



## Examining the disconnect between psychometric models and clinical reality of posttraumatic stress disorder



Mark Shevlin<sup>a,\*</sup>, Philip Hyland<sup>b</sup>, Thanos Karatzias<sup>c,d</sup>, Jonathan I Bisson<sup>e</sup>, Neil P. Roberts<sup>f</sup>

<sup>a</sup> Ulster University, School of Psychology and Psychology Research Institute, Magee Campus, Derry, Northern Ireland, Ireland

<sup>b</sup> National College of Ireland, School of Business, Dublin, Ireland

<sup>c</sup> Edinburgh Napier University, School of Health and Social Care, Edinburgh, UK

<sup>d</sup> NHS Lothian, Rivers Centre for Traumatic Stress, Edinburgh, UK

<sup>e</sup> Cardiff University, School of Medicine, Cardiff, UK

<sup>f</sup> Psychology and Counselling Directorate, Cardiff and Vale University Health Board, Cardiff, UK

### ARTICLE INFO

#### Article history:

Received 7 September 2016

Received in revised form 2 February 2017

Accepted 17 February 2017

Available online 20 February 2017

#### Keywords:

Posttraumatic stress disorder (PTSD)

Confirmatory factor analysis (CFA)

Childhood maltreatment

Psychiatric diagnosis

Psychometric modelling.

### ABSTRACT

There have been many factor analytic studies aimed at testing alternative latent structures of DSM-IV and DSM-5 posttraumatic stress disorder (PTSD) symptoms. The primary rationale for such studies is that determining the 'best' factor analytic model will result in better diagnoses if that structure is the basis for diagnostic decisions. However, there appears to be a disconnect between the factor analytic modelling and the diagnostic implications. In this study, we derived prevalence rates based on commonly reported models of PTSD, based on data from two clinical samples ( $N=434$ ), and also assessed if the different models generated consistent risk estimates in relation to the effects of childhood maltreatment. We found that the different models produced different prevalence rates, ranging from 64.5% to 83.9%. Furthermore, we found that the relationship between childhood maltreatment and 'diagnosis' varied considerably depending upon which latent symptom profile was adopted. It is argued that, given the maturity of this area of research, factor analytic studies of PTSD should now include information on the diagnostic implications of their findings.

© 2017 Elsevier Ltd. All rights reserved.

### 1. Examining the disconnect between psychometric models and clinical reality

There has been a plethora of factor analytic studies aimed at testing alternative latent structures of posttraumatic stress disorder (PTSD) symptoms. In a recent review [Armour, Müllerová, & Elhai \(2016\)](#) identified 112 research papers published since 1994 that used confirmatory factor analysis (CFA) to test alternative models based on DSM-IV and DSM-5 symptoms. An additional 70 studies were not included as they did not meet certain exclusion criteria such as failing to test alternative models, not using DSM based measures, or using data from children. It was concluded that a five-factor Dysphoric Arousal model ([Elhai et al., 2011](#)) was the best representation of DSM-IV symptoms and a seven-factor Hybrid model ([Armour et al., 2015](#)) was the best representation of DSM-5 symptoms.

This plethora of factor analytic studies is predicated on the assertion that the determination of the most accurate and parsimonious account of the latent structure of PTSD symptoms is a necessary prerequisite for successful diagnosis, accurate estimates of prevalence, identification of key etiological variables for the development and maintenance of PTSD, and precise assessments of treatment response (e.g., [Armour et al., 2016](#); [Elhai & Palmieri, 2011](#)). Despite such arguments for the importance of identifying the optimal symptom profile, none of the existing DSM-5 studies that have tested alternative symptom structures have provided a corresponding diagnostic algorithm for diagnosis (e.g., number of symptoms required from each cluster). The link between the psychometric structure of symptoms and diagnostic criteria has not been made. Consequently, it is currently unknown how the adoption of any of the alternative models of PTSD as diagnostic systems would influence diagnostic rates relative to the existing DSM-5 criteria. Furthermore, it also remains unknown whether adoption of any of the alternative PTSD models will affect the nature of the relationship between etiological variables and the likelihood of a PTSD diagnosis.

\* Corresponding author.

E-mail address: [m.shevlin@ulster.ac.uk](mailto:m.shevlin@ulster.ac.uk) (M. Shevlin).

The overall aims of this study are to provide the first examination of the diagnostic rates generated from the various existing factor analytically derived models of DSM-5 PTSD and determine if a specific traumatic event, namely childhood maltreatment, was differentially associated with PTSD depending on the model used to derive the diagnosis. Childhood maltreatment has been consistently shown to predict, and confer susceptibility, to subsequent PTSD. In order to meet the aims of this study we (1) tested seven alternative factor analytic models of DSM-5 PTSD symptoms (using both commonly employed methods of estimation), (2) proposed and applied diagnostic criteria for establishing symptom-based diagnoses in-line with the DSM-5 standards, and (3) assessed differential risk of 'diagnosis' according to each PTSD symptom profile based on childhood traumatic exposure.

## 2. Method

### 2.1. Participants and procedures

The participant group ( $N=434$ ) was comprised of two clinical samples from the United Kingdom (UK). One group ( $n=195$ ) were attendees of a National Health Service (NHS) trauma centre in Scotland who had been referred by a general practitioner, psychiatrist, or psychologist for psychological therapy. The second group ( $n=239$ ) were recruited via Wales' National Centre for Mental Health (NCMH), a research centre investigating a number of mental health conditions. NCMH participants were recruited through primary and secondary health services, and social media. Individuals who reported that they had previously been given a diagnosis of PTSD, or those who had screened positively for PTSD, and were aged 18 or over, were invited to join the 'All Wales PTSD Registry'. All individuals in the current study from the two research samples reported exposure to a traumatic event(s) fulfilling the DSM-5's 'Criterion A' requirement for diagnosis of PTSD, and experienced these symptoms for a period greater than one month fulfilling the DSM-5's 'Criterion F' requirement. Ethical approvals for data collection were separately provided by University and National Health Service ethical review boards.

Of the total sample, 56.5% were female ( $n=245$ ) and the average age was 44.85 years ( $SD=12.81$ ). The majority of the sample indicated their marital status to be single (42.2%,  $n=182$ ), 28.3% were married ( $n=122$ ), 12.5% were co-habiting with a partner ( $n=54$ ), and 16.9% were divorced ( $n=73$ ). Just over two-thirds of the current sample reported having been exposed to some form of traumatic exposure during their childhood (68.4%,  $n=290$ ). The mean number of traumatic life exposures was 6.11 ( $SD=3.08$ ) based on an amended version of the Life Events Checklist (Gray, Litz, Hsu, & Lombardo, 2004), which included two additional items inquiring about exposure to childhood sexual abuse and childhood physical abuse.

### 2.2. Measures

The PTSD Checklist for DSM-5 (PCL-5; Weathers et al., 2013) includes 20 self-report items that capture the DSM-5 PTSD symptoms. Respondents are asked to "...indicate how much you have been bothered by that problem in the past month" and respond using a five-point Likert scale (0 = *Not at all*, 1 = *A little bit*, 2 = *Moderately*, 3 = *Quite a bit*, 4 = *Extremely*). The psychometric properties of the PCL-5 have been assessed across multiple trauma-exposed samples and the scale has demonstrated satisfactory reliability and validity (e.g., Blevins, Weathers, Davis, Witte, & Domino, 2015; Bovin et al., 2015). Among the current sample the internal reliability for the full scale was satisfactory ( $\alpha=.92$ ), as was the reliability for each subscale: intrusions ( $\alpha=.88$ ), avoid-

ance ( $\alpha=.69$ ), negative alterations in cognitions and mood (NACM) ( $\alpha=.83$ ), and hyperarousal ( $\alpha=.76$ ). For the purposes of estimating quasi-diagnostic rates, each symptom was dichotomised to reflect its presence or absence. In-line with standard conventions for estimating the presence or absence of a symptom based on self-report data (e.g., Bovin et al., 2015; Elklit & Shevlin, 2007; Hansen, Hyland, Armour, Elklit, & Shevlin, 2015), a score of 2 (*Moderately*) or greater was used to indicate symptom endorsement.

Exposure to childhood maltreatment was based on the responses to two questions that asked about exposure "...to childhood physical abuse" and exposure "...to childhood sexual abuse or molestation". The questions used a 'Yes/No' response format. If a participant endorsed either, or both, questions they were coded as having experienced childhood maltreatment. Basic sociodemographic variables including age, gender, and marital status were also collected.

### 2.3. Analysis

First, seven alternative model solutions for the DSM-5's PTSD symptoms were assessed using CFA to determine the fit of each model. The item mapping is presented in Table 1. These analyses were conducted in Mplus 7.0 (Muthén & Muthén, 2013) with robust maximum likelihood estimation (MLR; Yuan & Bentler, 2000) treating the five-point Likert scale scores as being continuous and also using the robust weighted least squares estimator (WLSMV) based on the polychoric correlation matrix of latent continuous response variables. For both methods of estimation standard recommendations were followed to assess model fit (Hu & Bentler, 1998, 1999): a non-significant chi-square ( $\chi^2$ ), Comparative Fit Index (CFI; Bentler, 1990) and Tucker Lewis Index (TLI; Tucker & Lewis, 1973) values above .95 reflect excellent fit, while values above .90 reflect acceptable fit; Root-Mean-Square Error of Approximation with 90% confidence intervals (RMSEA 90% CI; Steiger, 1990) with values of .06 or less reflect excellent fit while values less than .08 reflect acceptable fit. For the models based on MLR estimation the Standardised Root-Mean-Square Residual (SRMR; Jöreskog & Sörbom, 1996) was also used with values of .06 or less indicating excellent fit while values less than .08 indicating acceptable fit. The Bayesian Information Criterion (BIC; Schwarz, 1978) was used to evaluate and compare models, with the smallest value indicating the best fitting model. In relation to the BIC Raftery (1996) suggested that a 2–6 point difference offers evidence of model superiority, a 6–10 point difference indicates strong evidence of model superiority, and a difference greater than 10 points indicates very strong evidence of model superiority. For WLSMV estimation the Weighted Root Mean Square Residual (WRMR) was also used with values less than 1 indicating acceptable model fit.

The second stage of the analysis sought to determine the probable self-report based prevalence rate of PTSD for each model based on a score of 2 (*Moderately*) or greater being used to indicate symptom endorsement. The DSM-5's criteria for PTSD diagnosis requires exposure to a traumatic event (Criterion A), the presence of one of five symptoms of Intrusions (Criterion B), one of two Avoidance symptoms (Criterion C), two of seven NACM symptoms (Criterion D), and two of six Hyperarousal symptoms (Criterion E). In addition, DSM-5 also requires that symptoms persist for more than one month (Criterion F), are associated with functional impairment (Criterion G), and are not due to substance use, medication or any other illness (Criterion H). For the purposes of this study only Criteria A-F were assessed for diagnostic purposes.

Given that the authors who developed the respective alternative symptom models of PTSD did not provide a corresponding symptom-based diagnostic algorithm, it was necessary for us to develop such criteria. In order to develop equitable and logically-derived symptom-based diagnostic criteria for each model we

**Table 1**  
Item mapping for the alternative DSM-5 PTSD factor models.

Symptoms	DSM-5 (4 factors)	Dysphoria (4 factors)	Dysphoric arousal (5 factors)	Anhedonia (6 factors)	External behaviours (6 factors)	Alternative dysphoria (6 factors)	Hybrid (7 factors)
B1: Unwanted memories	I	I	I	I	I	I	I
B2: Disturbing dreams	I	I	I	I	I	I	I
B3: Reliving	I	I	I	I	I	I	I
B4: Feeling upset	I	I	I	I	I	I	I
B5: Physical reactions	I	I	I	I	I	I	I
C1: Internal avoidance	A	A	A	A	A	A	A
C2: External avoidance	A	A	A	A	A	A	A
D1: Amnesia	N	D	N	N	N	D	N
D2: Negative self-beliefs	N	D	N	N	N	D	N
D3: Self-blame	N	D	N	N	N	D	N
D4: Negative feelings	N	D	N	N	N	D	N
D5: Loss of interest	N	D	N	AN	N	AN	AN
D6: Distant	N	D	N	AN	N	AN	AN
D7: No positive feelings	N	D	N	AN	N	AN	AN
E1: Aggression	H	D	DA	DA	EB	EB	EB
E2: Risky behaviour	H	D	DA	DA	EB	EB	EB
E3: On guard	H	H	AA	AA	AA	AA	AA
E4: Easily startled	H	H	AA	AA	AA	AA	AA
E5: Concentration	H	D	DA	DA	DA	D	DA
E6: Sleep problems	H	D	DA	DA	DA	D	DA

Note. I=intrusions; A=avoidance; N=negative alternations in cognition and mood; H=hyperarousal; D=dysphoria; DA=dysphoric arousal; AA=anxious arousal; AN=anhedonia; EB=externalized behaviour.

**Table 2**  
Number of symptoms required from each cluster for each PTSD symptom profile for diagnostic purposes.

PTSD symptom cluster	DSM-5	Dysphoria (4 factors)	Dysphoric arousal (5 factors)	Anhedonia (6 factors)	External behaviours (6 factors)	Alternative dysphoria (6 factors)	Hybrid (7 factors)
Intrusions	1/5	1/5	1/5	1/5	1/5	1/5	1/5
Avoidance	1/2	1/2	1/2	1/2	1/2	1/2	1/2
NACM	2/7	–	2/7	1/4	2/7	–	1/4
Hyperarousal	2/6	1/2	–	–	–	–	–
Dysphoria	–	3/11	–	–	–	2/6	–
Dysphoric Arousal	–	–	1/4	1/4	1/2	–	1/2
Anxious Arousal	–	–	1/2	1/2	1/2	1/2	1/2
Anhedonia	–	–	–	1/3	–	1/3	1/3
EB	–	–	–	–	1/2	1/2	1/2
Total symptoms required	6/20	6/20	6/20	6/20	7/20	7/20	7/20

Note. NACM = negative alternations in cognition and mood; EB = externalized behaviour.

adopted the DSM-5 conventions as a basis whereby for any symptom cluster including 2–5 symptoms, one symptom must be present for that cluster to be endorsed (as per requirements for Criteria B and C), and for any symptom cluster that contains 6 or more symptoms, two symptoms must be present for that cluster to be endorsed (as per requirements for Criteria D and E). One exception to this guiding principle was the requirement of 3 of the 11 ‘dysphoria’ symptoms from the four-factor Dysphoria model. This exception was made based on the large number of symptoms included within this cluster and to include a total of 6 out of 20 symptoms for diagnosis, as per the DSM-5 criteria. The symptom-based “diagnostic” requirements used are displayed in Table 2.

The final stage of the analysis sought to determine whether a history of childhood maltreatment was differentially associated with the risk of meeting symptom-based diagnosis of PTSD according to the alternative symptom profiles. Chi-square tests of independence and odds ratios (OR) with 95% confidence intervals were estimated between each PTSD model and the variable representing childhood maltreatment.

### 3. Results

The fit statistics for the alternative models of PTSD are reported in Table 3. Based on the fit statistics using MLR estimation the four-factor DSM-5 model of PTSD was judged to be an unsatisfactory representation of the sample data, failing to satisfy the criteria for

acceptable model fit according to the CFI and TLI indices. The DSM-5 model also produced the poorest fit of the seven models according to the RMSEA, SRMR, and BIC results. Of the alternative models, the ‘Dysphoria’, ‘Dysphoric Arousal’, ‘External Behaviours’, and ‘Alternative Dysphoria’ models exhibited acceptable model fit, while the ‘Anhedonia’ and ‘Hybrid’ models produced excellent model fit results. The fit statistics were all acceptable for the ‘Anhedonia’ model, but in addition it had the lowest BIC value, and on the basis of this it was considered the ‘best’ model. Based on WLSMV estimation the CFI and TLI suggested that all models were acceptable, with highest values for the ‘Anhedonia’ and ‘Hybrid’ models. Overall, the RMSEA values were higher than for MLR estimation, but the lowest values were for the ‘Anhedonia’ and ‘Hybrid’ models; these models also produced the lowest WRMR values. Both sets of fit statistics converged on the ‘Anhedonia’ and ‘Hybrid’ models being the best models.

#### 3.1. Diagnostic estimates

Table 3 also shows the estimated ‘diagnostic’ rates corresponding to the seven symptom algorithms. Considerable variation existed across the seven symptom profiles with the four-factor DSM-5 model producing the highest diagnostic rates (83.9%) and the seven-factor Hybrid model producing the lowest diagnostic rates (64.5%). The difference in diagnostic rates between these two models was statistically significant ( $z=6.10, p<.001$ ). The Anhe-

**Table 3**  
Model fit statistics for alternative models of DSM-5 PTSD and prevalence rates based on each symptom algorithm.

Model	Estimator	$\chi^2$	df	p	CFI	TLI	RMSEA (90% CI)	SRMR WRMR	BIC	DX %
DSM-5 model of PTSD	MLR	535.751	164	.000	.880	.860	.076	.053	23429	83.9
	WLSMV	702.554	164	.000	.933	.922	(.069–.083)	1.355		
Dysphoria model	MLR	456.101	164	.000	.905	.890	.067	.049	23332	68.5
	WLSMV	537.427	164	.000	.954	.946	(.060–.075)	1.151		
Dysphoric arousal model	MLR	447.909	160	.000	.907	.889	.068	.048	23345	78.4
	WLSMV	536.892	160	.000	.953	.944	(.060–.075)	1.140		
Anhedonia model	MLR	301.192	155	.000	.953	.942	.049	.040	23203	76.0
	WLSMV	368.918	155	.000	.973	.967	(.041–.057)	.893		
External Behaviours model	MLR	432.548	155	.000	.910	.890	.047	.047	23358	65.3
	WLSMV	526.510	155	.000	.954	.943	(.060–.075)	1.114		
Alternative Dysphoria model	MLR	371.971	155	.000	.930	.914	.060	.044	23288	65.0
	WLSMV	432.085	155	.000	.966	.958	(.052–.067)	.987		
Hybrid model	MLR	278.983	149	.000	.958	.946	.047	.038	23214	64.5
	WLSMV	345.73	149	.000	.976	.969	(.038–.056)	.845		

Note.  $N=434$ ;  $\chi^2$  = chi-square goodness of fit statistic;  $df$  = degrees of freedom;  $p$  = probability value; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA (90% CI) = Root-Mean-Square Error of Approximation with 90% confidence intervals; SRMR = Standardized Square Root Mean Residual; BIC = Bayesian Information Criterion; DX % = percentage of sample meeting “diagnosis” according to each symptom profile.

donia model, which displayed the best model fit results, had an associated estimated diagnostic rate of 76.0%. The estimated rate of diagnosis based on this symptom algorithm was statistically different from both the DSM-5 model (76.0% vs 83.9%,  $z=2.70$ ,  $p=.003$ ), and the similarly well-fitting Hybrid model (76.0% vs 64.5%,  $z=3.47$ ,  $p<.001$ ). These results indicated that model selection has considerable influence on the proportion of trauma survivors that may qualify for a PTSD diagnosis.

The chi-square results in Table 4 showed a statistically significant and positive relationship between exposure to childhood maltreatment and meeting symptom-based diagnostic requirements for each of the seven PTSD models. Consistent with the diagnostic prevalence results, the risk of PTSD diagnosis was strongest for the DSM-5 symptom profile ( $OR=3.50$ ), and weakest for the Hybrid model ( $OR=1.89$ ). Risk of PTSD ‘diagnosis’ according to the best-fitting Anhedonia model ( $OR=2.77$ ) was lower than the DSM-5 model but greater than the similarly well-performing Hybrid model. These findings indicate that the relationship between an established etiological risk-factor and ‘diagnosis’ varies considerably depending upon which latent symptom profile is adopted.

#### 4. Discussion

The overall aims of this study were to provide the first examination of the diagnostic rates generated from the various existing factor analytically derived models of DSM-5 PTSD and determine if an established risk-factor for PTSD diagnosis, namely childhood maltreatment, was differentially associated with PTSD depending on the model used to derive the diagnosis. Overall the results indicated that there was significant variation in diagnostic rates and the magnitude of the association between childhood maltreatment and PTSD was moderated by the model that was used to derive the diagnosis. This indicates that proposing alternative models of PTSD

and failing to address the diagnostic consequences does not “. . . aid in the development of accurate assessment instruments” (Armour et al., 2016, p. 70).

The fit of the factor analytic models tested in this study was largely consistent with much of the existing research findings; based on both MLR and WLSMV estimation the Anhedonia and Hybrid models had the best model fit and the fit of the DSM-5 model was relatively poor. Indeed, it could be argued that the fit of all models, with the exception of the DSM-5 model, were ‘acceptable’. This provides some evidence of construct validity. However, the differences in the diagnostic rates derived from these different models are notable. First, all alternatives to the DSM-5 model (83.9%) produced lower rates of diagnosis and the differences between the two best fitting models (Anhedonia and Hybrid) and the DSM-5 were statistically significant. Indeed, the rates for the Anhedonia (76%) and the Hybrid model (64.5%) were significantly lower than for the DSM-5, and these equate to a 9.4% and 22.8% reduction in prevalence relative to DSM-5, respectively. The External Behaviours model and the Alternative Dysphoria model resulted in a 12.2% and 22.5% decrease in prevalence respectively compared to the DSM-5, and the Dysphoric arousal model resulted in a prevalence rate closest to DSM-5 being 6.6% lower. These results can be compared to those of Ford, Elhai, Ruggiero, and Frueh (2009) who examined the effect of diagnostic alternations in the specification of a DSM-IV based diagnosis of PTSD. They estimated and compared the prevalence of PTSD from two large-scale, adult epidemiological datasets (National Comorbidity Survey-Replication and National Survey of Adolescents) using the DSM-IV three-factor PTSD diagnostic model, and the four-factor models proposed by King et al. (2009) and Simms, Watson, and Doebbeling (2002) and found that the differences were modest. The prevalence estimates based on the King (6.0%) and Simms (5.9%) models were lower than for the DSM-IV model (6.81%) based on the National Comorbidity Survey-Replication, and these equate to a 12.5% and 14.1% reduction,

**Table 4**  
Bivariate relationships between childhood maltreatment and risk of diagnosis for each model of PTSD.

	Childhood maltreatment (Yes)	$\chi^2$ (df) p	OR (95% CI)
DSM-5 PTSD	229 (62.1%)	20.04 (1) <.001	3.50 (1.98/6.18)
Dysphoria PTSD	192 (51.8%)	11.62 (1) <.001	2.23 (1.40/3.55)
Dysphoric arousal PTSD	216 (58.5%)	14.18 (1) <.001	2.74 (1.63/4.58)
Anhedonia PTSD	211 (57.2%)	16.70 (1) <.001	2.77 (1.68/4.57)
External Behaviours PTSD	180 (48.6%)	6.28 (1) <.05	1.79 (1.13/2.83)
Alternative Dysphoria PTSD	180 (48.6%)	7.11 (1) <.05	1.86 (1.17/2.93)
Hybrid PTSD	179 (48.4%)	7.55 (1) <.05	1.89 (1.20/2.98)

Note.  $\chi^2$  = Chi-square test of independence; df = degrees of freedom; p = statistical significance; OR (95% CI) = odds ratio with 95% confidence intervals.

respectively. Similarly, the King (6.9%) and Simms (5.6%) models were lower than for the DSM-IV model (8.2%) based on the National Survey of Adolescents, but these differences were larger, representing a 15.8% and 31.4% reduction, respectively. Overall, alterations to the DSM-IV and DSM-5 models of PTSD, based on factor analytic modelling, will likely result in lower prevalence rates, and some of these differences are likely to be large (e.g. Hybrid model and Alternative Dysphoria model) while for others the difference will likely be modest (Anhedonia model).

Ford et al. (2009) concluded that "...this research contributes to our understanding of the PTSD construct's symptom structure and how empirically supported PTSD factor models can translate to meaningful changes in diagnosing the disorder" (p. 750). In contrast, more recent factor analytic research has failed to make the connection between proposed psychometric models and their clinical implications in terms of how models map onto diagnosis and the impact that particular symptoms organisations have on estimates of prevalence. This issue has been identified previously; Elhai and Palmieri (2011) noted that in relation to CFA of PTSD symptoms "...Such research has implications for diagnostic algorithms for PTSD; incorporating knowledge from state-of-the-art CFA models in diagnosis can alter PTSD prevalence rates". This issue appears to have been subsequently ignored and the proliferation of CFA studies has continued with no serious attempt being made to bridge the gap between psychometric modelling and the clinical reality that the symptoms being modelled are the basis for diagnosis.

The implications of proposing a psychometric model upon which diagnosis should be based extends beyond (but is related to) the issue of prevalence. If a factor analytically derived model implies a particular diagnostic model and a diagnosis is made based on this model, then the association between known risk factors and the resultant diagnosis should not be wildly inconsistent with commonly accepted estimates of risk. For example, it is well known that obesity is a risk-factor for diabetes, and Class II obesity increases the odds of a diagnosis of diabetes by about 3 times (Ganz et al., 2014). If a new measure of diabetes was developed and the odds ratio for the association between Class II obesity and the new diagnosis of diabetes was much lower, say 1.5, or much higher, say 5, then questions about whether this new measure was actually measuring diabetes would be raised. Similarly, if different PTSD models produced different associations with a well-established risk-factor, childhood maltreatment, then it would be unclear if (what is commonly considered to be) PTSD is still being diagnosed. The results reported in Table 4 show that the choice of model which the diagnosis is based upon has a profound effect on the estimate of risk associated with childhood maltreatment. With a diagnosis based on the DSM-5 model the odds ratio is 3.50 and this is higher than for all other models. The odds ratio for the 'best' alternative models were much lower (Anhedonia OR = 2.77, Hybrid OR = 1.89). It is not possible to determine which estimate is 'correct', but the variability in the magnitude of the odds ratios indicates that there can be profound consequences when estimating the degree of risk associated with a diagnosis of PTSD.

It is interesting to note that the extensive factor analytic literature has failed to find consistent support for both the DSM-IV and DSM-5 specifications, and has instead suggested that the latent structure of PTSD symptoms is considerably more complex. Despite the efforts of the American Psychiatric Association to address this problem by expanding the number of symptom clusters from three to four in the 5th edition, the decision to introduce these additional symptoms in DSM-5 seems to have only exacerbated the complexity issue with the literature supporting six-factor (Anhedonia) and seven-factor (Hybrid) models of PTSD. The expanded DSM-5 PTSD symptom criteria has been criticised on many fronts, but most notably for leading to a bewilderingly heterogeneous disorder which Galatzer-Levy and Bryant (2013, p. 656) succinctly noted has "636,120 possible presentations". Frances and Nardo (2013) previously highlighted the bias within the DSM-5 to focus on diagnostic sensitivity (broadening diagnostic boundaries to avoid false-negatives and thus increasing the likelihood of false-positive diagnoses) over diagnostic specificity (tightening diagnostic boundaries to avoid false-positives and thus increasing the likelihood of false-negative diagnoses) and this is particularly prescient with respect to the diagnostic changes made to PTSD from DSM-IV to DSM-5. Researchers who propose alternative symptom structures of DSM-5 PTSD symptoms must grapple with this issue of diagnostic sensitivity versus specificity when articulating a potential symptom structure and (ideally) a corresponding diagnostic criterion. As described in Table 4, if one applies the criteria that we proposed in this study, four models require a total of 6 out of 20 symptoms to be endorsed to qualify for diagnosis, while three models require a total of 7 out of 20 symptoms to be endorsed to qualify for a diagnosis. As factor-analytically derived models become increasingly complex (e.g., the Hybrid model) the threshold for diagnosis would appear to increase. Consequently, increasingly complex models appear to favour diagnostic specificity rather than diagnostic sensitivity. This critical issue that is well-worn with regards to the DSM-5 conceptualisation of PTSD cannot be ignored by proponents of alternative models of PTSD.

Factor analytic research has an important role in psychological and psychiatric research by allowing us to develop and test the plausibility of alternative, or competing, models to better understand the natural clustering of symptoms of a particular disorder. This paper does not argue against the use of CFA in PTSD research per se, rather, the way that such analyses are conducted should reflect the maturity of an area of investigation. Maybe it is time that CFAs of PTSD symptoms begin to move from merely providing a statistical description of sample data to addressing the implications of the models in the context of diagnosis. There are many important issues to be addressed in relation to how a diagnostic algorithm should be established; should the number of required symptoms be the same for all models, are all symptoms clusters required to be included, should there be 'core symptoms' (such as the "Depressed mood" and "Loss of interest" symptoms for major depressive disorder)? Similarly, there are important research questions related to the way in which proposed diagnostic algorithms can be evaluated. Traditionally, receiver operating characteristic

(ROC) analysis can be used to identify cut-off scores that optimise sensitivity/specificity. However, the PTSD diagnosis is based on multiple symptom cluster scores (or counts) rather than a single score which would require multiple cut-off scores to be estimated in a multivariate, rather than univariate, ROC model. There is no agreed method by which this can be achieved, although there is some promising research using machine learning methods that may be appropriate for this purpose (Wang & Li, 2013; Galatzer-Levy, Karstoft, Statnikov, & Shalev, 2014). In addition, there is likely to be variability in the 'best' cut-off scores depending on which criterion is used, if indeed an agreed criterion variable can be decided upon.

This study had some limitations. First, the prevalence rates were all based on self-report data so the absence/presence of the symptoms cannot be clinically verified. Second, there are many different ways the diagnostic algorithms could have been developed and applied to the factor analytically derived models; we tried to develop logical and sensible diagnostic rules, but the authors of the factor models may have alternative specifications. Third, these analyses were all based on clinical treatment seeking samples with high rates of PTSD and traumatisation. These findings may not generalise to other populations. Fourth, the diagnostic rates and the odds ratios associated with childhood maltreatment are, to some degree, confounded in that higher prevalence rates are associated with larger effects for childhood maltreatment. Fifth, the questions used to assess exposure to childhood maltreatment did not include behavioural descriptions and so endorsement may depend on the participant's perception of 'abuse'. Finally, in order to fulfil the aims of this study PTSD was treated as a discrete diagnostic entity that can be defined in terms of the symptom set specified by the DSM-5. This approach fails to acknowledge the challenges to the diagnostic boundaries of the disorder (Maercker et al., 2013), the myriad potential patterns of comorbidity (Young, Lareau, & Pierre, 2014), the possibility that there is no single 'correct' model, but rather that different models represent different subpopulations (Elklit & Shevlin, 2007), and the evidence that different traumatic experiences can produce qualitatively (rather than strictly quantitatively) different patterns of symptom endorsement (Forbes et al., 2012).

Overall, this paper has argued that continued factor analytic studies of the latent structure of PTSD symptoms, without addressing the diagnostic implications, does not advance our understanding of how to diagnose PTSD, aid in the development of interventions, or help explain the co-morbidities with other disorders. Such research may tantalise with the promise of such valuable outcomes, but ultimately fails to deliver. Elhai and Palmieri (2011) offered sensible and pragmatic guidance on future directions for CFA studies of PTSD symptoms and we suggest that engaging with the diagnostic implications associated with factor analytic models should be added to their list of good practice.

## References

- Armour, C., Müllerová, J., & Elhai, J. D. (2016). A systematic literature review of PTSD's latent structure in the Diagnostic and Statistical Manual of Mental Disorders: DSM-IV to DSM-5. *Clinical Psychology Review, 44*, 60–74.
- Armour, C., Tsai, J., Durham, T. A., Charak, R., Biehn, T. L., & Elhai, J. D. (2015). Dimensional structure of DSM-5 posttraumatic stress symptoms: support for a hybrid anhedonia and externalizing behaviors model. *Journal of Psychiatric Research, 61*, 106–113.
- Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.
- Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress, 28*, 489–498. <http://dx.doi.org/10.1002/jts.22059>
- Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., & Keane, T. M. (2015). Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (PCL-5) in veterans. *Psychological Assessment, 27*, 1037–1047. <http://dx.doi.org/10.1037/pas0000254>
- Elhai, J. D., Biehn, T. L., Armour, C., Klopper, J. J., Frueh, B. C., & Palmieri, P. A. (2011). Evidence for a unique PTSD construct represented by PTSD's D1–D3 symptoms. *Journal of Anxiety Disorders, 25*, 340–345. <http://dx.doi.org/10.1016/j.janxdis.2010.10.007>
- Elhai, J. D., & Palmieri, P. A. (2011). Posttraumatic stress disorder symptom instruments and factor structure: An update on the current literature and advancing a research agenda. *Journal of Anxiety Disorders, 25*, 849–854. <http://dx.doi.org/10.1016/j.janxdis.2011.04.007>
- Elklit, A., & Shevlin, M. (2007). The structure of PTSD symptoms: A test of alternative models using confirmatory factor analysis. *British Journal of Clinical Psychology, 46*, 299–313.
- Forbes, D., Fletcher, S., Parslow, R., Phelps, A., O'Donnell, M., Bryant, R. A., et al. (2012). Trauma at the hands of another: Longitudinal study of differences in the posttraumatic stress disorder symptom profile following interpersonal compared with noninterpersonal trauma. *Journal of Clinical Psychiatry, 73*, 372–376. <http://dx.doi.org/10.4088/JCP.13m08374>
- Ford, J. D., Elhai, J. D., Ruggiero, K. J., & Frueh, B. C. (2009). Refining posttraumatic stress disorder diagnosis: Evaluation of symptom criteria with the National Survey of Adolescents. *Journal of Clinical Psychiatry, 70*, 748–755. <http://dx.doi.org/10.4088/JCP.08m04692>
- Frances, A. J., & Nardo, J. M. (2013). ICD-11 should not repeat the mistakes made by DSM-5. *British Journal of Psychiatry, 203*, 1–2. <http://dx.doi.org/10.1192/bjp.bp.113.127647>
- Galatzer-Levy, I. R., & Bryant, R. A. (2013). 636,120 ways to have posttraumatic stress disorder. *Perspectives on Psychological Science, 8*, 651–662. <http://dx.doi.org/10.1177/1745691613504115>
- Galatzer-Levy, I. R., Karstoft, K. I., Statnikov, A., & Shalev, A. Y. (2014). Quantitative forecasting of PTSD from early trauma responses: A machine learning application. *Journal of Psychiatric Research, 59*, 68–76. <http://dx.doi.org/10.1016/j.jpsychires.2014.08.017>
- Ganz, M. L., Wintfeld, N., Li, Q., Alas, V., Langer, J., & Hammer, M. (2014). The association of body mass index with the risk of type 2 diabetes: A case-control study nested in an electronic health records system in the United States. *Diabetology and Metabolic Syndrome, 50*. <http://dx.doi.org/10.1186/1758-5996-6-50>
- Gray, M., Litz, B., Hsu, J., & Lombardo, T. (2004). Psychometric properties of the Life Events Checklist (PDF). *Assessment, 11*, 330–341. <http://dx.doi.org/10.1177/1073191104269954>
- Hansen, M., Hyland, P., Armour, C., Elklit, A., & Shevlin, M. (2015). Less is more? Assessing the validity of the ICD-11 model of PTSD across multiple trauma samples. *European Journal of Psychotraumatology, 6*, 28766. <http://dx.doi.org/10.3402/ejpt.v6.28766>
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structural modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424–453.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.
- Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8 User's reference guide*. Chicago: Scientific Software.
- King, D. W., Orazem, R. J., King, L. A., Lauterbach, D., Hebenstreit, C. L., & Shalev, A. Y. (2009). Factor structure of posttraumatic stress disorder as measured by the Impact of Events Scale-Revised: Stability across cultures and time. *Psychological Trauma: Theory, Research, Practice, and Policy, 1*, 173–187.
- Maercker, A., Brewin, C. R., Bryant, R. A., Cloitre, M., van Ommeren, M., Jones, L. M., & ... Reed, G. M. (2013). Diagnosis and classification of disorders specifically associated with stress: proposals for ICD-11. *World Psychiatry, 12*(3), 198–206. <http://doi.org/10.1002/wps.20057>
- Muthén, L. K., & Muthén, B. O. (2013). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Raftery, A. E. (1996). Approximate Bayes factors and accounting for model uncertainty in generalized linear models. *Biometrika, 83*, 251–266.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.
- Simms, L. J., Watson, D., & Doebbeling, B. N. (2002). Confirmatory factor analyses of posttraumatic stress symptoms in deployed and nondeployed veterans of the Gulf war. *Journal of Abnormal Psychology, 111*, 637–647.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173–180. [http://dx.doi.org/10.1207/s15327906mbr2502\\_4](http://dx.doi.org/10.1207/s15327906mbr2502_4)
- Wang, M. C., & Li, S. (2013). ROC analysis for multiple markers with tree-based classification. *Lifetime Data Analysis, 19*, 257–277. <http://dx.doi.org/10.1007/s10985-012-9233-5>
- Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2013). *The PTSD Checklist for DSM-5 (PCL-5) Scale* available from the National Center for PTSD at [www.ptsd.va.gov](http://www.ptsd.va.gov).
- Young, G., Lareau, C., & Pierre, B. (2014). One quintillion ways to Have PTSD comorbidity: Recommendations for the disordered DSM-5. *Psychological Injury and Law, 7*, 61–74. <http://dx.doi.org/10.1007/s12207-014-9186-y>
- Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology, 30*, 165–200. <http://dx.doi.org/10.1111/0081-1750.00078>