ELSEVIER

Psychiatry Research 88 (1999) 63–70

PSYCHIATRY RESEARCH

www.elsevier.com/locate/psychres

Derivation of the SPAN, a brief diagnostic screening test for post-traumatic stress disorder

Samantha Meltzer-Brody[b], E. Churchill[b], Jonathan R.T. Davidson[a,*]

[a] *Department of Psychiatry, Duke University Medical Center, Anxiety and Traumatic Stress Program, Box 3812, Durham, NC 27710, USA*

[b] *Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC 27710, USA*

Received 20 January 1999; received in revised form 6 July 1999; accepted 4 August 1999

Abstract

The Davidson Trauma Scale (DTS) is a validated 17-item self-rating scale used in the diagnosis of post-traumatic stress disorder (PTSD), which is sensitive to the effects of treatment. It was felt that a shorter version of the scale might provide a better diagnostic screening tool. Subjects were drawn from a sample of 243 patients obtained from multiple cohorts that included a group of survivors of various forms of trauma, including natural disaster, rape and combat. All subjects had diagnostic assessments for PTSD with a clinical interview and completed the DTS. The data were randomly divided between two subsamples, and frequency and severity scores were calculated for the DTS. A four-item scale, the SPAN (named for its top four items: Startle, Physiological arousal, Anger, and Numbness), was developed. It demonstrated an efficiency of 0.88, sensitivity of 0.84, specificity of 0.91 and positive likelihood ratio of 9.1. In a replication sample, values were slightly lower but still acceptable (efficiency = 0.80). A subgroup of PTSD patients received either fluoxetine or placebo in a clinical trial, and a significant SPAN score improvement was observed on fluoxetine. The SPAN, which correlated significantly with the Impact of Events Scale, the Sheehan Disability Scale, and the Structured Interview of PTSD, was found to have a diagnostic accuracy of 88%.

Keywords: Trauma; Davidson Trauma Scale; Scale development; Psychometrics; PTSD

* Corresponding author. Tel.: +1-919-684-2880; fax: +1-919-684-8866.

0165-1781/99/$ - see front matter
PII: S0165-1781(99)00070-0

1. Introduction

Post-traumatic stress disorder (PTSD) is a major mental health issue having a lifetime prevalence in the US of between 8 and 12%. It frequently becomes a chronic problem, and there are substantial rates of psychiatric and medical comorbidity found in association with PTSD (Kessler et al., 1995). Since PTSD was first introduced into the DSM-III (American Psychiatric Association, 1980), the need for accurate diagnosis of PTSD has led to the development of instruments for measuring the disorder.

The structured clinical interview has been regarded as the gold standard of PTSD measurement (Gerardi et al., 1989). These interviews include the Structured Clinical Interview for DSM-III-R (Spitzer et al., 1990), the Structured Interview PTSD measure (SIP) (Davidson et al., 1997a), and the Clinician Administered PTSD Scale (CAPS) (Blake et al., 1995). All have demonstrated good psychometric validity and reliability criteria and have been used extensively in treatment outcome studies.

However, a major drawback of structured clinical interviews is that they are time consuming, and thus, briefer self-rating scales may be viewed as having potential clinical utility. A number of self-rating scales have been developed to assess the symptoms of PTSD while economizing on time. These include the Impact of Events Scale (IES) (Horowitz et al., 1979), the Mississippi Scales for Combat and Non-combat (Keane et al., 1988), the Short Form of the Mississippi Scale (Fontana and Rosenheck, 1994), the Penn Inventory (Hammarberg, 1992), the PTSD Checklist (PCL) (Weathers et al., 1993), the Post-traumatic Stress Diagnostic Scale (PDS) (Foa, 1995), and the Davidson Trauma Scale (DTS) (Davidson et al., 1997b).

Given the abundance of instruments and a rapidly expanding empirical literature, researchers and clinicians can now choose instruments tailored to their particular assessment needs. In addition, they can increase their confidence in assessment decisions by relying on converging information obtained from multiple measures in an assessment battery (Keane et al., 1987). In particular, the use of a structured clinical interview as well as self-rating scales has become widely accepted as a way of obtaining diagnosis and/or assessing severity and clinical significance of symptoms.

The Davidson Trauma Scale (DTS) is a validated 17-item self-rating scale of frequency and severity of each symptom, which is sensitive to the effects of treatment (Davidson et al., 1997b). The DTS reflects the symptoms diagnostic of PTSD as defined in DSM-IV (American Psychiatric Association, 1994). Major strengths of the DTS include its development in a broad population of men and women exposed to different types of trauma, its sensitivity to treatment-induced change across time, and its capability of distinguishing between treatments of differing effectiveness and its ability to predict later treatment response. The full DTS has been shown to distinguish between those with, and those without, a diagnosis of PTSD at a cutoff score of 40, with an efficiency of 0.83.

Since the 17 items of the DTS demonstrate a high level of intercorrelation with one another (Cronbach's alpha coefficient $= 0.90$), we believed that it might be possible to develop a much shorter version of the scale to serve as a diagnostic screening tool. The newly derived four-item scale (SPAN) forms the subject of this report.

2. Methods

2.1. Subjects

Item selection for the SPAN was drawn from a sample of 243 patients obtained from multiple cohorts which comprised both pharmacotherapy trials of lamotrigine ($n = 14$) (Hertzberg et al., unpublished); fluvoxamine ($n = 12$) (Davidson et al., 1998a); nefazodone ($n = 12$) (Davidson et al., 1998b); and fluoxetine vs. placebo ($n = 55$) (Connor et al., 1999); a family study of rape trauma ($n = 74$) (Davidson et al., 1998c); an evaluation of Hurricane Andrew survivors ($n = 53$) (Davidson et al., 1997b); and a group of combat veterans who were undergoing psychiatric evaluation ($n = 23$) (Davidson et al., 1997b). Drug trial subjects were assessed at study termination, i.e. where

some still had PTSD, while others who had recovered did not exhibit the diagnosis.

All subjects in the above studies had diagnostic assessments for PTSD either with the SCID or SIP. In addition, all subjects completed the 17-item DTS as part of their assessment battery. The DTS measures DSM-IV PTSD symptoms on a 0–4 range for frequency and a 0–4 range for severity, with the maximum possible total score being 136. The treatment outcome data were based upon an endpoint analysis of all treated patients who entered the trials.

The patients in the pharmacotherapy trial of fluoxetine vs. placebo ($n = 55$) were independently evaluated in order to assess the ability of the SPAN to measure treatment outcome. A general linear model (GLM) analysis of variance (ANOVA), with treatment group as the main effect and time as a repeated measure, was applied to answer this question.

2.2. Scale construction

From our sample of 243 patients, the data were randomly divided between two subsamples, which were identical in demographics and other characteristics. In our first 50% subsample, the derivation sample, we initially established which DTS items differed most between PTSD and non-PTSD groups by ranking the mean differences in item frequency score, severity score, and frequency plus severity scores between PTSD and non-PTSD groups. We then prospectively tested its performance in the second, remaining subsample. In these two sets of analyses, the prevalence of PTSD in our population was approximately 50%. We then re-ran the analyses using an assumed PTSD prevalence of 10% in order to determine how successful the scale would be when the PTSD prevalence was significantly less than a 50:50 split of PTSD to non-PTSD patients.

Since we wanted a priori to develop a brief scale of no longer than seven items, we proceeded sequentially through the ranking by taking the top three through top seven items of the DTS and calculated sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and diagnostic accuracy measures for each total score. We conducted this procedure separately for item frequency, severity and frequency plus severity scores. Thus, for each set of items (e.g. the top four ranked items), we examined the performance of different score thresholds (e.g. a total score of 1–16), using the measures described.

Likelihood ratios for the test results were also calculated. Likelihood ratio is the ratio of the likelihood of the disease being present given a positive test result, to the likelihood of no disease being present given a positive test result. It provides an index of the increased likelihood that a disease is present given a positive test result and is calculated by sensitivity/(1-specificity) (Rampes et al., 1998). A positive likelihood ratio (PLR) score of 3 is considered moderately positive and a score of 10 is considered strongly positive. Negative likelihood ratios (NLR) may also be calculated for a negative test result A negative likelihood ratio score of 0.1 is considered strongly negative, and a score of 1 is considered neutral (Rampes et al., 1998). We determined the optimum algorithm to be that associated with the highest diagnostic accuracy (or efficiency) score based upon ROC characteristics. The above algorithm established in the derivation sample was then prospectively applied to the remaining 50% of subjects, i.e. the replication sample.

Although the primary purpose of the SPAN is to serve as a brief, self-rated, diagnostic screen, we took advantage of the treatment outcome data available to us and examined their performance in a drug (fluoxetine) vs. placebo controlled trial (Connor et al., 1999).

Additionally, the SPAN was compared to other measures of PTSD including the DTS, Impact of Events Scale, Sheehan Disability Scale and SIP by using the Pearson correlation coefficient.

3. Results

A total of 243 patients were included in the analysis. The average age was 37 years (S.D. = 11.6). Of the total group, 72% were women, 28% were men, 77.6% were Caucasian and 17.4% were African–American, and 5% were of other ethnic

groups. All subjects were evaluated for a diagnosis of PTSD. Of the subjects 118 (48.6%) had a diagnosis of PTSD, while 125 subjects (51.4%) had no diagnosis of PTSD. The demographic data were calculated for both derivation and replication groups as shown in Table 1.

Evaluations of frequency, severity and combined scores showed that the severity scores provided the greatest discrimination and yielded the highest efficiency. In the interest of space, we present only the severity score analyses (other data available upon request). Table 2 presents symptoms ranked according to the calculated differences in severity scores between subjects with and without a PTSD diagnosis.

The optimal efficiency scores for each solution are shown in Table 3 and are based on the particular item set cutoff score that yielded highest efficiency. We found that the four-item solution at a cutoff score of 5 gave an efficiency of 0.88, whereas the five-item solution at a cutoff score of 4 had an efficiency of 0.86. Thus, we took the shorter of the two, since our interest was to develop a brief screening tool. Next, we present the different sensitivity analyses for each score threshold with those four items (Table 4).

These top four items consist of items 17 (startle), 14 (anger), 11 (numbness) and 5 (physically upset at exposure to reminders of the trauma) (scale available upon request). Each item has a

Table 1
Demographics

	Entire sample	Derivation sample	Replication sample
Number (*n*)	243	121	122
Female (%)	71.6	74	69
Caucasian (%)	77.6	78.5	76.7
African–American (%)	17.4	15.7	19.2
Mean age	37.0	36.9	37.1
Positive PTSD (%)	48.6	46.3	50.8

Table 2
Rank of differences in severity scores between subjects with and without PTSD

DTS item no.	Item description	Positive PTSD	Negative PTSD	Difference
17	Startle	2.36	0.37	1.99
14	Anger	2.46	0.55	1.91
11	Numbness	2.29	0.45	1.84
5	Physiological upset	2.14	0.37	1.77
3	Flashbacks	1.86	0.11	1.75
6	Avoiding thoughts/feelings	2.11	0.37	1.74
1	Painful memories	2.30	0.62	1.69
15	Poor concentration	2.29	0.60	1.69
9	Anhedonia	2.27	0.65	1.62
16	Hypervigilance	2.11	0.49	1.61
10	Estrangement	1.89	0.29	1.60
4	Emotional upset	2.23	0.66	1.57
2	Nightmares	1.71	0.17	1.55
12	Shortened lifespan	2.41	0.89	1.52
8	Amnesia	2.17	0.74	1.44
7	Avoiding activating feelings	1.54	0.12	1.41
13	Insomnia	2.23	0.86	1.37

Table 3
Optimal efficiency scores

No. of severity items	Severity score	Maximum efficiency
Top 3	4	0.843
Top 4	5	0.876
Top 5	4	0.860
Top 6	5	0.868
Top 7	6	0.868

score of 0–4 in terms of severity. In the first 50% subsample ($N = 121$), it was found that a total score of 5 or more of the items gave the greatest likelihood of having PTSD, with a sensitivity of 84%, specificity of 91%, maximum efficiency of 88%, PPV of 89% and NPV of 87%. The positive likelihood ratio was 9.1.

In the replication subsample ($N = 122$), again using a score of 5 as cutoff, sensitivity was 77%, specificity was 82%, efficiency was 80%, PPV was 81%, and NPV was 78%. The positive likelihood ratio was 4.3. Chi-square analysis of the replication subsample demonstrated that 81% of the subjects were correctly diagnosed with PTSD when the four-item DTS was compared to a clinical interview diagnosis. ($\chi^2 = 42.6$, d.f. = 1, $P < 0.001$).

When a 10% prevalence rate of PTSD was assumed, we again examined the performance of different score thresholds and found the maximum efficiency to be 86% for a four-item solution, 83% for a five-item solution, and 83% for the seven-item solution.

One subgroup of the original cohort, the fluoxetine vs. placebo patients ($N = 55$), was analyzed separately in order to look at the ability of the SPAN to measure treatment outcome. In the pre-treatment group, the mean SPAN score was 9.70 (S.D. = 3.43) in the placebo group ($N = 24$) and 8.88 (S.D. = 3.09) in the fluoxetine group ($N = 26$). At post-treatment, the mean SPAN score was 6.83 (S.D. = 4.72) in the placebo group and 2.35 (S.D. = 3.22) in the fluoxetine group. When a repeated measures ANOVA was applied

Table 4
Test performance of different severity scores for the four-item scale

Severity score	Sensitivity	Specificity	Efficiency	PPV	NPV
0	1.000	0.000	0.463	0.463	0.000
1	0.982	0.431	0.686	0.598	0.966
2	0.964	0.600	0.769	0.675	0.951
3	0.875	0.738	0.802	0.742	0.873
4	0.875	0.846	0.860	0.831	0.887
5	0.839	0.908	0.876	0.887	0.868
6	0.768	0.908	0.843	0.878	0.819
7	0.643	0.938	0.802	0.900	0.753
8	0.625	0.969	0.810	0.946	0.750
9	0.589	0.969	0.793	0.943	0.733
10	0.554	0.985	0.785	0.969	0.719
11	0.464	0.985	0.744	0.963	0.681
12	0.339	0.985	0.686	0.950	0.634
13	0.232	1.000	0.645	1.000	0.602
14	0.196	1.000	0.628	1.000	0.591
15	0.179	1.000	0.620	1.000	0.586
16	0.125	1.000	0.595	1.000	0.570

with treatment group as the main effect and time as a repeated measure, the SPAN was able to differentiate between the fluoxetine vs. placebo group ($F = 15.64$, d.f. $= 1$, $P < 0.005$).

The SPAN also correlates well with other measures of PTSD. A Pearson correlation coefficient was calculated for the SPAN compared with the DTS ($r = 0.96$, $P < 0.0001$), the Impact of Events Scale ($r = 0.85$, $P < 0.0001$), the Sheehan Disability Scale ($r = 0.87$, $P < 0.0001$) and the SIP ($r = 0.86$, $P < 0.0001$).

4. Discussion

We found that a short four-item version of the DTS closely corresponded to the diagnosis of PTSD by structured clinical interview and believe, therefore, that it could be effectively used to screen for the diagnosis. We refer to this scale as the SPAN, a readily usable acronym, which conveys its content in the following way: Startle, Physiological arousal, Anger and Numbness. Based upon SPAN scores, appropriate subjects could then receive a more in depth clinical evaluation. The four-item version of the DTS had an efficiency score or diagnostic accuracy of 88%. The positive likelihood ratios of 9.1 and 4.3 suggest that this four-item scale supports a good degree of confidence in the interpretation of a positive score as indicating the diagnosis of PTSD. Our patient population, which had a prevalence of PTSD of 50%, comprised a varied group of trauma survivors including natural disaster, rape and combat, leading us to believe that the results may generalize to a wide population. Moreover, when a 10% prevalence rate of PTSD was assumed, the four-item scale remained effective at screening for the diagnosis. Adding more scale items (i.e. a five- or seven-item scale) did not significantly increase the efficiency score or diagnostic accuracy. However, the next step to take would be a prospective study of the SPAN in a population that has a lower base rate of PTSD. Thus, the advantages of the four-item SPAN include its brevity and good performance in both derivation and replication samples.

A new short seven-item screening scale for PTSD was recently published (Breslau et al., 1999). This scale differs from the SPAN in the following ways. Firstly, it is an interview measure and not self-rated. Secondly, the scale utilizes a yes/no response to symptoms whereas the SPAN utilizes a Likert scale that measures symptom severity on a five-point measure. Lastly, the Breslau scale measures lifetime prevalence of PTSD, whereas the SPAN measures current PTSD symptomatology. Interestingly, of the seven items included in the Breslau scale, five are from the avoidance and numbing symptom group and two are from the arousal symptom group. Thus, in both the Breslau scale and the SPAN, re-experiencing symptoms was found to be less critical in screening for the diagnosis of PTSD.

A comparison of the positive likelihood ratio of the SPAN with other tests is useful. For example, screening for alcohol dependence in general medical practice has become widespread given the recent data suggesting that up to 20% of patients seen in primary care settings satisfy diagnostic criteria for alcohol abuse or dependence (Bradley, 1992). Various screening tests are available, but the CAGE scale is a four-question screening test commonly used to detect alcohol abuse. When a score of two or more was used as the cutoff, the CAGE had a positive likelihood ratio of four (sensitivity of 0.91 and specificity of 0.77) in a large study of 385 psychiatric inpatients (Bernadt et al., 1982). Another study looked at the validity of the CAGE in screening for alcohol dependence in a medical walk-in triage clinic in a VA hospital. This study screened 1667 male veterans and used only one or more yes responses to indicate a positive response and achieved a sensitivity of 86% and a specificity of 93% (Liskow et al., 1995). A positive likelihood ratio can be calculated $0.86/1 - 0.93$ to yield a positive likelihood ratio of 12.3. The above two studies illustrate the phenomenon that modifying the 'cut score' on the screening test informing a binary decision that a problem is present or not influences sensitivity, specificity, and consequently the likelihood ratio (Allen et al., 1995). In addition the expected prevalence of the condition in the population also strongly influences the likelihood ratios.

Depression rating scales are widely used to

screen and monitor treatment of depressed patients. The Brief Carroll Depression Scale (CDS) is a 12-statement short form derived from the original Carroll Depression Scale (CDS) (Carroll et al., 1981), developed to serve as an efficient screening instrument for depression in primary care settings (Carroll, personal communication, 1997). In a comparison of the screening performance of several self-rated depression scales including the Brief CDS, the long and short forms of the Beck Depression Inventory (BDI) (Beck et al., 1961; Beck and Beck, 1972); the Center for Epidemiological Studies of Depression (CES-D) scale (Radloff, 1977); the Zung Self-Rated Depression Scale (SDS) (Zung, 1965), and a Visual Analogue Scale, the Brief CDS had the highest positive likelihood ratio of 5.2. (Carroll, personal communication, 1998). In contrast to our results, no replication scores were presented; these tend to be lower, as we found with the DTS.

The dexamethasone suppression test (DST) has been used to help confirm a diagnostic impression of endogenous depression. In one study that looked at 843 patients with a mood disorder, the DST was found to have a sensitivity of 46.2% and a specificity of 89.9% in differentiating endogenous from non-endogenous major depressive episodes (Rush et al., 1996). A positive likelihood ratio could be calculated $0.46/1-89.8$ to yield 4.6.

The SPAN offers itself as a brief screening tool for the diagnosis of PTSD, although its utility remains to be assessed in measuring treatment outcome. The four-item scale evaluates startle, physiological arousal at reminders of the trauma, anger, and numbness, three of which symptoms are specific to PTSD. We believe that the SPAN offers the advantages of ease of administration, brevity and diagnostic accuracy as a screening instrument. Given the prevalence of PTSD in the general population and the extensive psychiatric and medical comorbidity associated with PTSD, a brief screening instrument should offer the clinician significant utility and improve the diagnosis of PTSD. In addition, the SPAN may be valuable in distinguishing between treatments of differing effectiveness.

Further work will be required to establish the clinical and research utility of the SPAN in other populations including a general primary care setting where the prevalence of PTSD may be lower than in the studies presented.

Acknowledgements

Thanks are due to Drs Thomas Mellman, Jean Beckham, Michael Hertzberg, and Larry Tupler. This study was supported, in part, by grants from the National Institute of Mental Health (1RO1-MH44740 and RO1-MH47488) to Dr Davidson.

References

Allen, J.P., Maisto, S.A., Connors, G.J., 1995. Self-report screening tests for alcohol problems in primary care. Archives of Internal Medicine 155 (16), 1726–1730.

American Psychiatric Association, 1980. Diagnostic and Statistical Manual of Mental Disorders. 3rd ed. Washington, DC.

American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington, DC.

Beck, A.T., Beck, R.W., 1972. Screening depressed patients in family practice: a rapid technique.. Postgraduate Medicine 52, 81–85.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., Erbaugh, J., 1961. An inventory for measuring depression.. Archives of General Psychiatry 4, 561–571.

Bernadt, M.W., Mumford, J., Taylor, C., Smith, B., Murray, R.M., 1982. Comparison of questionnaire and laboratory tests in the detection of excessive drinking and alcoholism. Lancet 1, 325–328.

Blake, D.D., Weathers, F.W., Nagy, L.M., Kaloupek, D.G., Gusman, F.D., Charney, D.S., Keane, T.M., 1995. The development of a clinician-administered PTSD scale. Journal of Traumatic Stress 8, 75–90.

Bradley, K.A., 1992. Screening and diagnosis of alcoholism in the primary care setting. Western Journal of Medicine 156, 166–171.

Breslau, N, Peterson, E.L., Kessler, R.C., Schultz, L.R., 1999. Short Screening Scale for DSM-IV Posttraumatic Stress Disorder. American Journal of Psychiatry 156, 908–911.

Carroll, B.J., Feinberg, M., Smouse, P.E., Rawson, S.G., Greden, J.F., 1981. The Carroll Rating Scale for Depression. I. Development reliability and validation. British Journal of Psychiatry 138, 194–200.

Connor, K.M, Sutherland, S.M., Tupler, L.A., Malik, M.L., Davidson, J.R.T., 1999. Fluoxetine in post-traumatic stress disorder. British Journal of Psychiatry, in press.

Davidson, J.R.T., Malik, M.A., Travers, J., 1997a. Structured Interview for PTSD (SIP): psychometric validation for DSM-IV criteria. Depression and Anxiety 5, 127–129.

Davidson, J.R.T., Book, S.W., Colket, J.T., Tupler, L.A., Roth,

S., Hertzberg, M., Mellman, T., Beckham, J.C., Smith, R.D., Davidson, R.M., Katz, R., Feldman, M.E., 1997b. Assessment of a new self-rating scale for post-traumatic stress disorder. Psychological Medicine 27, 143–160.

Davidson, J.R.T., Weisler, R.H., Malik, M., Tupler, L.A., 1998a. Fluvoxamine in civilians with posttraumatic stress disorder. Journal of Clinical Psychopharmacology 18 (1), 93–95.

Davidson, J.R.T., Weisler, R.H., Malik, M.L., Connor, K.M., 1998b. Treatment of posttraumatic stress disorder with nefazodone. International Clinical Psychopharmacology 13 (3), 111–113.

Davidson, J.R.T., Tupler, L.A., Wilson, W.H., Connor, K.M., 1998c. A family study of chronic post-traumatic stress disorder following rape trauma. Journal of Psychiatric Research 32, 301–309.

Foa, E.B., 1995. Posttraumatic Stress Diagnostic Scale. National Computer Systems, Minneapolis, MN.

Fontana, A., Rosenheck, R., 1994. A Short Form of the Mississippi Scale for Measuring Change in Combat-Related PTSD. Journal of Traumatic Stress 7 (3), 407–414.

Gerardi, R., Keane, T.M., Penk, W., 1989. Sensitivity and specificity in developing diagnostic tests of combat-related post-traumatic stress disorder (PTSD). Journal of Clinical Psychology 45, 691–703.

Hammarberg, M., 1992. Penn Inventory for Posttraumatic Stress Disorder: psychometric properties. Psychological Assessment 4, 67–71.

Horowitz, M., Wilner, N., Alvarez, W., 1979. Impact of Event Scale: a measure a subjective distress. Psychosomatic Medicine 41, 209–218.

Keane, T.M., Wolfe, J., Taylor, K.L., 1987. Post-traumatic stress disorder: evidence for diagnostic validity and methods of psychological assessment. Journal of Clinical Psychology 43, 32–43.

Keane, T.M., Caddell, J.M., Taylor, K.L., 1988. Mississippi Scale for Combat-Related Post-Traumatic Stress Disorder: three studies in reliability and validity. Journal of Consulting and Clinical Psychology 56, 85–90.

Kessler, R.C., Sonnega, A., Bromet, E., Hughes, M.A., Nelson, C.B., 1995. Posttraumatic stress disorder in the national comorbidity survey. Archives of General Psychiatry 52, 1048–1060.

Liskow, B., Campbell, J., Nickel, E.J., Powell, B.J., 1995. Validity of the CAGE questionnaire in screening for alcohol dependence in a walk-in (triage) clinic. Journal of Studies on Alcohol 56 (3), 277–281.

Radloff, L.S., 1977 The CES-D scale: a self report depression scale for research in the general population. Applied Psychological Measurement 1, 385–401.

Rampes, H., Warner, J.P., Blizard, R., 1998. How to appraise an article on diagnosis. Psychiatric Bulletin 22, 506–509.

Rush, A.J., Giles, D.E., Schlesser, M.A., Orsulak, P.J., Parker, C.R., Jr., Weissenburger, J.E., Crowley, G.T., Khatami, M., Vasavada, N., 1996. The dexamethasone suppression test in patients with mood disorders. Journal of Clinical Psychiatry 57 (10), 470–484.

Spitzer, R.L., Williams, J.B.W., Gibbons, M.L., First, M.B., 1990. Structured Clinical Interview for DSM III-R. American Psychiatric Press, Inc., Washington, DC.

Weathers, F.W., Litz, B.T., Herman, D.S., Huska, J.A., Keane, T.M., 1993. The PTSD checklist: reliability, validity and diagnostic utility. Paper presented at the Annual Meeting of the International Society for Traumatic Stress Studies, San Antonio, TX, October.

Zung, W.W.K., 1965. A self-rating depression scale. Archives of General Psychiatry 12, 63–70.