauma survivors. Separate measures of frequency and severity may also provide an advantage.

Methods. Three hundred and fifty-three men and women completed the Davidson Trauma Scale (DTS), a 17-item scale measuring each DSM-IV symptom of PTSD on 5-point frequency and severity scales. These subjects comprised war veterans, survivors of rape or hurricane and a mixed trauma group participating in a clinical trial. Other scales were included as validity checks as follows: Global ratings, SCL-90-R, Eysenck Scale, Impact of Event Scale and Structured Clinical Interview for DSM-III-R.

Results. The scale demonstrated good test–retest reliability (r = 0.86), internal consistency (r = 0.99). One main factor emerged for severity and a smaller one for intrusion. In PTSD diagnosed subjects, and the factor structure more closely resembled the traditional grouping of symptoms. Concurrent validity was obtained against the SCID, with a diagnostic accuracy of 83% at a DTS score of 40. Good convergent and divergent validity was obtained. The DTS showed predictive validity against response to treatment, as well as being sensitive to treatment effects.

Conclusions. The DTS showed good reliability and validity, and offers promise as a scale which is particularly suited to assessing symptom severity, treatment outcome and in screening for the likely diagnosis of PTSD.

INTRODUCTION

Post-traumatic stress disorder (PTSD) was first included in the Diagnostic and Statistical Manual of Mental Disorders, 3rd edition (DSM-III), in 1980 (American Psychiatric Association, 1980). Since then, several different rating scales have been introduced to measure PTSD. These scales are typically clinician administered and thus limited by requiring the time and expertise of a qualified interviewer. Examples of clinician-administered rating scales include the Structured Interview for PTSD (SI-PTSD) (Davidson et al. 1989), the Clinician Administered PTSD Scale (CAPS) (Blake et al. 1990), and the Structured Clinical Interview for DSM-III-R (SCID) (Spitzer et al. 1990). Although several self-rating scales exist in the literature, including the Impact of Event Scale (IES) (Horowitz et al. 1979), Mississippi Scale (Keane et al. 1988), Penn Inventory (Hammarberg, 1992), PTSD Inventory (Solomon et al. 1993) and PTSD Symptom Scale (Foa et al. 1993), these are all limited because reliability and validity have not been demonstrated in widely ranging populations. The IES, which has been adopted as a standard self-rating instrument, pre-dates DSM-III and...
fails to incorporate hyperarousal symptoms, an important component of the PTSD symptom complex. Moreover, none of the above self-rating scales separately addresses frequency and severity of symptoms.

We have developed the Davidson Trauma Scale (DTS), a self-rated scale tailored closely to the symptom definitions of the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) (American Psychiatric Association, 1994). The DTS is designed to evaluate symptoms of PTSD in individuals with a history of trauma. Its primary purposes are to measure symptom frequency and severity and to evaluate treatment, for example, measurement of symptom change over time, response prediction, and evaluation of differences between therapy modalities in the research setting. In this report, we present data on reliability and validity of the scale. We have administered the DTS to over 300 subjects from separate clinical research studies of rape, combat and natural disaster, as well as a sample of mixed trauma survivors participating in a double-blind, placebo-controlled pharmacotherapy study of PTSD. We evaluated the following measures of reliability and validity: test–retest reliability, internal consistency, and factorial, concurrent, convergent, discriminant and predictive validity.

**METHOD**

**Subjects**

We administered the DTS to 353 subjects who had taken part in one of four clinical research studies. Seventy-eight women participated in a study of rape victims (Study 1), 110 men in a study of war veterans (Study 2), 53 subjects in a study of Hurricane Andrew victims (Study 3), and 102 subjects in a multicentre clinical trial of an antidepressant drug in survivors of miscellaneous traumas (Study 4). Table 1 presents demographic breakdowns for the four samples. Each of these studies rendered some shared and some unique data, depending upon the measures involved (described below).

**Davidson Trauma Scale**

The DTS is composed of 17 items corresponding to each of the 17 DSM-IV symptoms. Items can be categorized as follows: items 1–4, 17: criteria B (intrusive re-experiencing); items 5–11: criteria C (avoidance and numbness); and items 12–16: criteria D (hyperarousal). For each item, the subject rates both frequency and severity during the previous week on a 5-point (0 to 4) scale for a total possible score of 136 points. Subscale scores can be computed separately for frequency and severity.

**Other clinical measures**

In addition to the DTS, subjects were assessed with several other scales useful for assessing validity. For concurrent validity, the SCID was administered as an independent validator. Positive and negative predictive value, sensitivity, specificity, and efficiency were determined. We also examined DTS scores as a function of ratings on the physician-rated Global Assessment of Severity (GASP) (Katz, 1995, personal communication). For convergent validity with other quantitative PTSD scales, the DTS total score was compared with total scores of both the CAPS and the IES, as well as with the subscales,

---

#### Table 1. Demographic characteristics of the study samples (N = 353). Study 1 constituted rape survivors, study 2 war veterans, study 3 Hurricane Andrew victims and study 4 miscellaneous trauma survivors participating in a multicentre antidepressant trial.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Study 1</th>
<th>Study 2</th>
<th>Study 3</th>
<th>Study 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age ± s.d.</td>
<td>26.4 ± 6.9</td>
<td>46.7 ± 7.7</td>
<td>40.8 ± 12.7</td>
<td>44.7 ± 7.4</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0</td>
<td>110</td>
<td>11</td>
<td>84</td>
</tr>
<tr>
<td>Female</td>
<td>78</td>
<td>0</td>
<td>42</td>
<td>18</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>20</td>
<td>73</td>
<td>33</td>
<td>Not available</td>
</tr>
<tr>
<td>Not married</td>
<td>58</td>
<td>37</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>66</td>
<td>66</td>
<td>36</td>
<td>Not available</td>
</tr>
<tr>
<td>Other</td>
<td>12</td>
<td>44</td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>
index score, and total score of the SCL-90-R (Derogatis, 1977). The Eysenck Personality Inventory Extroversion scale (Eysenck & Eysenck, 1968) was compared with the DTS to evaluate discriminant validity, as a null relationship was expected between the two measures. Finally, the Clinical Global Impressions (CGI) Improvement scale (Guy, 1976) was used to assess sensitivity to treatment effects and predictive validity.

Statistical analysis

Internal consistency was evaluated using Cronbach’s $\alpha$. Test–retest reliability and convergent and discriminant validity were assessed using Pearson product–moment correlations. Factorial validity was examined using exploratory factor analyses. Concurrent validity was assessed by comparing DTS scores of patients with and without SCID PTSD diagnoses via $t$ test. Sensitivity to treatment effects was examined using a general linear models (GLM) analysis, with CGI Improvement category as a grouping variable and time (treatment baseline and endpoint) as a repeated measure. Predictive validity was evaluated by regressing DTS scores against CGI Improvement ratings. Subsamples utilized for particular analyses are reported in pertinent results sections.

RESULTS

Test–retest reliability

Test–retest reliability was examined by comparing the baseline total DTS score with a DTS assessment two weeks later in subjects from the multicentre clinical trial who had been rated with ‘no change’ on the CGI Improvement scale ($N = 21$). The test–retest reliability coefficient was 0.86 ($P < 0.0001$).

Internal consistency

Cronbach’s $\alpha$ was used to evaluate internal consistency of the DTS in 241 patients recruited from the rape-victims study, the war veterans, and the study of Hurricane Andrew victims. For the 17 frequency and severity items, the coefficient was 0.99; for the frequency items alone it was 0.97 and for the severity items alone it was 0.98.

Factorial validity

Principal-components factor analysis of data from studies 1–3 ($N = 241$) revealed the presence of two main factors, one of which accounted for over 20% of the variance (eigenvalue = 24.19) and is interpreted as a severity factor. The second factor, with an eigenvalue of 1.34, accounted for a small amount of variance. This factor largely consisted of positive loadings on intrusive items and negative loadings on avoidance and numbing items (Table 2).

Concurrent validity

SCID diagnoses were used to assess concurrent validity. One hundred twenty-nine subjects were administered the SCID. These came from the rape-victims study, the study of war veterans and the Hurricane Andrew study. Sixty-seven of these subjects met a SCID diagnosis of current PTSD. These subjects had a mean ± S.D. total DTS score of 62.0 ± 38.0, whereas those who did not meet current PTSD criteria ($N = 62$) displayed significantly lower scores: 15.5 ± 13.8 ($t = 9.37$, $P < 0.0001$).

Sensitivity, specificity, predictive value and efficiency were calculated for all possible DTS
scores relative to a SCID-based diagnosis of PTSD as independent validator according to the definitions of Insel & Goodwin (1983). Table 4 shows five different threshold scores and their corresponding sensitivity (percentage with PTSD scoring at threshold or higher), specificity (percentage without PTSD scoring below threshold), predictive value of a positive test (percentage scoring at or above threshold who have PTSD), predictive value of a negative test (percentage scoring below threshold who do not have PTSD), and efficiency (percentage correctly classified as having PTSD or as not having PTSD). The highest efficiency was found at a total score of 40. The area under the curve (± standard error) was 0.88 (±0.02).

A self-rated scale which purports to assess symptom severity can also be evaluated relative to an independent severity measure. Such a marker was used in the multicentre clinical trial, wherein the GASP provided a 15-point rating of severity, ranging from 1–3 (minimal, within the range of normality), 4–6 (subclinical PTSD), 7–9 (clinical PTSD), 10–12 (severe PTSD) and 13–15 (very severe PTSD). Specific descriptors of each category were given to raters. Treatment endpoint scores were used as opposed to baseline measures because the former provided a greater distribution of severity. Mean DTS total scores for each GASP category were as follows: minimal (N = 17), 14.0 ± 13.8; subclinical (N = 27), 41.7 ± 28.1; clinical (N = 36), 78.5 ± 27.1;
severe (N = 15), 108.5 ± 15.4; very severe (N = 2), 114.0 ± 8.4. An ANOVA revealed highly significant differences between these five groups (F(4, 92) = 42.7, P < 0.0001).

Convergent and discriminant validity
To assess convergent validity with other PTSD rating scales, we compared the DTS total score with the CAPS and IES total scores. Subjects from studies 1 and 2 (N = 102) were included in this analysis. The correlations were 0.78 (P < 0.0001) for the CAPS and 0.64 (P < 0.0001) for the IES. With reference to the IES, we further parsed the DTS into three subsections cor-
Table 5. Correlations of DTS total score with SCL-90-R subscales, index score, and total score (N = 123)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Correlation*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Somatization</td>
<td>0.49</td>
</tr>
<tr>
<td>Obsessive-compulsive</td>
<td>0.52</td>
</tr>
<tr>
<td>Interpersonal</td>
<td>0.57</td>
</tr>
<tr>
<td>Depression</td>
<td>0.58</td>
</tr>
<tr>
<td>Anxiety</td>
<td>0.65</td>
</tr>
<tr>
<td>Hostility</td>
<td>0.44</td>
</tr>
<tr>
<td>Phobic avoidance</td>
<td>0.52</td>
</tr>
<tr>
<td>Paranoid ideation</td>
<td>0.51</td>
</tr>
<tr>
<td>Psychoticism</td>
<td>0.48</td>
</tr>
<tr>
<td>Index score</td>
<td>0.57</td>
</tr>
<tr>
<td>Total score</td>
<td>0.57</td>
</tr>
</tbody>
</table>

* All P values < 0.0001.

responding to re-experiencing/intrusions, avoidance/numbness and hyperarousal. The IES avoidance subsection correlated 0.52 (P < 0.0001) with the DTS numbness/avoidance items. The IES intrusion subsection correlated 0.77 (P < 0.0001) with the DTS reexperiencing/intrusion items.

As another measure of convergent validity, 123 subjects from the rape-victims and Hurricane Andrew studies completed the SCL-90-R. The total score, subscales, and index score were each correlated with the DTS total score. Results, presented in Table 5, revealed significant correlations between total DTS score and each of the SCL-90-R scales, with the highest association noted for the anxiety subscale.

The Eysenck Personality Inventory Extraversion subscale was correlated with the DTS total score to evaluate discriminant validity. The resulting coefficient was 0.04 (NS).

Predictive validity

To evaluate predictive validity, baseline DTS total scores were examined in relation to endpoint CGI ratings in Study 4 to explore whether initial severity on the DTS predicted response to double-blind treatment. Subjects rated at endpoint as very much improved (N = 26) scored 63.5 ± 28.6; those much improved (N = 24) scored 84.5 ± 26.0; minimally improved (N = 22) 84.9 ± 25.9; no change (N = 13) 84.5 ± 26.1; minimally to markedly worse (N = 12) 94.3 ± 29.0. A regression analysis examining CGI response predicted by baseline DTS score revealed a significant positive relationship (P < 0.005), although the R² was low (0.10).

Sensitivity to treatment effects

Sensitivity to treatment effects was evaluated by comparing the DTS total score in responders versus non-responders to the double-blind medication trial (Study 4). Response was defined as an endpoint CGI score of 1 or 2; non-response was defined as a CGI score of 3 or greater. A significant interaction was observed between responder status and DTS totals at baseline and endpoint (F(1, 95) = 31.75, P < 0.0001). The 50 CGI responders had mean DTS total scores of 73.6 ± 29.1 and 40.3 ± 32.9 at baseline and endpoint, respectively, whereas the 47 CGI non-responders had mean scores of 87.2 ± 26.5 and 85.8 ± 32.1, respectively.

DISCUSSION

The DTS was developed as a self-rating scale measuring frequency and severity of each DSM-IV symptom of PTSD in subjects having identified an unusually traumatic event, or set of events. The intrusive and avoidant items are asked with reference to the event, while the numbing, withdrawal and hyperarousal items are rated as present or absent without direct linking to the event – a slight departure from the DSM-IV and the CAPS, which require these symptoms to have arisen after the event. We judged that the ability of a subject to make an accurate assessment in this respect after either remote childhood trauma or chronic and persistent re-traumatization would be difficult. This point of view receives some support from the study of Solomon & Canino (1990), which demonstrated poor reliability of patient descriptions of such symptomatology. Because of this slight departure from the way in which DSM-IV symptoms are assessed, and because the DTS does not explicitly refer to avoided conversation as an example of avoidance, we do not present the DTS as a diagnostic instrument. However, it appears to be useful as a symptoms-severity measure, as a predictor and measure of treatment response, and as a diagnostic ally, in that its diagnostic efficiency reached 83%. As designed, the scale likewise fulfilled standard criteria for reliability and validity in a broad range of adult trauma victims.
Diagnostic assessment using the DTS, relative to the SCID, yielded respectable accuracy. At a score of 40, the positive predictive value, negative predictive value, and efficiency of 0.92, 0.79 and 0.83 were comparable to values for the Penn Inventory of 0.84, 0.92 and 0.86 and for the Mississippi Scale of 0.88, 0.87 and 0.88 (Hammarberg, 1992). While diagnostic efficiency of these latter scales was slightly higher, their values were based upon a more restricted trauma sample.

Factor analysis of the sample as a whole revealed that the factor structure of the DTS did not correspond to the group clusters of DSM-IV (i.e. intrusive, avoidant/numbing and hyperarousal). A single factor accounted for much of the variance and may be construed as a severity measure or a higher-order organizing factor. A second, smaller factor emerged consistent with intrusive symptoms. In yielding primarily a single higher-order factor, the DTS is similar to the combat Mississippi Scale (Keane et al. 1988) and different from our earlier report of the SI-PTSD (Davidson et al. 1989) and the self-rated civilian version of the Mississippi Scale (Vreven et al. 1995). When factor analysed in the subsample of subjects with current PTSD, a general severity factor emerged, along with numbing (factor II), avoidance/insomnia and absence of intrusion (factors III and IV), and hyperarousal symptoms (factor V). These principal factors are more in keeping with the DSM formulation of symptom groupings in PTSD.

The DTS performed well as a measure detecting change over time during a controlled treatment trial, as well as yielding a statistically significant prediction of treatment outcome. The latter finding, that is, that more severely symptomatic PTSD responded less well to a pharmacotherapy trial, echoes an earlier finding with amitriptyline, in which we found that higher baseline Hamilton anxiety and depression scores (Hamilton, 1959, 1960) were associated with poorer response on the CGI (Davidson et al. 1993). Also of relevance was the correlation in that study between high baseline PTSD symptoms and poor outcome measured by the self-rated IES.

The DTS was validated mostly against scales which either predated or were modelled after DSM-III-R. In ideal circumstances, DSM-IV analogues would have been preferred, but they were unavailable at the time. We, nevertheless, believe that the DTS offers a useful new assessment option for the following reasons. The sample in which we have tested the DTS is perhaps the most comprehensive of all used to evaluate PTSD symptom–severity scales, with victims of both sexes drawn from four different trauma groups (rape, combat, hurricane and a category of miscellaneous other traumata). We also provided an assessment of the scale in a controlled treatment-trial setting. One of the main purposes of the DTS is to measure PTSD severity and to detect treatment change over time. We believe our results demonstrate promise in this regard. Moreover, in unpublished data from our own group, the DTS proved to be more sensitive at detecting drug versus placebo differences than the IES, CGI or SI-PTSD in a double-blind trial. We can also adduce construct validity for the DTS relative to a biological marker of PTSD, namely, rapid eye movement (REM) density, as measured by polysomnography in the hurricane survivors. Severity of intrusive symptoms correlated significantly with REM density in a subsample of Study 3 (Mellman et al. 1995).

Further work remains in relation to the DTS, including developing and applying it to other populations (e.g. children, non-English speakers), additional investigation of its factor structure, and assessment of its ability to detect differences between active medication and placebo.

REFERENCES


BACKGROUND In post-traumatic stress disorder (PTSD) there is a need for self-rating scales that are sensitive to treatment effects and have been tested in a broad range of trauma survivors. Separate measures of frequency and severity may also provide an advantage.


