Journal of Anxiety Disorders 28 (2014) 402–409

Contents lists available at ScienceDirect

Journal of Anxiety Disorders

# Properties of Swedish posttraumatic stress measures after a disaster

Filip K. Arnberg*, Per-Olof Michel, Kerstin Bergh Johannesson

*National Centre for Disaster Psychiatry, Department of Neuroscience, Uppsala University, Akademiska sjukhuset ing 10, SE-751 85 Uppsala, Sweden*

ARTICLE INFO

*Article history:*
Received 18 December 2013
Received in revised form 17 February 2014
Accepted 18 February 2014
Available online 19 March 2014

*Keywords:*
Posttraumatic stress disorder
Longitudinal invariance
Confirmatory factor analysis
Sensitivity and specificity
Mental health
Stressful events

ABSTRACT

This study evaluated the properties of Swedish versions of self-report measures of posttraumatic stress disorder (PTSD), with emphasis on the Impact of Event Scale-Revised (IES-R). Survey data from adult survivors 1, 3, and 6 years after the 2004 Indian Ocean tsunami ($n = 1506$) included the IES-R (from which the IES-6 was derived) and the 12-item General Health Questionnaire (GHQ-12). The PTSD Checklist (PCL) was included in one survey. A structured clinical interview was performed after 6 years ($n = 142$). Factor analyses of the IES-R and PCL indicated that a dysphoric-arousal model provided good fit invariant across assessments. Both measures were accurate in excluding PTSD while all measures provided poorer positive predictive values. The IES-R, but not the IES-6 and GHQ-12, evidenced stability across assessments. In conclusion, the Swedish IES-R and PCL are sound measures of chronic PTSD, and the findings illustrate important temporal aspects of PTSD assessment.



## 1. Introduction

Disaster survivors are at risk of developing chronic posttraumatic stress disorder (PTSD; American Psychiatric Association [APA], 2013), which can persist for several years and is associated with significant comorbidity and disability (Arnberg, Johannesson, & Michel, 2013; Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995; Taft, Stern, King, & King, 1999). Screening for PTSD thus serves an important purpose after disasters and evaluations of measures in appropriate contexts are needed. At the same time, there is no solid consensus about how to conceptualize the PTSD construct in terms of its symptom clusters and there are gaps in our knowledge about its longitudinal stability (Armour, Carragher, & Elhai, 2013; Elhai & Palmieri, 2011; McHugh & Treisman, 2007).

PTSD was devised as a three-dimensional construct that includes intrusions, avoidance, and hyperarousal reactions (American Psychiatric Association, 1980). Since then, empirical studies have consistently found that models with four or five factors provide better fit than the DSM-IV model (Armour, Carragher, et al., 2013; Yufik & Simms, 2010). These models differ from DSM, and among themselves, in how they conceptualize the DSM avoidance and hyperarousal factors. King, Leskin, King, and Weathers (1998) found support for a four-factor numbing model, in which the DSM avoidance factor is split into two factors labeled effortful avoidance and emotional numbing. Simms, Watson, and Doebbeling (2002) argued that symptoms of emotional numbing were examples of general distress. In support of their claim, they found a good fit for a dysphoria model, in which emotional numbing is combined with the general symptoms in the hyperarousal factor (i.e., symptoms D1-D3 in DSM-IV: sleep difficulty, irritability, and concentration problems) into a factor labeled dysphoria.

More recently, a five-factor model (Elhai et al., 2011) has been proposed, in which avoidance is split into effortful avoidance and numbing, and hyperarousal into dysphoric arousal and anxious arousal. Elhai et al. (2011) noted that the general hyperarousal symptoms are conceptually different from both the hyperarousal and the numbing items: First, they differ from the remaining hyperarousal symptoms (i.e., startle and hypervigilance) that characterize anxious arousal prototypical for fear-based symptomatology. Second, symptoms D1-D3 represent agitation and restlessness, which corresponds poorly with the notion of dysphoria (Elhai et al., 2011). Note that nearly all confirmatory factor analyses (CFAs) have used measures mapping directly onto the DSM-IV symptom criteria. If a model can provide good fit also when applied to other PTSD measures it would strengthen the evidence and could prevent that the refinement process ends prematurely (King et al., 2009). Corroborations also across disaster types would strengthen the evidence, and of particular relevance to the present study, the dysphoric-arousal model seems to

* Corresponding author at: Akademiska sjukhuset ing 10 v3, SE-751 85, Uppsala, Sweden. Tel.: +46 18 611 88 98; fax: +46 18 611 21 53.
*E-mail addresses:* filip.arnberg@neuro.uu.se (F.K. Arnberg), po.michel@gmail.com (P.-O. Michel), kerstin.bergh.johannesson@akademiska.se (K.B. Johannesson).

http://dx.doi.org/10.1016/j.janxdis.2014.02.005
0887-6185/

outperform the four-factor models in disaster contexts such as tsunamis (Armour, Carragher, et al., 2013) and earthquakes (L. Wang, Zhang, Shi, Zhou, Li, et al., 2011).

A factor model should also provide similar fit across several assessments (i.e., longitudinal invariance), as it is essential, perhaps particularly for PTSD that has a clear temporal aspect, to establish that we assess the same construct regardless of timing (Elhai & Palmieri, 2011). The few studies that exist suggest that at least configural invariance seems plausible, but more research is needed (King et al., 2009; Krause, Kaltman, Goodman, & Dutton, 2007; M. Wang, Elhai, Dai, & Yao, 2012).

With regard to screening, widely used self-report measures such as the PTSD Checklist (PCL; Weathers, Litz, Herman, Huska, & Keane, 1993) and the Impact of Event Scale–Revised (IES-R; Weiss, 2004) have been evaluated in clinical and military settings whereas less is known of their properties in the context of disasters (for reviews, see Brewin, 2005; McDonald & Calhoun, 2010; Wilkins, Lang, & Norman, 2011). In addition, several studies have used other self-report measures as the reference standard (e.g., Creamer, Bell, & Failla, 2003; Olde, Kleber, van der Hart, & Pop, 2006), which seem to inflate the screening accuracy as compared to studies that use a structured clinical interview as reference (Beck et al., 2008).

Longitudinal stability is relevant also for diagnostic accuracy when screening for PTSD because of the long-standing notion of variability within and across symptom clusters as time passes from the event (Horowitz, Wilner, & Alvarez, 1979; Kessler et al., 1995). Longitudinal stability becomes particularly relevant to brief measures. A screening measure should be easy to administer, complete, and score (Brewin et al., 2002). Simple, brief measures would therefore be preferred over complex screening methods. Brevity is particularly important in contexts involving large groups or rapid assessments, whereas work with clinical populations may require comprehensive, detailed measures. Brief measures seem to perform as well as longer measures in cross-sectional comparisons (Brewin, 2005; Thoresen et al., 2010). However, abbreviated measures usually do not include all symptom clusters (Brewin et al., 2002; Ouimette, Wade, Prins, & Schohn, 2008), with the six-item abbreviation of the IES-R (IES-6) being an exception (Thoresen et al., 2010). The narrow coverage of symptoms in brief measures entails a greater risk of failing to detect symptoms prominent at the time of screening. To our knowledge, however, their stability has not been assessed (Brewin et al., 2002; Thoresen et al., 2010).

In addition to brief measures, there are potential benefits of screening for PTSD with measures of general distress. Fewer screening tools could lower costs and, importantly, yield higher response rates to screening programs. A general measure could potentially screen for a range of anxiety and mood disorders whilst having similar diagnostic accuracy to detect PTSD, as would PTSD-specific measures. The 12-item General Health Questionnaire (GHQ-12; Goldberg, 1972) performed slightly worse than a 4-item PTSD measure when given to veteran soldiers in primary care (Ouimette et al., 2008), although the differences between the measures were small and warrant further investigation.

In summary, the issue of longitudinal stability is unclear both in terms of symptom clusters and with regard to screening. In a national perspective, several widely used measures have not been evaluated. The present study aims to shed light on the properties of the Swedish versions of self-report measures for posttraumatic stress in a disaster context. We describe convergent and construct validity, using CFA to evaluate different models with the PCL and IES-R. Longitudinal invariance was investigated with the best fitting model for the IES-R. The diagnostic accuracy were evaluated for these measures together with the IES-6 and a measure of general distress, the GHQ-12, and we present mean scores for survivors and a minimally exposed comparison group.

## 2. Materials and methods

### 2.1. Procedure and participants

The present study is part of a national cohort study of Swedish survivors from the 2004 Southeast Asia tsunamis that followed from a massive earthquake in the Indian Ocean. Swedish authorities at the national airports registered all repatriated Swedish citizens from destinations in Southeast Asia, regardless of their actual disaster exposure or not, during the first 3 weeks after the disaster. The study was approved by the Regional Ethical Review Board in Uppsala, Sweden.

#### 2.1.1. Mail surveys

The assessments of the national cohort include a mail survey 1 year and 2 months after the disaster (T1, $n = 4932$; Johannesson et al., 2009) and two mail surveys of the respondents from the first survey: at 3 years and 1 month (T2, $n = 3457$; Johannesson, Lundin, Fröjd, Hultman, & Michel, 2011) and 6 years and 3 months after the event (T3, $n = 2643$; Johannesson, Arnberg, & Michel, 2012). Disaster exposure was established at T1 based on previous analyses of this sample (Johannesson et al., 2009) with 30 multiple-choice items. The participants were included in the exposed group if they had been caught by the waves or been severely injured; experienced life threat, death or life threat to close ones; or witnessed horrifying events (i.e., deceased people, survivors with severe injuries, people searching for others among corpses). Participants who had not experienced any of the above were not included in the analyses but served as a non-exposed comparison group.

In order to facilitate direct comparisons across measures and assessments we excluded participants with two or more missing values at any measure at any assessment ($n = 941$). However, it was common to all measures and assessments that participants had one missing value ($n = 356$, 4–8% of participants across measures). In order to prevent unnecessary loss of data participants with one missing item at any assessment were included and missing values were imputed.

The final survey sample included 1506 survivors and 541 comparisons. The survivors were 61% women, had a mean age of 48.8 years ($SD = 13.6$), 79% were mainly married or cohabiting, 48% had a university education, and 93% were either employed, students, or had retired from employment. The comparison group was similar to the survivors with respect to demographics (all $ps > .05$), included 56% women with a mean age of 50 years ($SD = 14.8$), 76% were cohabiting, and 94% were employed/students/retired. A majority (65%) of survivors indicated having been caught in or chased by the tsunamis, 28% were injured, 9% received hospital care, and 15% had been bereaved. In the immediate aftermath, 33% were unsure of the fate of their significant others, 39% were missing clothes/other belongings, and 62% witnessed several deceased bodies and injured people.

#### 2.1.2. Clinical interviews

The participants at T2 were asked for consent to participate in a telephone interview and 62% agreed to participate. The interviews were conducted at 6 years post-disaster with a random sample of 200 individuals from the exposed group. An examination of the rates of psychopathology has been reported previously (Arnberg et al., 2013). Those who consented to an interview and those who did not were highly similar in age (both groups $M = 45$ years; $t = 0.79$, $p = .4$, partial $\eta^2 < 0.001$), gender (43% vs. 41% women, $\chi^2 = 3.1$, $p = .22$, $\varphi = .03$), and employment status (both 94% employed, $\chi^2 = 0.47$, $p = .50$, $\varphi = .01$). However, participants who declined to be interviewed had somewhat higher scores on the IES-R at T2 ($M_{diff} = 3.65$, $t = 7.36$, $p < .001$, partial $\eta^2 = 0.01$). Of the 200 participants approached for interviews, 27 declined and 31 could