Systematic Review of Screening Instruments for Adults at Risk of PTSD

Chris R. Brewin

The development of effective methods of screening for posttraumatic stress disorder (PTSD) is important in the context of mass trauma, the geographical dispersion of victims, and the restricted availability of specialists in psychological trauma. The review focused on published English-language screening instruments for civilian PTSD consisting of 30 items or fewer and validated against structured clinical interviews. Thirteen instruments were identified meeting these criteria, all consisting of symptoms of traumatic stress. The review concluded that the performance of some currently available instruments is near to their maximal potential effectiveness, and that instruments with fewer items, simpler response scales, and simpler scoring methods perform as well as if not better than longer and more complex measures.

Major traumatic events often pose special challenges to mental health services, either because of the numbers of individuals affected; because of their geographical dispersion; because of the priority that must be given to immediate medical and social care delivered by nurses, physicians, and social workers; because of the difficulty in following up those affected; or because of the unavailability of specialist trauma clinicians and services. These circumstances mean that there is often an important role to be played by screening instruments that can be used by non-trauma specialists to detect adverse psychological responses. This article reviews the characteristics and performance of those instruments that are currently available to screen for the presence of posttraumatic stress disorder (PTSD).

Although there are a wide range of possible adverse psychological reactions that follow trauma, little research has been carried out post trauma on conditions other than PTSD. Although trauma increases the risk of several disorders, including PTSD, depression, and substance abuse (Resick, 2001), there are practical difficulties in determining that the symptoms detected by posttrauma screening instruments are attributable to the specific traumatic event of interest rather than to more general adversity or to pre-existing conditions. Screening for PTSD has an important advantage in that the existence of intrusion and avoidance symptoms tied to the trauma makes a much stronger case for the causal role of the event in contributing to the disorder. Current estimates of the prevalence of PTSD in trauma populations are highly diverse but confirm that it is a relatively common response. Some illustrative figures range from 47% in rape victims 12 weeks post assault, to 12% in road traffic accident victims, to between 5% and 8% among victims of some natural disasters (Resick, 2001).

Unlike diagnostic measures such as structured clinical interviews, screening instruments need not include items corresponding to specific diagnostic criteria but may be based on any measure (e.g., demographic, biological, or self-report items) that successfully predicts the criterion diagnosis. In medical, social, or primary care contexts where there is little specific expertise in psychological...
trauma, there are obvious benefits to screening instruments that are short and easy to administer. They should ideally contain the minimal number of items necessary for accurate case identification and should preferably not require respondents to ponder over large numbers of alternative scale points. They should be written in language that is easy to understand. Their purpose should be plain, and they should be acceptable to respondents. For ease of administration self-report questionnaires would appear to be the most flexible solution. If they are to be scored by nonspecialists, a method that would widen their applicability, simple decision rules for determining who passes and fails the screen would be at a premium. Also highly desirable for successful instruments is that they be applicable to populations who have a varying prevalence of PTSD and are experiencing different traumas.

Over and above these considerations, the instrument must be effective in ruling in respondents who are cases and ruling out respondents who are not. The performance of a screening instrument for a condition such as PTSD is generally assessed by reference to several criteria, of which two are most commonly encountered: sensitivity, that is, the probability that someone who has a PTSD diagnosis will have had a positive test result, and specificity, that is, the probability that someone who does not have a PTSD diagnosis will have had a negative test result. A good test will have a reasonable balance of sensitivity and specificity. A test for PTSD can be made highly sensitive by setting a very low threshold, such as two reexperiencing symptoms present at any time in the past month, with the result that almost everyone who has a PTSD diagnosis will exceed this threshold. However, many people will exceed this threshold even though they do not have the disorder, with the result that the specificity of the test will be correspondingly low. In the same way a test for PTSD can be made highly specific by setting a very high threshold, such as five avoidance or numbing symptoms present during the past week, with the result that almost nobody who does not have a PTSD diagnosis will exceed this threshold. However, many people will fall short of this threshold even though they do have the disorder, with the result that the sensitivity of the test will be correspondingly low.

Other criteria for evaluating screening tests are relevant to two slightly different but highly practical questions: What is the probability that someone who has a positive test result will report a diagnosis of PTSD, and what is the probability that someone who has a negative test result will not receive a PTSD diagnosis? The answer to the first question is given by the positive predictive power of the screening test, and the answer to the latter by the negative predictive power of the test. Finally, the performance of a test can also be expressed in terms of the percentage of respondents correctly classified by the test as having or not having PTSD, which is referred to as its overall efficiency.

Whereas sensitivity and specificity are independent of the prevalence of the disorder in the population, and so can readily be compared across studies, positive and negative predictive power are sensitive to population prevalence. If there are very few cases to detect, the positive predictive power of the test will suffer, whereas if the vast majority of the population are affected, its negative predictive power will be correspondingly limited (Baldessarini, Finklestein, & Arana, 1983). In other words, at low prevalences a negative test result is more likely to be correct, whereas at high prevalences a positive result is more likely to be correct.

Baldessarini and associates (1983) comment that highly specific tests (those having a low rate of false positive findings), even with moderate sensitivity, are particularly useful when test results are positive and when the prevalence of the condition is high. These conditions might obtain in a specialist service, or in screening of a sample exposed to a severe stressor involving high levels of injury and death. In such conditions highly specific tests should help to identify individuals requiring more intensive assessment. Conversely, highly sensitive tests (those having a low false negative rate), even with moderate specificity, are particularly useful when test results are negative and when the prevalence of the condition is low. These conditions might obtain in screening a population after a less severe stressor in which there was little injury and death. In such conditions highly sensitive tests should be useful in excluding individuals from further assessment.

In practice, what levels of test performance is it realistic to expect? One guide as to an upper limit of performance can be derived from a comparison of diagnoses yielded by the two most highly regarded interview assessments currently available, the Structured Clinical Interview for DSM-IV (SCID) PTSD module (First, Spitzer, Gibbon, & Williams, 1995) and the Clinician Administered PTSD Scale (CAPS; Blake et al., 1995). In a sample of 123 combat veterans, a CAPS total score of 65 was found to have a sensitivity of .84 and a specificity of .95 relative to a SCID diagnosis (Blake et al., 1995).

The vast majority of screening measures are symptom-based. A few studies have adopted general measures of psychopathology such as the General Health Questionnaire or the Millon Clinical Multiaxial Inventory to screen for PTSD (Craig & Olson, 1997; Darves-Bornoz, Pierre, Lepine, Degiovanni, & Gaillard, 1998).
that specifically target trauma symptoms. In principle, an associated feature of PTSD such as guilt or known risk factors (see Brewin, Andrews, & Valentine, 2000; Ozer, Best, Lipsey, & Weiss, 2002, for reviews) could also form the basis of screening instruments or could be incorporated into symptom-based measures in an attempt to increase levels of performance. Some of the risk factors (e.g., family history of psychiatric disorder) are difficult to measure, however, whereas others are not applicable to all types of trauma (e.g., peritraumatic emotions) or involve soliciting sensitive information (e.g., previous trauma history). Nor has consideration been given to how information from risk factors, associated features, or symptoms could be effectively combined in a single instrument.

The purpose of the review is to determine the applicability of general screening instruments for PTSD to civilian trauma populations, the level of performance that can be expected from currently available instruments, and the potential that there is for improvement. The review will also examine whether the length of the instrument or the simplicity of response and scoring methods is related to performance. All the instruments described have been validated against structured interviews for PTSD and have published information about their diagnostic utility. To enable the reader to evaluate the available information, details of the prevalence of PTSD in the original sample studied and whether the instrument has been validated on an independent sample are included. This is particularly critical in those cases in which an optimal cutoff score has been determined post hoc from the existing data set, thus potentially inflating the diagnostic validity of the instrument. Where appropriate I also draw attention to those instruments that have been evaluated within the first year post trauma and that are therefore particularly relevant to the context of early intervention.

Method

Selection Criteria

Articles published in English in peer-reviewed journals in which the performance of screening instruments with the potential to detect current PTSD was compared with that of structured clinical interviews for PTSD and that reported information (such as sensitivity and specificity) about the instrument’s diagnostic validity were sought. Instruments had to have been used with adults and be relevant to any trauma population (i.e., not be worded in a way or rely on information specific to individual traumas, such as combat, injury, or earthquake). Those longer than 30 items were omitted as being likely to be too time-consuming to be of value in a screening context. Studies using the instruments were not included if they were limited to current members of the armed services or specific clinical populations such as the brain-injured or substance abusers.

Search Strategy

Three electronic databases were searched in December 2003 for articles meeting the specified criteria. Medline was searched by using the expressions “screen* and instrument* and PTSD,” yielding 31 articles, and “instrument* and valid* and diagnos* and PTSD,” yielding 50 articles. ISI Web of Science was searched by using the expressions “screen* and instrument* and PTSD,” yielding 56 articles, and “instrument* and valid* and diagnos* and PTSD,” yielding 49 articles. Pilots (a specialist trauma database managed by the National Center for PTSD in White River Junction, Vermont) was searched with the expression “screen$ DE (PTSD and ASSESSMENT INSTRUMENTS),” in which the free term “screen$” was combined with descriptors specified by the database. This search yielded 97 articles, including chapters and theses. In addition a hand search of books on trauma assessment (Carlson, 1997; Stamm, 1996; Wilson & Keane, 2003), of back issues of the Journal of Traumatic Stress, and of the reference lists of articles meeting study selection criteria was conducted. The search generated 19 articles meeting the selection criteria, containing a total of 22 relevant datasets. All the instruments consisted of posttraumatic symptoms.

Information Extracted

From the data sets the following information about each instrument was extracted and is reported in Table 1: how the presence of PTSD was determined by the instrument, including any cutoff score employed and whether this was determined post hoc; number of items; sample composition, including gender information, size, and prevalence of PTSD; sensitivity, specificity, positive and negative predictive power, and overall efficiency of the instrument. When these performance data were not all reported, they were calculated by the author from the information provided. Instruments are reported according to the date when the first published article meeting the criteria appeared.

Results

In all, 13 separate instruments, ranging from 4 to 30 items in length, were identified in the 22 data sets.
Table 1. Sensitivity, Specificity, and Power to Predict Posttraumatic Stress Disorder in Adults of Different Screening Instruments

<table>
<thead>
<tr>
<th>Authors</th>
<th>Instrument</th>
<th>Number of Items</th>
<th>Sample (Size)</th>
<th>Prevalence of PTSD</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive Predictive Power</th>
<th>Negative Predictive Power</th>
<th>Overall Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neal et al. (1994)</td>
<td>IES/cutoff 35</td>
<td>15</td>
<td>Mixed trauma (70)</td>
<td>51%</td>
<td>.89</td>
<td>.88</td>
<td>.89</td>
<td>.88</td>
<td>.89</td>
</tr>
<tr>
<td>Wohlfarth et al. (2003)</td>
<td>IES/cutoff 35</td>
<td>15</td>
<td>Crime victims (79)</td>
<td>13%</td>
<td>.89</td>
<td>.94</td>
<td>.67</td>
<td>.99</td>
<td>.94</td>
</tr>
<tr>
<td>Blanchard et al. (1996)</td>
<td>PCL-C/cutoff 44</td>
<td>17</td>
<td>Mixed trauma (40)</td>
<td>45%</td>
<td>.94</td>
<td>.86</td>
<td>.85</td>
<td>.95</td>
<td>.90</td>
</tr>
<tr>
<td>Andrykowski et al. (1998)</td>
<td>PCL-C/cutoff 50</td>
<td>17</td>
<td>Breast cancer (82)</td>
<td>6%</td>
<td>.60</td>
<td>.99</td>
<td>.75</td>
<td>.97</td>
<td>.96</td>
</tr>
<tr>
<td>Manne et al. (1998)</td>
<td>PCL-C</td>
<td>17</td>
<td>Mothers of cancer survivors (65)</td>
<td>6%</td>
<td>1.00</td>
<td>.92</td>
<td>.44</td>
<td>1.00</td>
<td>.92</td>
</tr>
<tr>
<td>Walker et al. (2002)</td>
<td>PCL-C/cutoff 30</td>
<td>17</td>
<td>HMO members (261)</td>
<td>11%</td>
<td>.82</td>
<td>.76</td>
<td>.30</td>
<td>.97</td>
<td>.77</td>
</tr>
<tr>
<td>Foa et al. (1997)</td>
<td>PDS</td>
<td>30</td>
<td>Mixed trauma (248)</td>
<td>52%</td>
<td>.89</td>
<td>.75</td>
<td>.79</td>
<td>.86</td>
<td>.82</td>
</tr>
<tr>
<td>Sheeran &amp; Zimmerman (2002)</td>
<td>PDS/cutoff 27</td>
<td>17</td>
<td>Psychiatric outpatients (774)</td>
<td>11%</td>
<td>.67</td>
<td>.91</td>
<td>.49</td>
<td>.96</td>
<td>.88</td>
</tr>
<tr>
<td>Wohlfarth et al. (2003)</td>
<td>PSS-SR</td>
<td>17</td>
<td>Crime victims (79)</td>
<td>13%</td>
<td>.90</td>
<td>.84</td>
<td>.45</td>
<td>.98</td>
<td>.85</td>
</tr>
<tr>
<td>Davidson et al. (1997a)</td>
<td>DTS/cutoff 40</td>
<td>17</td>
<td>Mixed trauma (129)</td>
<td>52%</td>
<td>.69</td>
<td>.95</td>
<td>.92</td>
<td>.79</td>
<td>.83</td>
</tr>
<tr>
<td>Meltzer-Brody et al. (2000)</td>
<td>SPAN/cutoff 5</td>
<td>4</td>
<td>Mixed trauma (121)</td>
<td>46%</td>
<td>.84</td>
<td>.91</td>
<td>.89</td>
<td>.87</td>
<td>.88</td>
</tr>
<tr>
<td>Meltzer-Brody et al. (2000)</td>
<td>SPAN/cutoff 5</td>
<td>4</td>
<td>Mixed trauma (122)</td>
<td>51%</td>
<td>.77</td>
<td>.82</td>
<td>.81</td>
<td>.78</td>
<td>.80</td>
</tr>
<tr>
<td>Carlier et al. (1998)</td>
<td>SRS-PTSD</td>
<td>17</td>
<td>Airplane crash survivors (136)</td>
<td>26%</td>
<td>.83</td>
<td>.80</td>
<td>.60</td>
<td>.93</td>
<td>.80</td>
</tr>
<tr>
<td>Fullerton et al. (2000)</td>
<td>BPTSD-6/cutoff 4</td>
<td>6</td>
<td>MVA victims (122)</td>
<td>50%</td>
<td>.77</td>
<td>.87</td>
<td>.75</td>
<td>.88</td>
<td>.84</td>
</tr>
<tr>
<td>Carlson (2001)</td>
<td>SPTSS/cutoff 4.0</td>
<td>17</td>
<td>Psychiatric inpatients (114)</td>
<td>74%</td>
<td>.94</td>
<td>.60</td>
<td>.87</td>
<td>.78</td>
<td>.85</td>
</tr>
<tr>
<td>Cross &amp; McCane (2001)</td>
<td>PTSD-Q/cutoff 60</td>
<td>17</td>
<td>College students (76)</td>
<td>41%</td>
<td>.81</td>
<td>.82</td>
<td>.76</td>
<td>.86</td>
<td>.82</td>
</tr>
<tr>
<td>Scragg et al. (2001)</td>
<td>Pennclinic 35</td>
<td>26</td>
<td>Trauma clinic referrals (80)</td>
<td>75%</td>
<td>.90</td>
<td>.60</td>
<td>.87</td>
<td>.66</td>
<td>.82</td>
</tr>
<tr>
<td>Brewin et al. (2002)</td>
<td>TSQ/cutoff 6</td>
<td>10</td>
<td>Train crash survivors (41)</td>
<td>34%</td>
<td>.86</td>
<td>.93</td>
<td>.86</td>
<td>.93</td>
<td>.90</td>
</tr>
<tr>
<td>Brewin et al. (2002)</td>
<td>TSQ/cutoff 6</td>
<td>10</td>
<td>Crime victims (157)</td>
<td>27%</td>
<td>.76</td>
<td>.97</td>
<td>.91</td>
<td>.92</td>
<td>.92</td>
</tr>
<tr>
<td>Chou et al. (2003)</td>
<td>DRPST/cutoff 3</td>
<td>7</td>
<td>Earthquake victims (461)</td>
<td>8%</td>
<td>.98</td>
<td>.97</td>
<td>.76</td>
<td>1.00</td>
<td>.97</td>
</tr>
<tr>
<td>Hovens et al. (2002)</td>
<td>SRIP/cutoff 52</td>
<td>22</td>
<td>Mixed trauma (76)</td>
<td>54%</td>
<td>.86</td>
<td>.71</td>
<td>.78</td>
<td>.81</td>
<td>.78</td>
</tr>
<tr>
<td>van Zelst et al. (2003)</td>
<td>SRIP/cutoff 39</td>
<td>22</td>
<td>Older adults (422)</td>
<td>3%</td>
<td>.74</td>
<td>.81</td>
<td>.07</td>
<td>.99</td>
<td>.90</td>
</tr>
<tr>
<td>Mean performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. PTSD = posttraumatic stress disorder; MVA = motor vehicle accident; HMO = health maintenance organization; IES = Impact of Event Scale; PCL-C = PTSD Checklist (Civilian Version); PDS = Posttraumatic Stress Diagnostic Scale; PSS-SR = Posttraumatic Stress Symptom Scale (Self-Report Version); DTS = Davidson Trauma Scale; SRS-PTSD = Self-Rating Scale for Posttraumatic Stress Disorder; SPAN = Startle, Physiological Arousal, Anger, and Numhness items from the Davidson Trauma Scale; SPTSS = Screen for Posttraumatic Stress Symptoms; PTSD-Q = Posttraumatic Stress Disorder Questionnaire; Penn = Penn Inventory for Posttraumatic Stress Disorder; TSQ = Trauma Screening Questionnaire; DRPST = Disaster-Related Psychological Screening Test; SRIP = Self-Rating Inventory for Posttraumatic Stress Disorder.

aCutoff determined post hoc.
bAll or almost-all-female sample.
cScored by using symptom cluster method.
dAdministered within 1 year of trauma.
eInformation supplied by authors.
Impact of Event Scale (IES)

The original 15-item version of the IES (Horowitz, Wilner, & Alvarez, 1979) contains questions about intrusion and avoidance relative to a specified event that are answered on a 4-point scale. Although it predates the introduction of the PTSD diagnosis, the scale has the advantage of being very widely used as a measure of traumatic stress. Horowitz and colleagues (1979) suggested a cutoff of 19 as indicative of a high level of clinical concern, but Neal and coworkers (1994) found empirically that the best screening performance was obtained with a cutoff score of 35. Wohlfarth, van den Brink, Winkel, and ter Smitten (2003) in their study of crime victims found perfect sensitivity but much lower levels of specificity when using a cutoff score of 19. Overall diagnostic performance improved from 83% to 94% when using 35 as the cutoff.

PTSD Checklist—Civilian Version (PCL-C)

This measure, developed by Weathers and associates (1991), requires respondents to indicate the degree to which they have been bothered during the past month by the 17 PTSD symptoms indicated in the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV). Symptoms are related to a particular traumatic event, and respondents answer 17 questions about the nature of the stressor, corresponding to Criterion A. They then answer 17 questions about the frequency in the past month of the symptoms mentioned in Criteria B, C, and D, using a 4-point scale ranging from 0 (not at all) to 5 (extremely). A diagnosis of PTSD can be based on endorsement (a symptom rating of at least 3) of at least one reexperiencing symptom, three avoidance symptoms, and two arousal symptoms (the symptom cluster method of scoring). Alternatively, a cutoff score of 50 has been proposed to suggest a PTSD diagnosis (Weathers & Ford, 1996).

More studies have used the PCL-C as a screening instrument than any other, albeit almost exclusively with women. Blanchard, Jones-Alexander, Buckley, & Forneris (1996) examined the diagnostic efficiency of the PCL-C with adults who had been involved in accidents or sexual assaults. When using the suggested cutoff score of 50 the performance of the PCL-C was good, but reducing the cutoff to 44 resulted in still better performance (see Table 1 for details). Andrykowski, Cordova, Studts, and Miller (1998) used the PCL-C with women who had breast cancer and Manne, Du Hamel, Galleli, Sorgen, and Redd (1998) with the mothers of pediatric cancer patients. Both studies had a very low prevalence of PTSD, which reduced the positive predictive power somewhat, but negative predictive power and overall efficiency were high. Studies employing the PCL-C with less highly selected female samples (e.g., Walker, Newman, Dobie, Ciechanowski, and Katon, 2002) have generally found that cutoff scores considerably lower than 50 yield optimal performance.

Posttraumatic Stress Symptom Scale—Self-Report Version (PSS-SR) and Posttraumatic Diagnostic Scale (PDS)

The development of the PSS-SR was reported by Foa, Riggs, Dancu, and Rothbaum (1993). Its 17 items correspond to the symptoms of PTSD defined in the DSM-III-R and are answered on a 4-point scale. Foa, Cashman, Jaycox, and Perry (1997) revised the PSS-SR in accordance with DSM-IV. The PDS includes 12 preliminary items inquiring about the occurrence of specific traumatic experiences. After nominating the event that has disturbed them most in the past month, respondents answer four questions about the nature of the stressor, corresponding to Criterion A. They then answer 17 questions about the frequency in the past month of the symptoms mentioned in Criteria B, C, and D, using a 4-point scale ranging from 0 (not at all or only one time) to 3 (five or more times a week/almost always). Finally, nine questions assess impairment. A diagnosis of PTSD is based on the presence of a trauma satisfying Criterion A, endorsement (rating of 1 or higher) of at least one reexperiencing symptom, three avoidance symptoms, and two arousal symptoms; duration of at least 1 month, and impairment in at least one area of functioning. Foa and associates (1997) reported a validation study of a PDS-based diagnosis against a SCID (interview-based) diagnosis and found the PDS to have high sensitivity. Subsequently Sheeran and Zimmerman (2002) employed the PDS with a sample of psychiatric outpatients. Restricting their analysis to the 17 symptom items, they found overall performance was maximized with a cutoff of 27, but this method resulted in much lower sensitivity and higher specificity than had previously been reported by Foa and colleagues (1997). Using the symptom cluster method of scoring made little difference to performance. Finally, Wohlfarth and coworkers (2003) reported the results of using the original PSS-SR to screen for PTSD in crime victims. When using the symptom cluster method of scoring, sensitivity was higher than specificity, as found by Foa and associates (1997).

Davidson Trauma Scale (DTS)

Davidson and colleagues (1997a) reported the development of this PTSD symptom scale based on DSM-IV. The DTS consists of 17 items corresponding to each of the DSM-IV symptoms. For each item, respondents rate both frequency and severity during the previous week on
0–4 scales, yielding a total score of 136. Davidson and coworkers (1997a) reported a validation study of the DTS against SCID diagnoses. When using a cutoff score of 40 determined post hoc, the DTS demonstrated high specificity and positive predictive power, with rather lower sensitivity and negative predictive power.

**SPAN**

This 4-item measure was derived from the DTS by Meltzer-Brody, Churchill, and Davidson (1999). A mixed group of 243 patients completed the DTS and were administered a structured diagnostic interview. The sample was divided into two, so that results from the first subsample could be replicated on the second. Good discrimination was obtained by using severity scores from the items measuring startle, physiological arousal to trauma cues, anger, and numbness (from which the name of the scale was derived). In the replication sample, which used the same severity score of 5 as a cutoff, screening performance compared favorably with that of the full DTS, although the sensitivity of the SPAN was higher and its specificity lower than that of the DTS. Similar results were obtained when a PTSD prevalence of 10% was assumed.

**Self-Rating Scale for Posttraumatic Stress Disorder (SRS-PTSD)**

This measure was designed by Carlier, Lamberts, Van Uchelen, and Gersons (1998) to be an abridged version of Davidson, Smith, and Kudler’s (1989) Structured Interview for PTSD, adapted to meet DSMIII-R criteria. It contains 17 items corresponding to the DSM-III-R symptoms of PTSD. Respondents rate the severity of each symptom on a 3-point scale anchored with 0 (not at all) and 2 (very much). Symptoms are rated present if the item is scored 1 or higher, or in some cases 2 or higher (see Davidson et al., 1989). In a sample of air crash survivors the SRS-PTSD, scored by the symptom cluster method, demonstrated a good balance between sensitivity and specificity, with high levels of both.

**Brief DSMPTSD-III-R and DSMPTSD-IV (BPTSD-6)**

The DSMPTSD-III-R (Ursano, Fullerton, & Kao, 1995) and the DSMPTSD-IV (Fullerton et al., 2000) are questionnaires developed to measure PTSD in community samples. They are based on the IES and the Symptom Checklist (SCL-90-R: Derogatis, 1983) as core instruments, supplemented by 12 PTSD-specific items. These 12 items are scored in accordance with the SCL-90-R on a 5-point scale ranging from 0 (not at all) to 4 (extremely). Fullerton and associates (2000) reported on the use of these 12 items alone in screening for PTSD. Both 12-item (BPTSD-12) and 6-item (BPTSD-6) versions were used, with similar results. For example, using a cutoff score of at least 4 on the BPTSD-6 yielded overall efficiency of 84%.

**Screen for Posttraumatic Stress Symptoms (SPTSS)**

This was developed by Carlson (2001) to provide an instrument that was not tied to a single traumatic event (e.g., the reexperiencing item is worded, “I suddenly feel like I am back in the past, in a bad situation that I was once in, and it’s like it was happening all over again”). The 17 items are matched to DSM-IV symptoms and are answered on an 11-point scale anchored with 0 (never) and 10 (always). When using the recommended post-hoc cutoff score of 4.0, the measure had high sensitivity but rather low specificity. The symptom cluster scoring method, with a score of 5 indicating the presence of a symptom, yielded comparable results.

**Posttraumatic Stress Disorder Questionnaire (PTSD-Q)**

Cross and McCanne (2001) updated the PTSD Interview (Watson, Juba, Manifold, Kucala, & Anderson, 1991) and converted the questions to an interview format. The 17 items corresponding to DSM-IV criteria are answered on a 7-point scale anchored with 1 (no, never) and 7 (extremely, always). A response of 4 (somewhat, commonly) qualifies as symptom endorsement. In a sample of college women the PTSD-Q demonstrated a good balance between sensitivity and specificity, with high levels of both. A cutoff of 60 was superior to scoring with the symptom cluster method.

**Penn Inventory**

Each item in the Penn Inventory (Hammarberg, 1992) comprises four sentences, modeled on the Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). The meanings of the series of sentences measure the presence or absence of PTSD symptoms as well as their degree, frequency, or intensity. Respondents choose one sentence per item, and each sentence contributes a score from 0 to 3 the total. Hammarberg (1992) recommends a cutoff score of 35 to indicate the presence of PTSD. In their sample of referrals to a trauma clinic
with a high base rate of PTSD, Scragg, Grey, Lee, Young, and Turner (2001) found the Penn Inventory to have high sensitivity but low specificity. They suggested that the instrument might be better suited to screening populations with a low prevalence of disorder, in which its high sensitivity would be an advantage, rather than as a diagnostic tool for specific traumatic stress symptoms.

**Trauma Screening Questionnaire (TSQ)**

The measure consists of the 10 reexperiencing and arousal items from the PSS-SR (Foa et al., 1993), modified to provide only two response options. Respondents indicate whether or not they have experienced each symptom at least twice in the past week. A study of train crash survivors demonstrated that a post-hoc criterion of endorsing any six (or more) symptoms yielded high levels of sensitivity and specificity compared to those of the CAPS diagnostic interview. This criterion was validated against a diagnosis derived from the PSS in a separate sample of crime victims, in which overall performance was similar: sensitivity was somewhat lower and specificity somewhat higher (Brewin et al., 2002).

**Disaster-Related Psychological Screening Test (DRPST)**

The original version of the instrument contained 17 items corresponding to PTSD symptoms and 9 items corresponding to symptoms of major depression. Symptoms are endorsed as present or absent and are not rated for duration or severity. Chou and colleagues (2003) described the empirical derivation of a 7-item screening scale based on a study of earthquake survivors that produced the best prediction of PTSD. This consisted of three reexperiencing symptoms, three arousal symptoms, and one arousal symptom. A cutoff score of 3, determined post hoc, yielded very high levels of sensitivity and specificity.

**Self-Rating Inventory for Posttraumatic Stress Disorder (SRIP)**

The original version of the scale consisted of 52 items based on DSM-III-R criteria and including items measuring Disorders of Extreme Stress Not Otherwise Specified. The most recent version of the scale (Hovens, Bramsen, & van der Ploeg, 2002) consists of 22 items based on DSM-IV symptoms and rated on a 4-point intensity scale anchored with 1 (*not at all*) and 4 (*very much*). Hovens and colleagues proposed a cutoff score of 52 based on a sample of trauma victims, most of whom were survivors of World War II and of the Dutch–Indonesian conflict. This yielded good sensitivity but rather lower levels of specificity. A subsequent study with a community sample of older adults (van Zelst et al., 2003) found low sensitivity at this cutoff and optimal performance with a cutoff score of 39.

**Discussion**

The mean diagnostic efficiency of the screening instruments reviewed was 86.5%. The performance of several approached the level of agreement between the SCID and the CAPS that has previously been obtained, although most of the instruments that achieved such high performance did so by changing cutoff scores post hoc or having a very low prevalence rate or in both ways. Two instruments (the IES and TSQ) consistently performed well and were the only ones that both had been validated on independent samples and had been tested within 1 year of a traumatic event. Although the number of studies conducted with any one instrument was small, these figures suggest that an efficiency ceiling has been reached. The first conclusions of the review are, therefore, that screening using a small number of core symptoms is potentially highly effective in a wide variety of trauma populations and that significant further gains are unlikely to be achieved by incorporating other risk factors or symptoms into the measures. The other conclusion from the review is that measures with fewer items, simpler response scales, and simpler methods of scoring perform as well as if not better than longer measures requiring more complex ratings. In particular, measures that were based on the 17 DSM-IV symptoms or that used the symptom threshold scoring method were not noticeably superior to other instruments.

On average the overall efficiency of instruments based on the 17 DSM-IV symptoms was around 85%, although a higher figure was sometimes achieved by changing cutoff scores post hoc or by having a very low prevalence rate. The instrument of this type that has been studied most is the PTSD Checklist (PCL-C), which has been tested across several different (predominantly female) populations and with several different prevalence rates. Performance has been quite variable; the best results have been obtained in homogeneous trauma samples. In less selected samples, lower cutoff scores have been found to be necessary, and performance has been weaker. Other promising instruments in this group are the PSS-SR and its successor, the PDS. When using the symptom cluster scoring method, this instrument has high sensitivity but somewhat lower specificity, perhaps because it
Investigators (e.g., Brewin, 2003) have proposed that clinical debriefing, to as many trauma survivors as possible in the immediate aftermath of the events. These interventions involve encouraging individuals to talk about the details of their experience in a supportive group or individual context, guided by appropriately trained professionals. Brief early interventions, although often popular with recipients who choose them, may suffer from low take up (Kadet, 2002) and are not associated with objectively better mental health outcomes (Bisson, 2003; Rose, Bisson, & Wessely, 2002; van Emmerik, Kamphuis, Huulsbosch, & Emmelkamp, 2002). In light of this evidence, several investigators (e.g., Brewin, 2003) have proposed that clinicians should refrain from providing symptom-oriented early intervention in favor of monitoring survivors for psychological disorder and intervening more intensively, using empirically validated treatments, should a significant clinical condition develop. The success of such a public health strategy depends on the existence of empirically validated screening instruments.

Among the most important limitations of the screening literature are the lack of replication and of systematic investigation of the properties of a particular instrument, whether with different trauma populations, different ethnic groups, different trauma prevalences, or different times post trauma. In particular, nothing is known about the impact of gender and cultural background on screening performance. Relatively few studies have been conducted within 6 months of a traumatic event, and so the relevance of these instruments to the context of early intervention is unclear. Immediate assessment is probably inadvisable, as the level of symptoms immediately post trauma is not a good predictor of the risk of later disorder (Greenberg, 1995; Rothbaum & Foa, 1993; Shalev, 1992), and it has been suggested that screening should take place at least 1 month post trauma (Brewin, 2003). Other critical issues are the acceptability and uptake of such instruments. There is evidence that individuals who are difficult to contact and assess may have a particularly high rate of PTSD, as Weisaeth (1989) found in his study of survivors of a factory explosion and fire.

There are as yet no published screening instruments that have been shown to meet all the criteria put forward at the beginning of this article, but it is likely that the next few years will confirm that existing measures do offer good predictive power across a range of trauma populations and prevalence rates. The evidence reviewed is encouraging inasmuch as symptom-based measures are already approaching a threshold of efficiency. No one instrument appears to be uniquely sensitive or specific, and thus in screening situations requiring greater sensitivity or greater specificity it would seem appropriate to adjust the thresholds of existing measures down or up, respectively, rather than use different measures. Brief questionnaire instruments that have 10 or fewer items and simple yes/no response scales appear to perform at a level similar to that of much longer instruments. As such instruments are generally easier to understand, complete, and score, their use should yield considerable benefit. These findings promise to be extremely useful in helping to monitor victims in the early months post trauma and identifying who requires more intensive intervention. Simple screening tools should improve the recognition of the disorder in a variety of nonspecialist settings. To give just one example, they would facilitate the development in the near future of early intervention programs that involve encouraging individuals to talk about the details of their experience in a supportive group or individual context, guided by appropriately trained professionals.
future of computer-based screening in primary care, with links to self-help materials and Internet-based treatment packages.

Acknowledgment

This article is based on a report presented at a Roundtable Meeting on Assessment and Screening Following Disasters and Mass Violence jointly organized by the National Institute of Mental Health, the Substance Abuse and Mental Health Services Administration, and the Department of Veterans Affairs National Center for PTSD (Bethesda, Maryland, August 26–27, 2003).

References